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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 August 2021 (19.08.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/163706 A1

(51) International Patent Classification:
G16B 20/00 (2006.01) *G16B 40/00* (2019.01)
G06N 20/00 (2019.01) *G06N 3/08* (2006.01)
G16H 50/20 (2018.01) *G16H 50/70* (2018.01)

(21) International Application Number:
PCT/US2021/018263

(22) International Filing Date:
16 February 2021 (16.02.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/977,015	14 February 2020 (14.02.2020)	US
63/014,515	23 April 2020 (23.04.2020)	US
63/052,363	15 July 2020 (15.07.2020)	US
63/145,305	03 February 2021 (03.02.2021)	US

(71) Applicant: **CARIS MPI, INC.** [US/US]; 750 West John Carpenter Freeway, Suite 800, Irving, Texas 75039 (US).

(72) Inventors: **ABRAHAM, Jim**; 1112 Longford Circle, Southlake, Texas 76092 (US). **SPETZLER, David**; 6045 N. 40th Street, Paradise Valley, Arizona 85253 (US).

(74) Agent: **DEYOUNG, Janice Kugler** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: PANOMIC GENOMIC PREVALENCE SCORE

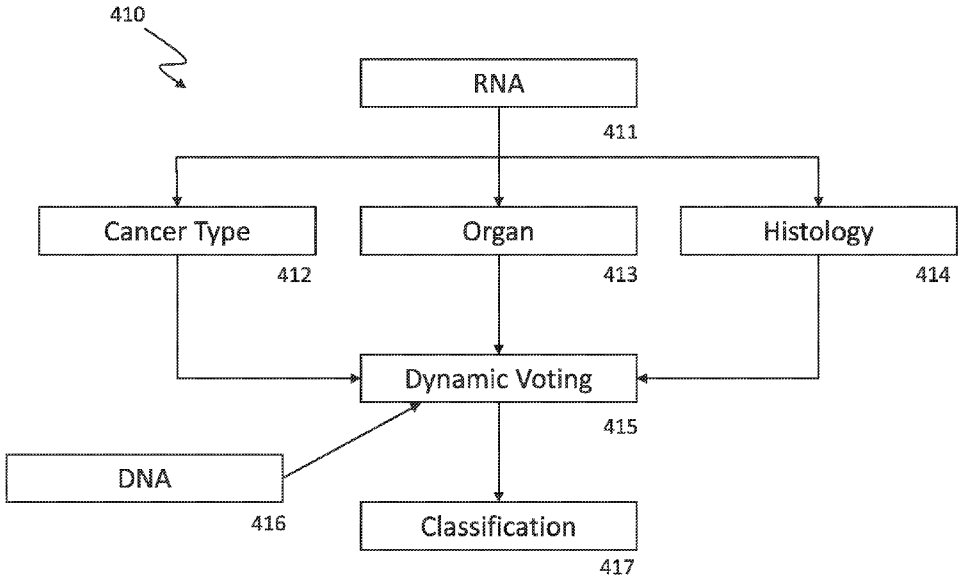


FIG. 4B

(57) Abstract: Comprehensive molecular profiling provides a wealth of data concerning the molecular status of patient samples. Such data can be compared to patient response to treatments to identify biomarker signatures that predict response or non-response to such treatments. Here, we used molecular profiling data to identify biomarker signatures (biosignatures) that predict a tumor primary lineage, cancer category or type, organ group and/or histology. The signature may use genomic and transcriptome level information.

WO 2021/163706 A1

[Continued on next page]

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

PANOMIC GENOMIC PREVALENCE SCORE

CLAIM OF PRIORITY

This application claims the benefit of U.S. Provisional Patent Application Serial Nos. 62/977,015, filed on February 14, 2020; 63/014,515, filed on April 23, 2020; 63/052,363, filed on July 15, 2020; and 63/145,305, filed on February 3, 2021; the entire contents of which applications are hereby incorporated by reference in their entirety.

This application is related to International Patent Publication WO/2020/146554, entitled Genomic Profiling Similarity and based on International Patent Application PCT/US2020/012815 filed on January 8, 2020, the entire contents of which application is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

The present disclosure relates to the fields of data structures, data processing, and machine learning, and their use in precision medicine, e.g., tumor characterization including without limitation the use of molecular profiling to predict an attribute of a biological sample such as the primary origin, organ type, histology and/or cancer type.

BACKGROUND

Carcinoma of Unknown Primary (CUP) represents a clinically challenging heterogeneous group of metastatic malignancies in which a primary tumor remains elusive despite extensive clinical and pathologic evaluation. Approximately 2-4% of cancer diagnoses worldwide comprise CUP. *See, e.g.,* Varadhachary. New Strategies for Carcinoma of Unknown Primary: the role of tissue of origin molecular profiling. Clin Cancer Res. 2013 Aug 1;19(15):4027-33. In addition, some level of diagnostic uncertainty with respect to an exact tumor type classification is a frequent occurrence across oncologic subspecialties. Efforts to secure a definitive diagnosis can prolong the diagnostic process and delay treatment initiation. Furthermore, CUP is associated with poor outcome which might be explained by use of suboptimal therapeutic intervention. Immunohistochemical (IHC) testing is the gold standard method to diagnose the site of tumor origin, especially in cases of poorly differentiated or undifferentiated tumors. Assessing the accuracy in challenging cases and performing a meta-analysis of these studies reported that IHC analysis had an accuracy of 66% in the characterization of metastatic tumors. *See, e.g.,* Brown RW, et al. Immunohistochemical identification of tumor markers in metastatic adenocarcinoma: a diagnostic adjunct in the determination of primary site. Am J Clin Pathol 1997, 107:12e19; Dennis JL, et al. Markers of adenocarcinoma characteristic of the site of origin: development of a diagnostic algorithm. Clin Cancer Res 2005, 11:3766e3772; Gamble AR, et al. Use of tumour marker immunoreactivity to identify primary site of metastatic cancer. BMJ 1993, 306:295e298; Park SY, et al. Panels of immunohistochemical markers help

determine primary sites of metastatic adenocarcinoma. Arch Pathol Lab Med 2007, 131:1561e1567; DeYoung BR, Wick MR. Immunohistologic evaluation of metastatic carcinomas of unknown origin: an algorithmic approach. Semin Diagn Pathol 2000, 17:184e193; Anderson GG, Weiss LM. Determining tissue of origin for metastatic cancers: meta-analysis and literature review of immunohistochemistry performance. Appl Immunohistochem Mol Morphol 2010, 18:3e8. Since therapeutic regimes can be dependent upon diagnosis, this represents an important unmet clinical need.

To address these challenges, assays aiming at tissue-of-origin (TOO) identification based on assessment of differential gene expression have been developed and tested clinically. However, integration of such assays into clinical practice is hampered by relatively poor performance characteristics (from 83% to 89%) and limited sample availability. *See, e.g.,* Pillai R, et al. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. J Mol Diagn 2011, 13:48e56; Rosenwald S, et al. Validation of a microRNA-based qRT-PCR test for accurate identification of tumor tissue origin. Mod Pathol 2010, 23:814e823; Kerr SE, et al. Multisite validation study to determine performance characteristics of a 92-gene molecular cancer classifier. Clin Cancer Res 2012, 18:3952e3960; Kucab JE, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019 May 2;177(4):821-836.e16. For example, a recent commercial RNA-based assay has a sensitivity of 83% in a test set of 187 tumors and confirmed results on only 78% of a separate 300 sample validation set. See Hainsworth JD, et al, Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. J Clin Oncol. 2013 Jan 10;31(2):217-23. This may, at least in part, be a consequence of limitations of typical RNA-based assays in regards to normal cell contamination, RNA stability, and dynamics of RNA expression. Thus, there is a need for more robust approaches to TOO identification to aid cancer patients, particularly but not limited to CUP.

Machine learning models can be configured to analyze labeled training data and then draw inferences from the training data. Once the machine learning model has been trained, sets of data that are not labeled may be provided to the machine learning model as an input. The machine learning model may process the input data, e.g., molecular profiling data, and make predictions about the input based on inferences learned during training. The present disclosure further provides a voting methodology to combine multiple classifier models to achieve more accurate classification than that achieved by use a single model.

Comprehensive molecular profiling provides a wealth of data concerning the molecular status of patient samples. We have performed such profiling on well over 100,000 tumor patients from practically all cancer lineages. Patient and molecular data can be processed using machine learning algorithms to identify additional biomarker signatures that can be used to characterize various phenotypes of interest. Here, this "next generation profiling" (NGP) approach has been applied to

build models to predict an attribute of a biological sample, including without limitation such as the primary origin, organ type, histology and/or cancer type.

SUMMARY

Comprehensive molecular profiling provides a wealth of data concerning the molecular status of patient samples. Such data can be compared to patient response to treatments to identify biomarker signatures that predict response or non-response to such treatments. Herein we provide systems and methods to predict attributes of a patient sample, including without limitation a tissue-of-origin (TOO).

In an aspect, the disclosure provides a data processing apparatus for generating input data structure for use in training a machine learning model to predict at least one attribute of a biological sample, wherein the at least one attribute is selected from the group comprising a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the data processing apparatus including one or more processors and one or more storage devices storing instructions that when executed by the one or more processors cause the one or more processors to perform operations, the operations comprising: obtaining, by the data processing apparatus one or more biomarker data structures and one or more sample data structures; extracting, by the data processing apparatus, first data representing one or more biomarkers associated with the sample from the one or more biomarker data structures, second data representing the sample data from the one or more sample data structures, and third data representing a predicted at least one attribute; generating, by the data processing apparatus, a data structure, for input to a machine learning model, based on the first data representing the one or more biomarkers and the second data representing the predicted at least one attribute and sample; providing, by the data processing apparatus, the generated data structure as an input to the machine learning model; obtaining, by the data processing apparatus, an output generated by the machine learning model based on the machine learning model's processing of the generated data structure; determining, by the data processing apparatus, a difference between the third data representing a predicted predicted at least one attribute for the sample and the output generated by the machine learning model; and adjusting, by the data processing apparatus, one or more parameters of the machine learning model based on the difference between the third data representing a predicted predicted at least one attribute for the sample and the output generated by the machine learning model. In some embodiments, the set of one or more biomarkers include one or more biomarkers listed in any one of **Tables 121-129**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof. In some embodiments, the set of one or more biomarkers include each of the biomarkers. In some embodiments, the set of one or more biomarkers includes at least one of these biomarkers, optionally wherein the set of one or more biomarkers comprises each of the biomarkers in

Table 118, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**.

In an aspect, the disclosure provides a data processing apparatus for generating input data structure for use in training a machine learning model to predict at least one attribute of a biological sample, wherein the at least one attribute is selected from the group comprising a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the data processing apparatus including one or more processors and one or more storage devices storing instructions that when executed by the one or more processors cause the one or more processors to perform operations, the operations comprising: obtaining, by the data processing apparatus, a first data structure that structures data representing a set of one or more biomarkers associated with a biological sample from a first distributed data source, wherein the first data structure includes a key value that identifies the sample; storing, by the data processing apparatus, the first data structure in one or more memory devices; obtaining, by the data processing apparatus, a second data structure that structures data representing data for the at least one attribute for the sample having the one or more biomarkers from a second distributed data source, wherein the data for the at least one attribute includes data identifying a sample, at least one attribute, and an indication of the predicted at least one attribute, wherein second data structure also includes a key value that identifies the sample; storing, by the data processing apparatus, the second data structure in the one or more memory devices; generating, by the data processing apparatus and using the first data structure and the second data structure stored in the memory devices, a labeled training data structure that includes (i) data representing the set of one or more biomarkers and the sample, and (ii) a label that provides an indication of a predicted at least one attribute, wherein generating, by the data processing apparatus and using the first data structure and the second data structure includes correlating, by the data processing apparatus, the first data structure that structures the data representing the set of one or more biomarkers associated with the sample with the second data structure representing predicted at least one attribute data for the sample having the one or more biomarkers based on the key value that identifies the subject; and training, by the data processing apparatus, a machine learning model using the generated label training data structure, wherein training the machine learning model using the generated labeled training data structure includes providing, by the data processing apparatus and to the machine learning model, the generated label training data structure as an input to the machine learning model. In some embodiments, the operations further comprise: obtaining, by the data processing apparatus and from the machine learning model, an output generated by the machine learning model based on the machine learning model's processing of the generated labeled training data structure; and determining, by the data processing apparatus, a difference between the output generated by the machine learning model and the label that provides an indication of the predicted at least one attribute. In some embodiments, the operations further comprise: adjusting, by the data processing apparatus, one or more parameters of the machine learning model based on the determined difference between the output generated by the

machine learning model and the label that provides an indication of the predicted at least one attribute. In some embodiments, the set of one or more biomarkers include one or more biomarkers listed in any one of **Tables 121-129**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof. In some embodiments, the set of one or more biomarkers include each of the biomarkers. In some embodiments, the set of one or more biomarkers includes at least one of these biomarkers, optionally wherein the set of one or more biomarkers comprises each of the biomarkers in **Table 118**, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**.

The disclosure also provides a method comprising steps that correspond to each of the operations described above. The disclosure also provides a system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described above. The disclosure also provides a non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described above.

In an aspect, the disclosure provides a method for determining at least one attribute of a biological sample, wherein the at least one attribute is selected from the group comprising a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the method comprising: for each particular machine learning model of a plurality of machine learning models that have each been trained to perform an prediction operation between received input data representing a sample and the at least one attribute: providing, to the particular machine learning model, input data representing a sample of a subject, wherein the sample was obtained from tissue or an organ of the subject; and obtaining output data, generated by the particular machine learning model based on the particular machine learning model's processing the provided input data, that represents a probability or likelihood that the sample represented by the provided input data corresponds to the at least one attribute; providing, to a voting unit, the output data obtained for each of the plurality of machine learning models, wherein the provided output data includes data representing initial sample attributes determined by each of the plurality of machine learning models; and determining, by the voting unit and based on the provided output data, the predicted at least one attribute. In some embodiments, the predicted at least one attribute is determined by applying a majority rule to the provided output data, by using the provided output data as input into a dynamic voting model, or a combination thereof. In some embodiments, the determining, by the voting unit and based on the provided output data, the predicted at least one attribute comprises: determining, by the voting unit, a number of occurrences of each initial attribute class of the multiple candidate attribute classes; and selecting, by the voting unit, the initial attribute class of the multiple candidate attribute classes having the highest number of

occurrences. In some embodiments, each machine learning model of the plurality of machine learning models comprises a random forest classification algorithm, boosted tree, support vector machine, logistic regression, k-nearest neighbor model, artificial neural network, naïve Bayes model, quadratic discriminant analysis, Gaussian processes model, or any combination thereof. In some embodiments, each machine learning model of the plurality of machine learning models comprises a random forest classification algorithm. In some embodiments, each machine learning model of the plurality of machine learning models comprises a boosted tree classification algorithm. In some embodiments, the plurality of machine learning models includes multiple representations of a same type of classification algorithm. In some embodiments, the input data represents a description of (i) sample attributes and (ii) origins. In some embodiments, the multiple candidate attribute classes include at least one class for prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas, rectum, colon, breast, intrahepatic bile duct, cecum, gastroesophageal junction, frontal lobe, kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin. In some embodiments, the multiple candidate attribute classes include at least at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or all 21 of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma. In some embodiments, the sample attributes includes one or more biomarkers for the sample, wherein optionally the one or more biomarkers comprises one or more biomarkers listed in any one of **Tables 121-129**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof. In some embodiments, the set of one or more biomarkers include each of the biomarkers. In some embodiments, the set of one or more biomarkers includes at least one of these biomarkers, optionally wherein the set of one or more biomarkers comprises each of the biomarkers in **Table 118**, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**. In some embodiments, the input data further includes data representing a description of the sample and/or subject. The disclosure also provides a system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described above. The disclosure also provides a non-transitory computer-readable medium storing software comprising

instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described above.

1. In an aspect, the disclosure provides a method for classifying a biological sample, the method comprising: obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample; obtaining, by one or more computers, second data representing another initial classification for the biological sample that were previously determined based on DNA sequences of the biological sample; providing, by one or more computers, at least a portion of the first data and the second data as an input to a dynamic voting engine that has been trained to predict a target biological sample classification based on processing of multiple initial biological sample classifications; processing, by one or more computers, the provided input data through the dynamic voting engine; obtaining, by one or more computers, output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the provided input data; and determining, by one or more computers, a target biological sample classification for the biological sample based on the obtained output data. In some embodiments, the obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample comprises: obtaining data representing a cancer type classification for the biological sample based the RNA sequences of the biological sample; obtaining data representing an organ from which the biological sample originated based on the RNA sequences of the biological sample; and obtaining data representing a histology for the biological sample based on the RNA sequences of the biological sample, and wherein providing at least a portion of the first data and the second data as an input to the dynamic voting engine comprises: providing the obtained data representing the cancer type classification, the obtained data representing the organ from which the biological sample originated, the obtained data representing the histology, and the second data as an input to the dynamic voting engine. In some embodiments, the dynamic voting engine comprises one or more machine learning model. In some embodiments, training the dynamic voting engine comprises: obtaining a labeled training data item that includes (I) one or more initial classifications that include data indicating a cancer classification type, data indicating an initial organ of origin, data indicating a histology, or data indicating output of a DNA analysis engine and (II) a target biological sample classification, generating training input data for input to the dynamic voting engine based on the obtained training data item, processing the generated training input data through the dynamic voting engine, obtaining output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the generated training input data, and adjusting one or more parameters of the dynamic voting engine based on the level of similarity between the output data and the label of the obtained training data item.

In some embodiments, previously determining an initial classification for the biological sample based on DNA sequences of the biological sample comprises: receiving, by one or more computers, a biological signature representing the biological sample that was obtained from a cancerous neoplasm in a first portion of a body, wherein the model includes a cancerous biological signature for each of multiple different types of cancerous biological samples, wherein each of the cancerous biological signatures include at least a first cancerous biological signature representing a molecular profile of a cancerous biological sample from the first portion of one or more other bodies and a second cancerous biological signature representing a molecular profile of a cancerous biological sample from a second portion of one or more other bodies; performing, by one or more computers and using a pairwise-analysis model, pairwise analysis of the biological signature using the first cancerous biological signature and the second cancerous biological signature; generating, by one or more computers and based on the performed pairwise analysis, a likelihood that the cancerous neoplasm in the first portion of the body was caused by cancer in a second portion of the body; and storing, by one or more computers, the generated likelihood in a memory device. The disclosure also provides a system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described above. The disclosure also provides a non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described above.

In an aspect, the disclosure provides a method comprising: (a) obtaining a biological sample from a subject having a cancer; (b) performing at least one assay on the sample to assess one or more biomarkers, thereby obtaining a biosignature for the sample; (c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises at least one pre-determined biosignature indicative of at least one attribute, and wherein the at least one attribute of the cancer is selected from the group comprising primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof; (d) processing, by one or more computers, the provided biosignature through the model; and (e) outputting from the model a prediction of the at least one attribute of the cancer.

In the methods provided herein, the biological sample may comprise formalin-fixed paraffin-embedded (FFPE) tissue, fixed tissue, a core needle biopsy, a fine needle aspirate, unstained slides, fresh frozen (FF) tissue, formalin samples, tissue comprised in a solution that preserves nucleic acid or protein molecules, a fresh sample, a malignant fluid, a bodily fluid, a tumor sample, a tissue sample, or any combination thereof. In some embodiments, the biological sample comprises cells from a solid tumor, a bodily fluid, or a combination thereof. In some embodiments, the bodily fluid comprises a malignant fluid, a pleural fluid, a peritoneal fluid, or any combination thereof. In some embodiments, the bodily fluid comprises peripheral blood, sera, plasma, ascites, urine, cerebrospinal fluid (CSF), sputum, saliva, bone marrow, synovial fluid, aqueous humor, amniotic fluid, cerumen,

breast milk, broncheoalveolar lavage fluid, semen, prostatic fluid, Cowper's fluid, pre-ejaculatory fluid, female ejaculate, sweat, fecal matter, tears, cyst fluid, pleural fluid, peritoneal fluid, pericardial fluid, lymph, chyme, chyle, bile, interstitial fluid, menses, pus, sebum, vomit, vaginal secretions, mucosal secretion, stool water, pancreatic juice, lavage fluids from sinus cavities, bronchopulmonary aspirates, blastocyst cavity fluid, or umbilical cord blood.

In the methods provided herein, performing the at least one assay in step (b) may comprise determining a presence, level, or state of a protein or nucleic acid for each of the one or more biomarkers, wherein optionally the nucleic acid comprises deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or a combination thereof. In some embodiments, the presence, level or state of at least one of the proteins is determined using a technique selected from immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof, wherein optionally the presence, level or state of all of the proteins is determined using the technique; and/or the presence, level or state of at least one of the nucleic acids is determined using a technique selected from polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, whole genome sequencing, whole transcriptome sequencing, or any combination thereof, wherein optionally the presence, level or state of all of the nucleic acids is determined using the technique. In some embodiments, the state of the nucleic acid comprises a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof. In some embodiments, the state of the nucleic acid consists of or comprises a copy number. In some embodiments, the at least one assay comprises next-generation sequencing, wherein optionally the next-generation sequencing is used to assess: i) at least one of the genes, genomic information / signatures, and fusion transcripts in any of **Tables 121-130**, or any combination thereof; ii) at least one of the genes and/or transcripts in any table selected from **Tables 117-120**, INSM1, and any combination thereof; iii) the whole exome or substantially the whole exome; iv) the whole transcriptome or substantially the whole transcriptome; v) at least one gene in any table selected from **Tables 2-116**, and any combination thereof; or vi) any combination thereof.

In the methods provided herein, predicting the at least one attribute of the cancer may comprise determining a probability that the attribute is each member of a plurality of such attributes and selecting the attribute with the highest probability.

In some embodiments of the methods provided herein, the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, or all 38 of prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas,

rectum, colon, breast, intrahepatic bile duct, cecum, gastroesophageal junction, frontal lobe, kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin. In some embodiments, the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or all 21 of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma. In some embodiments, the cancer/disease type consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or all 28 of adrenal cortical carcinoma; bile duct, cholangiocarcinoma; breast carcinoma; central nervous system (CNS); cervix carcinoma; colon carcinoma; endometrium carcinoma; gastrointestinal stromal tumor (GIST); gastroesophageal carcinoma; kidney renal cell carcinoma; liver hepatocellular carcinoma; lung carcinoma; melanoma; meningioma; Merkel; neuroendocrine; ovary granulosa cell tumor; ovary, fallopian, peritoneum; pancreas carcinoma; pleural mesothelioma; prostate adenocarcinoma; retroperitoneum; salivary and parotid; small intestine adenocarcinoma; squamous cell carcinoma; thyroid carcinoma; urothelial carcinoma; uterus. In some embodiments, the organ group consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or all 17 of adrenal gland; bladder; brain; breast; colon; eye; female genital tract and peritoneum (FGTP); gastroesophageal; head, face or neck, NOS; kidney; liver, gallbladder, ducts; lung; pancreas; prostate; skin; small intestine; thyroid. In some embodiments, the histology consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or all 29 of adenocarcinoma, adenoid cystic carcinoma, adenosquamous carcinoma, adrenal cortical carcinoma, astrocytoma, carcinoma, carcinosarcoma, cholangiocarcinoma, clear cell carcinoma, ductal carcinoma in situ (DCIS), glioblastoma (GBM), GIST, glioma, granulosa cell tumor, infiltrating lobular carcinoma, leiomyosarcoma, liposarcoma, melanoma, meningioma, Merkel cell carcinoma, mesothelioma, neuroendocrine, non-small cell carcinoma, oligodendroglioma, sarcoma, sarcomatoid carcinoma, serous, small cell carcinoma, squamous.

In some embodiments of the methods provided herein, the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, wherein optionally the at least one attribute is a cancer/disease type, comprises selections of biomarkers according to **Table 118**, wherein optionally: i. a pre-determined biosignature indicative of adrenal cortical carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from INHA, MIB1, SYP, CDH1, NKX3-1, CALB2, KRT19, MUC1, S100A5,

CD34, TMPRSS2, KRT8, NCAM2, ARG1, TG, NCAM1, SERPINA1, PSAP, TPM3, and ACVRL1; ii. a pre-determined biosignature indicative of bile duct, cholangiocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from HNF1B, VIL1, SERPINA1, ESR1, ANO1, SOX2, MUC4, S100A2, KRT5, KRT7, CNN1, AR, ENO2, S100A9, NKX2-2, SATB2, PSAP, S100A6, CALB2, and TMPRSS2; iii. a pre-determined biosignature indicative of breast carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, ANKRD30A, KRT15, KRT7, S100A2, PAX8, MUC4, KRT18, HNF1B, S100A1, PIP, SOX2, MDM2, MUC5AC, PMEL, TFF1, KRT16, KRT6B, S100A6, and SERPINB5; iv. a pre-determined biosignature indicative of central nervous system (CNS) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT18, KRT8, SOX2, ANO1, NCAM1, PDPN, NKX2-2, KRT19, S100A14, S100A11, S100A1, MSH2, CEACAM1, GPC3, ERBB2, TG, KRT7, CGB3, and S100A2; v. a pre-determined biosignature indicative of cervix carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ESR1, CDKN2A, CCND1, LIN28A, PGR, SMARCB1, CEACAM4, S100B, FUT4, PSAP, MUC2, MDM2, NCAM1, SATB2, TNFRSF8, CD79A, S100A13, VHL, CD3G, and TPSAB1; vi. a pre-determined biosignature indicative of colon carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, KRT7, MUC2, KRT20, MUC1, SATB2, VIL1, CEACAM5, CDH17, S100A6, CEACAM20, KRT6B, TFF3, FUT4, BCL2, KRT6A, KRT18, CEACAM18, TFF1, and MLH1; vii. a pre-determined biosignature indicative of endometrium carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, PGR, ESR1, VHL, CALD1, LIN28B, NAPSA, KRT5, S100A6, DES, FLI1, DSC3, S100P, CEACAM16, PDPN, ARG1, TLE1, WT1, BCL6, and MLH1; viii. a pre-determined biosignature indicative of gastrointestinal stromal tumor (GIST) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ANO1, SDC1, KRT19, MUC1, KRT8, ACVRL1, KIT, CDH1, S100A2, KRT7, ERBB2, S100A16, ENO2, S100A9, TPSAB1, KRT17, PAX8, PGR, ESR1, and VHL; ix. a pre-determined biosignature indicative of gastroesophageal carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FUT4, CDX2, SERPINB5, MUC5AC, AR, TFF1, NCAM2, TFF3, ISL1, ANO1, VIL1, PAX8, SOX2, CEACAM6, S100A13, ENO2, NAPSA, TPSAB1, S100B, and CD34; x. a pre-determined biosignature indicative of kidney renal cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, CDH1, CDKN2A, S100P, S100A14, HAVCR1, HNF1B, KL, KRT7, MUC1, POU5F1, VHL, PAX2, AMACR, BCL6, S100A13, CA9, MDM2, SALL4, and SYP; xi. a pre-determined biosignature indicative of liver hepatocellular carcinoma consists of, comprises, or comprises at least,

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, CEACAM16, KRT19, AFP, MUC4, CEACAM5, MSH2, BCL6, DSC3, KRT15, S100A6, CEACAM20, GPC3, MUC1, CD34, VIL1, ERBB2, POU5F1, KRT18, and KRT16; xii. a pre-determined biosignature indicative of lung carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NAPSA, SOX2, CEACAM7, KRT7, S100A10, CEACAM6, S100A1, PAX8, AR, VHL, S100A13, CD99L2, KRT5, MUC1, CEACAM1, SFTPA1, TMPRSS2, TFF1, KRT15, and MUC4; xiii. a pre-determined biosignature indicative of melanoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT8, PMEL, KRT19, MUC1, MLANA, S100A14, S100A13, MITF, S100A1, VIM, CDKN2A, ACVRL1, MS4A1, POU5F1, TPM1, UPK3A, S100P, GATA3, and CEACAM1; xiv. a pre-determined biosignature indicative of meningioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SDC1, KRT8, ANO1, VIM, S100A14, S100A2, CEACAM1, MSH2, PGR, KRT10, TP63, CD5, INHA, CDH1, CCND1, MDM2, KRT16, SPN, SMARCB1, and S100A9; xv. a pre-determined biosignature indicative of Merkel cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, ERBB2, S100A12, S100A14, MYOG, SDC1, KRT7, S100PBP, MME, TMPRSS2, CEACAM5, CPS1, CR1, MUC4, CEACAM4, CA9, ENO2, FLI1, LIN28B, and MLANA; xvi. a pre-determined biosignature indicative of neuroendocrine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, ISL1, ENO2, POU5F1, TFF3, SYP, TPM4, S100A1, S100Z, MUC4, MPO, DSC3, CEACAM4, S100A7, ERBB2, CDX2, S100A11, KRT10, CEACAM5, and CEACAM3; xvii. a pre-determined biosignature indicative of ovary granulosa cell tumor consists of, comprises, or comprises at least, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FOXL2, SDC1, MSH6, MUC1, KRT8, PGR, MME, SERPINA1, FLI1, S100B, CEACAM21, AMACR, KRT1, SFTPA1, TPM1, CALCA, S100A11, NCAM1, ISL1, and ENO2; xviii. a pre-determined biosignature indicative of ovary, fallopian, peritoneum consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from WT1, PAX8, INHA, TFE3, S100A13, FOXL2, TLE1, MSLN, POU5F1, CEACAM3, ALPP, S100A10, FUT4, NKX3-1, CEACAM5, SOX2, ESR1, ENO2, ACVRL1, and SYP; xix. a pre-determined biosignature indicative of pancreas carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, GATA3, ANO1, SERPINA1, ISL1, MUC5AC, FUT4, SMAD4, CD5, CALB2, S100A4, SMN1, ESR1, HNF1B, AMACR, MSH2, PDPN, MSLN, TFF1, and KRT6C; xx. a pre-determined biosignature indicative of pleural mesothelioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from UPK3B, CALB2, WT1, SMARCB1, PDPN, INHA, CEACAM1, MSLN, KRT5, CA9, S100A13, SF1, CDH1,

CDKN2A, FLI1, SYP, CEACAM3, CPS1, SATB2, and BCL6; xxi. a pre-determined biosignature indicative of prostate adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT7, KLK3, NKX3-1, AMACR, S100A5, MUC1, MUC2, UPK3A, KL, CPS1, MSLN, PMEL, CNN1, SERPINA1, KRT2, CGB3, TMPRSS2, CEACAM6, SDC1, and AR; xxii. a pre-determined biosignature indicative of retroperitoneum consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT18, KRT8, TPM1, S100A14, CD34, TPM4, CDH1, CNN1, SDC1, AR, MDM2, KIT, TLE1, CPS1, CDK4, UPK3A, TMPRSS2, TPM3, and CEACAM1; xxiii. a pre-determined biosignature indicative of salivary and parotid consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ENO2, PIP, TPM1, KRT14, S100A1, ERBB2, TFF1, ALPP, DSC3, CTNNB1, CALB2, SALL4, ANO1, CEACAM16, HNF1B, KIT, ARG1, CEACAM18, TMPRSS2, and HAVCR1; xxiv. a pre-determined biosignature indicative of small intestine adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, DES, MUC2, CDH17, CEACAM5, SERPINA1, KRT20, HNF1B, ESR1, ARG1, CD5, TLE1, PMEL, SOX2, SFTPA1, MME, CD99L2, MPO, S100P, and CA9; xxv. a pre-determined biosignature indicative of squamous cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, SOX2, KRT6A, KRT17, S100A1, CD3G, SFTPA1, AR, KRT5, SDC1, KRT20, DSC3, CNN1, MSH2, ESR1, S100A2, SERPINB5, PDPN, S100A14, and TPM3; xxvi. a pre-determined biosignature indicative of thyroid carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TG, PAX8, CPS1, S100A2, TPSAB1, CALB2, HNF1B, INHA, ARG1, CNN1, CDK4, VIM, CEACAM5, TLE1, TFF3, KRT8, S100P, FOXL2, MUC1, and GATA3; xxvii. a pre-determined biosignature indicative of urothelial carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, UPK2, KRT20, MUC1, S100A2, CPS1, TP63, CALB2, MITF, S100P, SERPINA1, DES, CTNNB1, MSLN, SALL4, VHL, KRT7, CD2, PAX8, and UPK3A; and/or xxviii. a pre-determined biosignature indicative of uterus consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT18, NCAM1, DES, FOXL2, CD79A, S100A14, ESR1, MSLN, MITF, UPK3B, TPM1, ENO2, S100P, MLH1, KRT8, CDH1, TPM4, SATB2, and MDM2.

In some embodiments of the methods provided herein, the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, wherein optionally the at least one attribute is an organ type, comprises selections of biomarkers according to **Table 119**; wherein optionally: i. a pre-determined biosignature indicative of adrenal gland consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from INHA, CDH1, SYP, MIB1, CALB2, KRT8, PSAP, KRT19, NCAM2, NKX3-1, ARG1,

SERPINA1, CD34, TPM3, S100A7, ACVRL1, PMEL, CR1, ERG, and PECAM1; ii. a pre-determined biosignature indicative of bladder consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, KRT20, UPK2, CPS1, SALL4, SERPINA1, DES, CALB2, MUC1, S100A2, MSLN, MITF, PAX8, S100A10, CNN1, UPK3A, CD3G, NAPSA, CD2, and MME; iii. a pre-determined biosignature indicative of brain consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT8, ANO1, S100B, S100A14, SOX2, PDPN, CEACAM1, S100A2, NCAM1, MSH2, KRT18, NKX2-2, WT1, S100A1, GPC3, TLE1, CD5, S100Z, S100A16, and PGR; iv. a pre-determined biosignature indicative of breast consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, ANKRD30A, KRT15, KRT7, S100A2, S100A1, MUC4, HNF1B, KRT18, SOX2, PIP, PAX8, MDM2, KRT16, MUC5AC, S100A6, TP63, TFF1, KRT5, and SERPINA1; v. a pre-determined biosignature indicative of colon consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, KRT7, MUC2, KRT20, MUC1, CEACAM5, CDH17, TFF3, KRT18, KRT6B, VIL1, SATB2, S100A6, SOX2, S100A14, HAVCR1, FUT4, ERG, HNF1B, and PTPRC; vi. a pre-determined biosignature indicative of eye consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PMEL, MLANA, MITF, BCL2, S100A13, S100A2, S100A10, S100A1, MIB1, SOX2, ENO2, S100A16, VIM, VHL, PDPN, WT1, S100B, KRT7, KRT10, and PSAP; vii. a pre-determined biosignature indicative of female genital tract and peritoneum (FGTP) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, ESR1, WT1, PGR, CDKN2A, FOXL2, KRT5, TPM4, SMARCB1, DES, TMPRSS2, CDK4, GATA3, AR, S100A13, MSH2, ANO1, CALB2, MS4A1, and CCND1; viii. a pre-determined biosignature indicative of gastroesophageal consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, ANO1, FUT4, SERPINB5, SPN, NCAM2, VIL1, CD34, ENO2, TFF3, AR, S100A13, TPM1, CEACAM6, SOX2, PAX8, MUC5AC, CDH1, S100A11, and ISL1; ix. a pre-determined biosignature indicative of head, face or neck, NOS consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT5, DSC3, TP63, HNF1B, MUC5AC, PAX5, KRT15, PGR, S100A6, TMPRSS2, MME, S100B, ENO2, CEACAM8, SALL4, ANO1, GATA3, LIN28B, CD99L2, and UPK3A; x. a pre-determined biosignature indicative of kidney consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, CDH1, HNF1B, S100A14, HAVCR1, CDKN2A, S100P, KL, KRT7, S100A13, VHL, PAX2, POU5F1, MUC1, AMACR, ENO2, MDM2, WT1, SYP, and AR; xi. a pre-determined biosignature indicative of liver, gallbladder, ducts consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, VIL1, HNF1B, ANO1, ESR1,

SOX2, MUC4, S100A2, ENO2, CNN1, POU5F1, KRT5, S100A9, UPK3B, PSAP, KRT7, KL, TMPRSS2, SATB2, and S100A14; xii. a pre-determined biosignature indicative of lung consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NAPSA, SOX2, SFTPA1, VHL, S100A1, S100A10, AR, TMPRSS2, CD99L2, CEACAM7, CEACAM6, KRT6A, KRT7, NCAM2, TP63, CEACAM1, MUC4, KRT20, CNN1, and ISL1; xiii. a pre-determined biosignature indicative of pancreas consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, ANO1, SERPINA1, GATA3, ISL1, MUC5AC, SMAD4, FUT4, CD5, SMN1, NKX2-2, TFF1, AMACR, SOX2, HNF1B, S100Z, MSLN, DES, S100A4, and CALB2; xiv. a pre-determined biosignature indicative of prostate consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KLK3, KRT7, NKX3-1, AMACR, CPS1, S100A5, UPK3A, KL, MUC1, CGB3, MUC2, TMPRSS2, MSLN, PMEL, S100A10, SERPINA1, KRT20, SFTPA1, BCL6, and TFF1; xv. a pre-determined biosignature indicative of skin consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT8, PMEL, KRT7, KRT19, GATA3, MDM2, AMACR, TPM1, TLE1, CEACAM19, CEACAM16, MLANA, TMPRSS2, AR, TFF3, BCL6, CR1, NCAM1, and MS4A1; xvi. a pre-determined biosignature indicative of small intestine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MUC2, CDH17, FLI1, KRT20, CDX2, CD5, KRT7, MPO, CNN1, DSC3, DES, ANO1, S100A1, CALD1, TFF1, SPN, MITF, TMPRSS2, CALB2, and CEACAM16; and/or xvii. a pre-determined biosignature indicative of thyroid consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, TG, CPS1, SERPINB5, INHA, ARG1, CNN1, CEACAM5, TPSAB1, CALB2, HNF1B, VIM, CDK4, S100P, S100A2, LIN28B, TFF3, CGA, TLE1, and TPM3.

In some embodiments of the methods provided herein, the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, wherein optionally the at least one attribute is a histology, comprises selections of biomarkers according to **Table 120**; wherein optionally: i. a pre-determined biosignature indicative of adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TMPRSS2, HNF1B, KRT5, MUC1, CEACAM5, MUC5AC, CDH17, TP63, ALPP, GATA3, CEACAM1, TFF3, S100A1, KRT8, PDX1, KRT17, CDH1, KLK3, CPS1, and S100A2; ii. a pre-determined biosignature indicative of adenoid cystic carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT14, KIT, TPM3, CGA, SMAD4, CTNNB1, DSC3, S100A6, TP63, TPM1, CALD1, MIB1, CD2, CDH1, ANO1, ENO2, CD3G, TPM2, CEACAM1, and BCL2; iii. a pre-determined biosignature indicative of adenosquamous carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, SFTPA1, OSCAR,

KRT19, KRT15, NAPSA, GPC3, MS4A1, S100A12, ERG, CEACAM6, VHL, SOX2, SERPINA1, KRT6A, CDKN2A, CD3G, PIP, NCAM2, and CEACAM7; iv. a pre-determined biosignature indicative of adrenal cortical carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MIB1, INHA, CDH1, SYP, CALB2, NKX3-1, KRT19, ERBB2, MUC1, ARG1, VIM, CD34, CALD1, S100A9, MSLN, S100A10, CD5, PMEL, SDC1, and TP63; v. a pre-determined biosignature indicative of astrocytoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, SOX2, NCAM1, MUC1, S100A4, KRT17, KRT8, S100A1, TPM4, CNN1, TPM2, OSCAR, AR, SDC1, SALL4, SMN1, SFTPA1, KIT, CA9, and S100A9; vi. a pre-determined biosignature indicative of carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, MITF, MUC5AC, PDPN, VIL1, CEACAM5, CDH1, CDH17, IL12B, S100P, KRT20, KRT7, SPN, TMPRSS2, ENO2, NKX2-2, PMEL, IMP3, BCL6, and S100A8; vii. a pre-determined biosignature indicative of carcinosarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT6B, GPC3, MSLN, MUC1, S100A6, S100A2, MME, CDKN2A, CDH1, FOXL2, KRT7, CALB2, SFTPA1, ERG, PGR, KRT17, NAPSA, CALD1, LIN28B, and KIT; viii. a pre-determined biosignature indicative of cholangiocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, HNF1B, VIL1, TFF1, ENO2, NKX2-2, FUT4, MUC4, MLH1, TMPRSS2, WT1, KL, KRT7, ESR1, MDM2, SFTPA1, SMN1, KRT18, UPK3B, and COQ2; ix. a pre-determined biosignature indicative of clear cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from POU5F1, HAVCR1, CEACAM6, HNF1B, PAX8, NAPSA, CD34, MYOG, FOXL2, MITF, S100P, S100A9, S100A14, S100Z, WT1, CDH1, TTF1, SYP, MLH1, and KRT16; x. a pre-determined biosignature indicative of ductal carcinoma in situ (DCIS) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, HNF1B, DES, MME, ANKRD30A, SATB2, SOX2, NCAM2, PAX8, CEACAM4, PIP, MUC4, NKX3-1, SERPINA1, KRT20, KIT, NCAM1, KRT14, S100A2, and CDKN2A; xi. a pre-determined biosignature indicative of glioblastoma (GBM) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT18, PDPN, NKX2-2, SOX2, NCAM1, KRT8, ERBB2, KRT15, KRT19, GATA3, CDKN2A, BCL6, S100A14, KRT10, UPK3A, SF1, CA9, CCND1, and KRT5; xii. a pre-determined biosignature indicative of GIST consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ANO1, SDC1, MUC1, KRT19, KRT8, ACVRL1, KIT, ERBB2, CDH1, CEACAM19, FUT4, TFF3, S100A16, S100A13, ISL1, S100A9, TPSAB1, KRT18, IMP3, and KRT3; xiii. a pre-determined biosignature indicative of glioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,

13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT8, S100B, SYP, NCAM2, CD3G, SDC1, SOX2, CEACAM1, POU5F1, MIB1, SATB2, MDM2, NCAM1, KRT7, CGB3, CPS1, PDPN, CALCA, ERBB2, and TNFRSF8; xiv. a pre-determined biosignature indicative of granulosa cell tumor consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FOXL2, SDC1, MSH6, KRT18, KRT8, MME, FLI1, S100A9, CALCA, S100B, CCND1, CEACAM21, TLE1, SERPINA1, S100A11, SFTPA1, SYP, NCAM2, CD3G, and SOX2; xv. a pre-determined biosignature indicative of infiltrating lobular carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDH1, GATA3, S100A1, TFF3, CA9, MUC1, NKX3-1, ANKRD30A, SOX2, S100A5, MUC4, KRT7, OSCAR, MME, SERPINA1, CDK4, AR, CEACAM3, BCL6, and KRT5; xvi. a pre-determined biosignature indicative of leiomyosarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT8, KRT18, CNN1, TPM4, FOXL2, TPM2, TPM1, CD79A, CALB2, SATB2, S100A5, DES, S100A14, KRT2, ERBB2, PDPN, ENO2, CD2, and CALD1; xvii. a pre-determined biosignature indicative of liposarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT18, MDM2, CDK4, CDH1, KRT19, KRT7, PDPN, CD34, TPM4, CR1, ACVRL1, MME, KRT8, AMACR, CEACAM5, S100B, OSCAR, LIN28A, S100A12, and SDC1; xviii. a pre-determined biosignature indicative of melanoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, PMEL, KRT19, KRT8, MUC1, S100A14, MLANA, S100A13, TPM1, MITF, VIM, CEACAM19, POU5F1, SATB2, CPS1, CDKN2A, KRT10, AR, ACVRL1, and LIN28A; xix. a pre-determined biosignature indicative of meningioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SDC1, KRT8, S100A14, ANO1, CEACAM1, VIM, KRT10, PGR, MSH2, CD5, S100A2, CDH1, TP63, SMARCB1, KRT16, S100A10, S100A4, DSC3, CCND1, and GATA3; xx. a pre-determined biosignature indicative of Merkel cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, ERBB2, MME, MYOG, CPS1, KRT7, SALL4, S100A12, S100A14, S100PBP, CR1, SMAD4, CEACAM5, MUC4, CA9, KRT10, SYP, CCND1, MSLN, and MLANA; xxi. a pre-determined biosignature indicative of mesothelioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from UPK3B, CALB2, PDPN, SMARCB1, MSLN, KRT5, CEACAM3, WT1, INHA, CEACAM1, CA9, TLE1, SATB2, CDH1, MUC2, CDKN2A, CEACAM18, MSH2, DSC3, and PTPRC; xxii. a pre-determined biosignature indicative of neuroendocrine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, NCAM1, S100A11, ENO2, S100A1, SYP, MUC1, TFF3, S100Z, PAX8, ERBB2, ESR1, S100A10, CEACAM5, SDC1, MUC4, MPO, S100A4, S100A7, and TP63; xxiii. a pre-determined biosignature

indicative of non-small cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ESR1, TMPRSS2, AR, S100A1, SFTPA1, MSLN, SOX2, ENO2, TP63, SMAD4, PTPRC, ISL1, CEACAM7, CEACAM20, S100Z, INHA, NCAM1, MUC2, TFF3, and PAX8; xxiv. a pre-determined biosignature indicative of oligodendroglioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, KRT18, CD2, S100A11, SYP, CDH1, S100A4, S100A14, CEACAM1, S100PBP, SDC1, SALL4, UPK2, COQ2, TPM2, CD99L2, TTF1, CD79A, INHA, and VIM; xxv. a pre-determined biosignature indicative of sarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, KRT19, S100A14, NKX2-2, KRT2, KRT7, SATB2, MYOG, CALD1, CEACAM19, CA9, KRT15, CDKN2A, S100P, WT1, TMPRSS2, S100A7, SERPINB5, DSC3, and ENO2; xxvi. a pre-determined biosignature indicative of sarcomatoid carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MME, VIM, S100A14, CD99L2, S100A11, NKX3-1, SATB2, CPS1, MSLN, SFTPA1, POU5F1, CDH1, OSCAR, S100A5, IMP3, CEACAM1, PMS2, NCAM2, KRT15, and S100A12; xxvii. a pre-determined biosignature indicative of serous consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from WT1, PAX8, KRT7, CDKN2A, MSLN, ACVRL1, SATB2, CDK4, DSC3, AR, S100A16, ANO1, S100A5, SDC1, IMP3, SERPINA1, KRT4, ESR1, FOXL2, and KRT15; xxviii. a pre-determined biosignature indicative of small cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, ISL1, PAX5, KIT, MUC4, S100A10, MUC1, CTNNB1, MITF, NKX2-2, S100A11, SMN1, MSLN, S100A6, BCL2, SYP, KL, CGB3, TPSAB1, TFF3; and/or xxix. a pre-determined biosignature indicative of squamous consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, KRT5, KRT17, SOX2, AR, CD3G, KRT6A, S100A1, DSC3, SERPINB5, HNF1B, SDC1, S100A6, TPSAB1, KRT20, HAVCR1, TTF1, MSH2, PMS2, and CNN1. The system and methods provided herein envision any combination of predetermined biosignatures above. See, e.g., **FIGs. 4A-C** and related text.

If making selections of biomarkers from within the pre-determined biosignatures provided herein, one may choose biomarkers that provide the most informative predictions. For example, one may choose the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features, e.g., 3 or 5 or 10 or 20 features, or at least 3 or 5 or 10 or 20 features, with the highest Importance value for each pre-determined biosignature listed in **Tables 118-120**.

In some embodiments of the methods provided herein, performing the at least one assay to assess the one or more biomarkers in step (b), including without limitation those described above with respect to **Tables 118-120**, comprises assessing the markers in the at least one pre-determined biosignature using DNA analysis and/or expression analysis, wherein: i. the DNA analysis consists of

or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof; ii. the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof; and/or iii. the expression analysis consists of or comprises analysis of RNA, where optionally: i. the RNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, amount, level, expression level, presence, or any combination thereof; and/or ii. the RNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole transcriptome sequencing, or any combination thereof; iv. the expression analysis consists of or comprises analysis of protein, where optionally: i. the protein analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, fusion, amplification, amount, level, expression level, presence, or any combination thereof; and/or ii. the protein analysis is performed using immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof; and/or v. any combination thereof. In some embodiments, performing the assay to assess the one or more biomarkers in step (b) comprises assessing the markers in the at least one pre-determined biosignature using: a combination of the DNA analysis and the RNA analysis; a combination of the DNA analysis and the protein analysis; a combination of the RNA analysis and the protein analysis; or a combination of the DNA analysis, the RNA analysis, and the protein analysis. In some embodiments, performing the assay to assess the one or more biomarkers in step (b) comprises RNA analysis of messenger RNA transcripts.

In some embodiments of the methods provided herein, the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, optionally a cancer type or primary tumor origin, comprises selections of biomarkers according to at least one of **FIGs. 6I-AC**; wherein optionally: i. a pre-determined biosignature indicative of breast adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, CDH1, PAX8, KRAS, ELK4, CCND1, MECOM, PBX1, CREBBP, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, NY-BR-1, KRT15, CK7, S100A2, RCCMa, MUC4, CK18, HNF1B and S100A1; ii. a pre-determined biosignature indicative of central nervous system cancer comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from IDH1, SOX2, OLIG2, MYC, CREB3L2, SPECC1, EGFR, FGFR2, SETBP1, and ZNF217, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from S100B, CK18,

CK8, SOX2, DOG1, CD56, PDPN, NKX2-2, CK19, and S100A14; iii. a pre-determined biosignature indicative of cervical adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, MECOM, RPN1, U2AF1, GNAS, RAC1, KRAS, FL11, EXT1, and CDK6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from ER, p16, CYCLIND1, LIN28A, PR, SMARCB1, CEACAM4, S100B, CD15, and PSAP; iv. a pre-determined biosignature indicative of cholangiocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, ARID1A, MAF, KRAS, CACNA1D, SPEN, SETBP1, CDK12, LHFPL6, and MDS2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from HNF1B, VILLIN, ANTITRYPSIN, ER, DOG1, SOX2, MUC4, S100A2, KRT5, and CK7; v. a pre-determined biosignature indicative of colon adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from APC, CDX2, KRAS, SETBP1, FLT3, LHFPL6, CDKN2A, FLT1, ASXL1, and CDKN2B, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CDX2, CK7, MUC2, CK20, MUC1, SATB2, VILLIN, CEACAM5, CDK17, and S100A6; vi. a pre-determined biosignature indicative of gastroesophageal adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CDX2, ERG, TP53, KRAS, U2AF1, ZNF217, CREB3L2, IRF4, TCF7L2, and LHFPL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CD15, CDX2, MASPIN, MUC5AC, AR, TFF1, NCAM2, TFF3, ISL1, and DOG1; vii. a pre-determined biosignature indicative of gastrointestinal stromal tumor (GIST) comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from c-KIT (KIT), TP53, MAX, PDGFRA, TSHR, MSI2, SPEN, JAK1, SETBP1, and CDH11, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from DOG1, CD138, CK19, MUC1, CK8, ACVRL1, KIT, E-CADHERIN, S100A2, and CK7; viii. a pre-determined biosignature indicative of hepatocellular carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from HLF, CACNA1D, HMGN2P46, KRAS, FANCF, PRCC, ERG, FLT1, FGFR1, and ACSL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from ANTITRYPSIN, CEACAM16, CK19, AFP, MUC4, CEACAM5, MSH2, BCL6, DSC3, and KRT15; ix. a pre-determined biosignature indicative of lung adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from NKX-2, KRAS, TP53, TPM4, CDX2, TERT, FOXA1, SETBP1, CDKN2A, and LHFPL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from Napsin A, SOX2, CEACAM7, CK7, S100A10, CEACAM6, S100A1, RCCMa, AR and VHL; x. a pre-determined biosignature indicative of melanoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from IRF4, SOX10, TP53, BRAF, FGFR2, TRIM27, EP300, CDKN2A, LRP1B, and NRAS, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from S100B, CK8, HMB-45, CD19, MUC1, MLANA, S100A14, S100A13, MITF, and S100A1; xi. a pre-determined biosignature indicative of meningioma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CHEK2, TP53, MYCL,

THRAP3, MPL, EBF1, EWSR1, PMS2, FLI1, and NTRK2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CD138, CK8, DOG1, VIM, S100A14, S100A2, CEACAM1, MSH2, PR, and KRT10; xii. a pre-determined biosignature indicative of ovarian granulosa cell tumor comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXL2, TP53, EWSR1, CBFB, SPECC1, BCL3, MYH9, TSHR, GID4, and SOX2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXL2, CD138, MSH6, MUC1, CK8, PR, MME, ANTITRYPSIN, FLI1, and S100B; xiii. a pre-determined biosignature indicative of ovarian & fallopian tube adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, MECOM, KRAS, TPM4, RAC1, ASXL1, EP300, CDX2, RPN1, and WT1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from WT1, RCCMa, INHIBIN-alpha, TFE3, S100A13, FOLX2, TLE1, MSLN, POU5F1, and CEACAM3; xiv. a pre-determined biosignature indicative of pancreas adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from KRAS, CDKN2A, CDKN2B, FANCF, IRF4, TP53, ASXL1, SETBP1, APC, and FOXO1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from PDX1, GATA3, DOG1, ANTITRYPSIN, ISL1, MUC5AC, CD15, SMAD4, CD5, and CALB2; xv. a pre-determined biosignature indicative of prostate adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXA1, PTEN, KLK2, FOXO1, GATA2, FANCA, LHFPL6, KRAS, ETV6, and ERCC3, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CK7, PSA, NKX3-1, AMACR, S100A5, MUC1, MUC2, UPK3A, KL and HEPPAR-1; xvi. a pre-determined biosignature indicative of renal cell carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from VHL, TP53, EBF1, MAF, RAF1, CTNNA1, XPC, MUC1, KRAS, and BTG1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RCCMa, E-CADHERIN, p16, S100P, S100A14, HAVCR1, HNF1B, KL, CK7, and MUC1; xvii. a pre-determined biosignature indicative of squamous cell carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, SOX2, KLHL6, CDKN2A, LPP, CACNA1D, TFRC, KRAS, RPN1, and CDX2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from P63, SOX2, CK6, KRT17, S100A1, CD3G, SFTPA1, AR, KRT5, and CD138; xviii. a pre-determined biosignature indicative of thyroid cancer comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from BRAF, NKX2-1, TP53, MYC, KDSR, TRRAP, CDX2, KRAS, FHIT, and SETBP1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from THYROGLOBULIN, RCCMa, HEPPAR-1, S100A2, TPSAB1, CALB2, HNF1B, INHIBIN-alpha, ARG1, and CNN1; xix. a pre-determined biosignature indicative of urothelial carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, ASXL1, CDKN2B, TP53, CTNNA1, CDKN2A, KRAS, IL7R, CREBBP, and VHL, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, UPII, CK20, MUC1,

S100A2, HEPPAR-1, P63, CALB2, MITF, and S100P; xx. a pre-determined biosignature indicative of uterine endometrial adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from PTEN, PAX8, PIK3CA, CCNE1, TP53, MECOM, ESR1, CDX2, CDKN2A, and KRAS, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RCCMa, PR, ER, VHL, CALD1, LIN28B, Napsin A, KRT5, S100A6, and DES; and/or xxi. a pre-determined biosignature indicative of uterine sarcoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RB1, SPECC1, FANCC, TP53, CACNA1D, JAK1, ETV1, PRRX1, PTCH1, and HOXD13, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CK19, CK18, CD56, DES, FOXL2, CD79A, S100A14, ER, MSLN, and MITF. In some embodiments, the DNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof. In some embodiments, the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof. In some embodiments, the expression analysis consists of or comprises analysis of RNA. In some embodiments, the RNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, amount, level, expression level, presence, or any combination thereof. In some embodiments, the RNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole transcriptome sequencing, or any combination thereof. In some embodiments, the expression analysis consists of or comprises analysis of protein. In some embodiments, the protein analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, fusion, amplification, amount, level, expression level, presence, or any combination thereof. In some embodiments, the protein analysis is performed using immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof. Any useful combination of such analyses is contemplated by the invention.

In the methods provided herein, the at least one pre-determined biosignature may comprise or may further comprise, as the case may be, selections of biomarkers according to any one of **Tables 2-116** assessed using DNA analysis. In some embodiments, the DNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA) or any combination thereof. In some embodiments, the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization,

microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof. In some embodiments, the at least one pre-determined biosignature comprising selections of biomarkers according to any one of **Tables 2-116** comprises:

i. a pre-determined biosignature indicative of adrenal cortical carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 2**; ii. a pre-determined biosignature indicative of anus squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 3**; iii. a pre-determined biosignature indicative of appendix adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 4**; iv. a pre-determined biosignature indicative of appendix mucinous adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 5**; v. a pre-determined biosignature indicative of bile duct NOS cholangiocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 6**; vi. a pre-determined biosignature indicative of brain astrocytoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 7**; vii. a pre-determined biosignature indicative of brain astrocytoma anaplastic origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 8**; viii. a pre-determined biosignature indicative of breast adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 9**; ix. a pre-determined biosignature indicative of breast carcinoma NOS consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 10**; x. a pre-determined biosignature indicative of breast infiltrating duct adenocarcinoma origin consisting of, comprising, or comprising

at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 11**; xi. a pre-determined biosignature indicative of breast infiltrating lobular adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 12**; xii. a pre-determined biosignature indicative of breast metaplastic carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 13**; xiii. a pre-determined biosignature indicative of cervix adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 14**; xiv. a pre-determined biosignature indicative of cervix carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 15**; xv. a pre-determined biosignature indicative of cervix squamous carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 16**; xvi. a pre-determined biosignature indicative of colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 17**; xvii. a pre-determined biosignature indicative of colon carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 18**; xviii. a pre-determined biosignature indicative of colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 19**; xix. a pre-determined biosignature indicative of conjunctiva malignant melanoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 20**; xx. a pre-determined biosignature indicative of duodenum and ampulla adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41,

42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 21**; xxi. a pre-determined biosignature indicative of endometrial endometrioid adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 22**; xxii. a pre-determined biosignature indicative of endometrial adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 23**; xxiii. a pre-determined biosignature indicative of endometrial carcinosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 24**; xxiv. a pre-determined biosignature indicative of endometrial serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 25**; xxv. a pre-determined biosignature indicative of endometrium carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 26**; xxvi. a pre-determined biosignature indicative of endometrium carcinoma undifferentiated origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 27**; xxvii. a pre-determined biosignature indicative of endometrium clear cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 28**; xxviii. a pre-determined biosignature indicative of esophagus adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 29**; xxix. a pre-determined biosignature indicative of esophagus carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 30**; xxx. a pre-determined biosignature indicative of esophagus squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 31**; xxxi. a pre-determined biosignature indicative of

extrahepatic cholangio common bile gallbladder adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 32**; xxxii. a pre-determined biosignature indicative of fallopian tube adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 33**; xxxiii. a pre-determined biosignature indicative of fallopian tube carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 34**; xxxiv. a pre-determined biosignature indicative of fallopian tube carcinosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 35**; xxxv. a pre-determined biosignature indicative of fallopian tube serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 36**; xxxvi. a pre-determined biosignature indicative of gastric adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 37**; xxxvii. a pre-determined biosignature indicative of gastroesophageal junction adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 38**; xxxviii. a pre-determined biosignature indicative of glioblastoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 39**; xxxix. a pre-determined biosignature indicative of glioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 40**; xl. a pre-determined biosignature indicative of gliosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 41**; xli. a pre-determined biosignature indicative of head, face or neck NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,

21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 42**; xlii. a pre-determined biosignature indicative of intrahepatic bile duct cholangiocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 43**; xliii. a pre-determined biosignature indicative of kidney carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 44**; xliv. a pre-determined biosignature indicative of kidney clear cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 45**; xlv. a pre-determined biosignature indicative of kidney papillary renal cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 46**; xlvi. a pre-determined biosignature indicative of kidney renal cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 47**; xlvii. a pre-determined biosignature indicative of larynx NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 48**; xlviii. a pre-determined biosignature indicative of left colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 49**; xlix. a pre-determined biosignature indicative of left colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 50**; l. a pre-determined biosignature indicative of liver hepatocellular carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 51**; li. a pre-determined biosignature indicative of lung adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected

from **Table 52**; lii. a pre-determined biosignature indicative of lung adenosquamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 53**; liii. a pre-determined biosignature indicative of lung carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 54**; liv. a pre-determined biosignature indicative of lung mucinous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 55**; lv. a pre-determined biosignature indicative of lung neuroendocrine carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 56**; lvi. a pre-determined biosignature indicative of lung non-small cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 57**; lvii. a pre-determined biosignature indicative of lung sarcomatoid carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 58**; lviii. a pre-determined biosignature indicative of lung small cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 59**; lix. a pre-determined biosignature indicative of lung squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 60**; lx. a pre-determined biosignature indicative of meninges meningioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 61**; lxi. a pre-determined biosignature indicative of nasopharynx NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 62**; lxii. a pre-determined biosignature indicative of oligodendroglioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 63**; lxiii. a pre-determined biosignature indicative of oligodendroglioma aplastic origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 64**; lxiv. a pre-determined biosignature indicative of ovary adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 65**; lxv. a pre-determined biosignature indicative of ovary carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 66;** lxvi. a pre-determined biosignature indicative of ovary carcinosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 67**; lxvii. a pre-determined biosignature indicative of ovary clear cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 68**; lxviii. a pre-determined biosignature indicative of ovary endometrioid adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 69**; lxix. a pre-determined biosignature indicative of ovary granulosa cell tumor NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 70**; lxx. a pre-determined biosignature indicative of ovary high-grade serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 71**; lxxi. a pre-determined biosignature indicative of ovary low-grade serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 72**; lxxii. a pre-determined biosignature indicative of ovary mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 73**; lxxiii. a pre-

determined biosignature indicative of ovary serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 74**; lxxiv. a pre-determined biosignature indicative of pancreas adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 75**; lxxv. a pre-determined biosignature indicative of pancreas carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 76**; lxxvi. a pre-determined biosignature indicative of pancreas mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 77**; lxxvii. a pre-determined biosignature indicative of pancreas neuroendocrine carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 78**; lxxviii. a pre-determined biosignature indicative of parotid gland carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 79**; lxxix. a pre-determined biosignature indicative of peritoneum adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 80**; lxxx. a pre-determined biosignature indicative of peritoneum carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 81**; lxxxi. a pre-determined biosignature indicative of peritoneum serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 82**; lxxxii. a pre-determined biosignature indicative of pleural mesothelioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 83**; lxxxiii. a pre-determined biosignature indicative of prostate adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6,

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 84**; lxxxiv. a pre-determined biosignature indicative of rectosigmoid adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 85**; lxxxv. a pre-determined biosignature indicative of rectum adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 86**; lxxxvi. a pre-determined biosignature indicative of rectum mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 87**; lxxxvii. a pre-determined biosignature indicative of retroperitoneum dedifferentiated liposarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 88**; lxxxviii. a pre-determined biosignature indicative of retroperitoneum leiomyosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 89**; lxxxix. a pre-determined biosignature indicative of right colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 90**; xc. a pre-determined biosignature indicative of right colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 91**; xci. a pre-determined biosignature indicative of salivary gland adenoidcystic carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 92**; xcii. a pre-determined biosignature indicative of skin Merkel cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 93**; xciii. a pre-determined biosignature indicative of skin nodular melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 94**; xciv. a pre-determined biosignature indicative of skin squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 95**; xcv. a pre-determined biosignature indicative of skin melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 96**; xcvi. a pre-determined biosignature indicative of small intestine gastrointestinal stromal tumor (GIST) NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 97**; xcvii. a pre-determined biosignature indicative of small intestine adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 98**; xcviii. a pre-determined biosignature indicative of stomach gastrointestinal stromal tumor (GIST) NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 99**; xcix. a pre-determined biosignature indicative of stomach signet ring cell adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 100**; c. a pre-determined biosignature indicative of thyroid carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 101**; ci. a pre-determined biosignature indicative of thyroid carcinoma anaplastic NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 102**; cii. a pre-determined biosignature indicative of papillary carcinoma of thyroid origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 103**; ciii. a pre-determined biosignature indicative of tonsil oropharynx tongue squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected

from **Table 104**; civ. a pre-determined biosignature indicative of transverse colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 105**; cv. a pre-determined biosignature indicative of urothelial bladder adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 106**; cvi. a pre-determined biosignature indicative of urothelial bladder carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 107**; cvii. a pre-determined biosignature indicative of urothelial bladder squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 108**; cviii. a pre-determined biosignature indicative of urothelial carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 109**; cix. a pre-determined biosignature indicative of uterine endometrial stromal sarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 110**; cx. a pre-determined biosignature indicative of uterus leiomyosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 111**; cxi. a pre-determined biosignature indicative of uterus sarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 112**; cxii. a pre-determined biosignature indicative of uveal melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 113**; cxiii. a pre-determined biosignature indicative of vaginal squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 114**; cxiv. a pre-determined biosignature indicative of vulvar squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7,

8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 115**; and/or cxv. a pre-determined biosignature indicative of skin trunk melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 116**. In some embodiments, the selections of biomarkers according to any one of **Tables 2-116** comprises the top 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the feature biomarkers with the highest Importance value in the corresponding table/s. In some embodiments, the selections of biomarkers according to any one of **Tables 2-116** comprises the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 feature biomarkers with the highest Importance value in the corresponding table/s. In some embodiments, the selections of biomarkers according to any one of **Tables 2-116** comprises at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 40%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 feature biomarkers with the highest Importance value in the corresponding table/s. In some embodiments, the selections of biomarkers according to any one of **Tables 2-116** comprises at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, or 100 feature biomarkers with the highest Importance value in the corresponding table.

If making selections of biomarkers from within the pre-determined biosignatures provided herein, one may choose biomarkers that provide the most informative predictions. For example, one may choose the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 features, e.g., 3 or 5 or 10 or 20 or 25 features, or at least 3 or 5 or 10 or 20 or 25 features, with the highest Importance value for each pre-determined biosignature listed in **Tables 2-116**.

In some embodiments of the methods provided herein, step (b) comprises determining a gene copy number for at least one member of the biosignature, and step (d) comprises processing the gene copy number. In some embodiments, step (b) comprises determining a sequence for at least one member of the biosignature, and step (d) comprises processing the sequence. In some embodiments, step (b) comprises determining a sequence for a plurality of members of the biosignature, and step (d)

comprises comparing the sequence to a reference sequence (e.g., wild type) to identify microsatellite repeats, and identifying members of the biosignature that have microsatellite instability (MSI. In some embodiments, step (b) comprises determining a sequence for a plurality of members of the biosignature, and step (d) comprises comparing the sequence to a reference sequence (e.g., wild type) to identify a tumor mutational burden (TMB. In some embodiments, step (b) comprises determining an mRNA transcript level for at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 genes in any one of Tables 117-120, and/or INSM1, and step (d) comprises processing the transcript levels. In some embodiments, a gene copy number, CNV or CNA of a gene in the biosignature is determined by measuring the copy number of at least one proximate region to the gene, wherein optionally the proximate region comprises at least one location in the same sub-band, band, or arm of the chromosome wherein the gene is located.

In some embodiments of the methods provided herein, the one or more biomarkers in the biosignature are assessed as described in their corresponding table, including without limitation **Tables 2-116** or **Tables 117-120**.

In some embodiments of the methods provided herein, the model comprises a plurality of intermediate models, wherein the plurality of intermediate models comprises at least one pairwise comparison module and/or at least one multi-class classification model. In some embodiments, the model calculates a statistical measure that the biosignature corresponds to at least one of the at least one pre-determined biosignatures. In some embodiments, the processing in step (d) comprises a pairwise comparison between candidate pre-determined biosignatures, and a probability is calculated that the biosignature corresponds to either one of the pairs of the at least one pre-determined biosignatures; and/or using at least one multi-class classification model to assess the biosignature. In some embodiments, the pairwise comparison between the two candidate primary tumor origins and/or the multi-class classification model is determined using a machine learning classification algorithm, wherein optionally the machine learning classification algorithm comprises a boosted tree. In some embodiments, the pairwise comparison between the two candidate primary tumor origins is applied to at least one pre-determined biosignature supplied herein, e.g., with respect to **Tables 2-116**; and/or the multi-class classification model is applied to at least one pre-determined biosignature supplied herein, e.g., with respect to **Tables 118-120**.

In some embodiments, the methods supplied herein further comprise determining intermediate model predictions, wherein the intermediate model predictions comprise: a cancer type determined by the joint pairwise comparisons between at least one pair of pre-determined biosignatures supplied herein, e.g., with respect to **Tables 2-116**; a cancer/disease type determined by an intermediate multi-class model applied to at least one pre-determined biosignature supplied herein, e.g., with respect to **Table 118**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28

of the pre-determined biosignatures in **Table 118**; an organ group type determined by an intermediate multi-class model applied to at least one pre-determined biosignature supplied herein, e.g., with respect to **Table 119**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27 of the pre-determined biosignatures in **Table 119**; and/or a histology determined by an intermediate multi-class model applied to at least one pre-determined biosignature supplied herein, e.g., with respect to **Table 120**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or 29 of the pre-determined biosignatures in **Table 120**. In some embodiments, the processing in step (d) comprises inputting the outputs of each of the utilized intermediate multi-class models into a final predictor model that provides the prediction in step (e), wherein optionally the final predictor model comprises a machine learning algorithm, wherein optionally the machine learning algorithm comprises a boosted tree.

As described herein, the predicted at least one attribute of the cancer provided by the systems and methods herein can be provided at a desired level of granularity. In some embodiments, the predicted at least one attribute of the cancer comprises at least one of adrenal cortical carcinoma; anus squamous carcinoma; appendix adenocarcinoma, NOS; appendix mucinous adenocarcinoma; bile duct, NOS, cholangiocarcinoma; brain astrocytoma, anaplastic; brain astrocytoma, NOS; breast adenocarcinoma, NOS; breast carcinoma, NOS; breast infiltrating duct adenocarcinoma; breast infiltrating lobular carcinoma, NOS; breast metaplastic carcinoma, NOS; cervix adenocarcinoma, NOS; cervix carcinoma, NOS; cervix squamous carcinoma; colon adenocarcinoma, NOS; colon carcinoma, NOS; colon mucinous adenocarcinoma; conjunctiva malignant melanoma, NOS; duodenum and ampulla adenocarcinoma, NOS; endometrial adenocarcinoma, NOS; endometrial carcinosarcoma; endometrial endometrioid adenocarcinoma; endometrial serous carcinoma; endometrium carcinoma, NOS; endometrium carcinoma, undifferentiated; endometrium clear cell carcinoma; esophagus adenocarcinoma, NOS; esophagus carcinoma, NOS; esophagus squamous carcinoma; extrahepatic cholangio, common bile, gallbladder adenocarcinoma, NOS; fallopian tube adenocarcinoma, NOS; fallopian tube carcinoma, NOS; fallopian tube carcinosarcoma, NOS; fallopian tube serous carcinoma; gastric adenocarcinoma; gastroesophageal junction adenocarcinoma, NOS; glioblastoma; glioma, NOS; gliosarcoma; head, face or neck, NOS squamous carcinoma; intrahepatic bile duct cholangiocarcinoma; kidney carcinoma, NOS; kidney clear cell carcinoma; kidney papillary renal cell carcinoma; kidney renal cell carcinoma, NOS; larynx, NOS squamous carcinoma; left colon adenocarcinoma, NOS; left colon mucinous adenocarcinoma; liver hepatocellular carcinoma, NOS; lung adenocarcinoma, NOS; lung adenosquamous carcinoma; lung carcinoma, NOS; lung mucinous adenocarcinoma; lung neuroendocrine carcinoma, NOS; lung non-small cell carcinoma; lung sarcomatoid carcinoma; lung small cell carcinoma, NOS; lung squamous carcinoma; meninges meningioma, NOS; nasopharynx, NOS squamous carcinoma; oligodendroglioma, anaplastic; oligodendroglioma, NOS; ovary adenocarcinoma, NOS; ovary

carcinoma, NOS; ovary carcinosarcoma; ovary clear cell carcinoma; ovary endometrioid adenocarcinoma; ovary granulosa cell tumor, NOS; ovary high-grade serous carcinoma; ovary low-grade serous carcinoma; ovary mucinous adenocarcinoma; ovary serous carcinoma; pancreas adenocarcinoma, NOS; pancreas carcinoma, NOS; pancreas mucinous adenocarcinoma; pancreas neuroendocrine carcinoma, NOS; parotid gland carcinoma, NOS; peritoneum adenocarcinoma, NOS; peritoneum carcinoma, NOS; peritoneum serous carcinoma; pleural mesothelioma, NOS; prostate adenocarcinoma, NOS; rectosigmoid adenocarcinoma, NOS; rectum adenocarcinoma, NOS; rectum mucinous adenocarcinoma; retroperitoneum dedifferentiated liposarcoma; retroperitoneum leiomyosarcoma, NOS; right colon adenocarcinoma, NOS; right colon mucinous adenocarcinoma; salivary gland adenoid cystic carcinoma; skin melanoma; skin melanoma; skin merkel cell carcinoma; skin nodular melanoma; skin squamous carcinoma; skin trunk melanoma; small intestine adenocarcinoma; small intestine gastrointestinal stromal tumor, NOS; stomach gastrointestinal stromal tumor, NOS; stomach signet ring cell adenocarcinoma; thyroid carcinoma, anaplastic, NOS; thyroid carcinoma, NOS; thyroid papillary carcinoma of thyroid; tonsil, oropharynx, tongue squamous carcinoma; transverse colon adenocarcinoma, NOS; urothelial bladder adenocarcinoma, NOS; urothelial bladder carcinoma, NOS; urothelial bladder squamous carcinoma; urothelial carcinoma, NOS; uterine endometrial stromal sarcoma, NOS; uterus leiomyosarcoma, NOS; uterus sarcoma, NOS; uveal melanoma; vaginal squamous carcinoma; vulvar squamous carcinoma; and any combination thereof. In some embodiments, the predicted at least one attribute of the cancer comprises at least one of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma. In some embodiments, the predicted at least one attribute of the cancer comprises at least one of bladder; skin; lung; head, face or neck (NOS); esophagus; female genital tract (FGT); brain; colon; prostate; liver, gall bladder, ducts; breast; eye; stomach; kidney; and pancreas. In some embodiments, the sample comprises a cancer of unknown primary (CUP).

In an aspect, provided herein is a method of predicting at least one attribute of a cancer, the method comprising: (a) obtaining a biological sample from a subject having a cancer, wherein the biological sample can be a biological sample such as described above; (b) performing at least one assay to assess one or more biomarkers in the biological sample to obtain a biosignature for the sample, wherein the at least one assay can be as described above; (c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises at least one intermediate model, wherein the at least one intermediate model comprises: (1) an first intermediate model trained to process DNA data using the predetermined biosignatures

supplied herein with respect to **Tables 2-116**; (2) a second intermediate model trained to process RNA data using the predetermined biosignatures supplied herein with respect to **Table 118**; (3) a third intermediate model trained to process RNA data using the predetermined biosignatures supplied herein with respect to **Table 119**; and/or (4) a fourth intermediate model trained to process RNA data using the predetermined biosignatures supplied herein with respect to **Table 120**; (d) processing, by one or more computers, the provided biosignature through each of the plurality of intermediate models in part (c), providing the output of each of the plurality of intermediate models into a final predictor model, and processing by one or more computers, the output of each of the plurality of intermediate models through the final predictor model; and (e) outputting from the final predictor model a prediction of the at least one attribute of the cancer. In some embodiments, the predicted at least one attribute of the cancer is a tissue-of-origin selected from the group consisting of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, uterine sarcoma, and any combination thereof. In some embodiments, step (b) comprises performing DNA analysis by sequencing genomic DNA from the biological sample, wherein the DNA analysis is performed for the genes in **Tables 2-116**. In some embodiments, step (b) comprises performing RNA analysis by sequencing messenger RNA transcripts from the biological sample, wherein the RNA analysis is performed for the genes in **Table 117** or **Tables 118-120**. In some embodiments, the at least one of the at least one intermediate model and final predictor model comprises a machine learning module, wherein optionally the machine learning module comprises one or more of a random forest, support vector machine, logistic regression, K-nearest neighbor, artificial neural network, naïve Bayes, quadratic discriminant analysis, and Gaussian processes models, wherein optionally the machine learning module comprises an XGBoost decision-tree-based ensemble machine learning algorithm.

The prediction of the at least one attribute of the cancer made using the systems and methods provided herein may be used in various settings. See, e.g., **Example 3** herein. In some embodiments, the prediction is used to confirm a diagnosis. In some embodiments, the prediction is used to change a diagnosis. In some embodiments, the prediction is used to perform a quality check. In some embodiments, the prediction is used to indicate additional molecular testing to be performed.

In some embodiments of the methods of the invention, the predicted at least one attribute comprises an ordered list, wherein optionally the list is ordered using a statistical measure. For example, the list may be ordered by confidence in the prediction. In some embodiments, the methods provided herein further comprise determining whether the prediction of the at least one attribute meets

a threshold level, wherein optionally the threshold level is related to a probability of the prediction and/or a confidence in the prediction.

In some embodiments, the methods provided herein further comprise generating a molecular profile that identifies the presence, level, or state of the biomarkers in the biosignature, e.g., whether each biomarker has a copy number alteration and/or mutation; and/or a TMB level, MSI, LOH, or MMR status; and/or expression level, wherein the expression level comprises that of at least one transcript and/or protein level. See, e.g., **Example 1** for more details.

In some embodiments, the methods provided herein further comprise selecting at least one treatment for the patient based at least in part upon the classified at least one attribute of the cancer, wherein optionally the treatment comprises administration of immunotherapy, chemotherapy, or a combination thereof.

In an aspect, provided herein is a method comprising preparing a report, wherein the report comprises a summary or overview of the molecular profile generated herein, e.g., as described above, wherein the report identifies the classified at least one attribute of the cancer, wherein optionally the report further identifies the at least one treatment selected according to the methods provided herein, e.g., as described above. In some embodiments, the report is computer generated, is a printed report and/or a computer file, and/or is accessible via a web portal.

Further provided herein is a system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform operations described with reference to the methods described above. Relatedly, also provided herein is a non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform operations with reference to the methods described above.

In an aspect, provided herein is a system for identifying a lineage for a cancer, the system comprising: (a) at least one host server; (b) at least one user interface for accessing the at least one host server to access and input data; (c) at least one processor for processing the inputted data; (d) at least one memory coupled to the processor for storing the processed data and instructions for carrying out operations with reference to the methods described above; and (e) at least one display for displaying the classified primary origin of the cancer. In some embodiments, the system further comprise at least one memory coupled to the processor for storing the processed data and instructions for selecting treatment and/or generating molecular profiling reports as described herein. In some embodiments, the at least one display comprises a report comprising the classified at least one attribute of the cancer.

In an aspect, provided herein is a system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising: one or more processors and one or more memory units storing instructions that,

when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising: obtaining, by the system, a sample biological signature representing the sample that was obtained from the body, wherein the sample comprises cancer cells; providing, by the system, the sample biological signature as an input to a model, wherein: the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or the model is a multi-class model wherein the classes comprise different attributes; and receiving, by the system, an output generated by the model that represents data indicating a likely attribute of the sample obtained from the body based on the pairwise analysis. In another aspect, provided herein is a system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising: one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising: obtaining, by the system, a sample biological signature representing the sample that was obtained from the body; providing, by the system, the sample biological signature as an input to a model, wherein: the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or the model is a multi-class model wherein the classes comprise different attributes; and receiving, by the system, an output generated by the model that represents data indicating a probability that an attribute identified by the particular biological signature identifies a likely attribute of the sample. In still another aspect, provided herein is a system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising: one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising: obtaining, by the system, a sample biological signature representing a biological sample that was obtained from the cancer sample in a first portion of the body, wherein the sample biological signature includes data describing a plurality of features of the biological sample, wherein the plurality of features include data describing the first portion of the body; providing, by the system, the sample biological signature as an input to a model, wherein: the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or the model is a multi-class model wherein the classes comprise different attributes; and receiving, by the system, an output generated by the model that represents data indicating a likely attribute of the sample obtained from the body. In some

embodiments, the sample obtained from the body is a biological sample as described above. In some embodiment, the at least one attribute is a primary tumor origin, cancer/disease type, organ group, and/or histology as described above. In some embodiments, the sample biological signature includes data representing features obtained based on performance of an assay to assess one or more biomarkers in the cancer sample, wherein optionally the assay is according to at least one assay described above. In some embodiments, the operations further comprise: determining, based on the output generated by the model, a proposed cancer treatment. In some embodiments, each of the multiple different biological signatures comprise pre-identified biosignatures as described above, e.g., with respect to **Tables 2-116** or **Tabled 118-120**. In some embodiments, the operations further comprise: receiving, by the system, an output generated by the model that represents a likelihood that the sample obtained from the body in a first portion of the body originated from a cancer in a second portion of the body. In some embodiments, further comprising determining, by the system and based on the received output, whether the received output generated by the model satisfies one or more predetermined thresholds; and based on the determining, by the system, that the received output satisfies the one or more predetermined thresholds, determining, by the system, that the cancerous neoplasm in the first portion of the body originated from a cancer in a second portion of the body or that the cancerous neoplasm in the first portion of the body did not originate from a cancer in a second portion of the body. In some embodiments, the received output generated by the model includes a matrix data structure, wherein the matrix data structure includes a cell for each feature of the plurality of features evaluated by the pairwise model, wherein each of the cells includes data describing a probability that the corresponding feature indicates that the cancerous neoplasm in the first portion of the body was caused by cancer in the second portion of the first body.

In an aspect, provided herein is a system for identifying at least one attribute of a cancer, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising: one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising: receiving, by the system storing a model that is configured to perform analysis of a biological signature, a sample biological signature representing a biological sample that was obtained from a cancerous neoplasm in a first portion of a body, wherein the model includes a cancerous biological signature for each of multiple different types of cancerous biological samples, wherein the cancerous biological signatures include at least a first cancerous biological signature representing a molecular profile of a cancerous biological sample from the first portion of one or more other bodies; performing, by the system and using the model, analysis of the sample biological signature using the cancerous biological signatures; generating, by the system and based on the performed analysis, a likelihood that the cancerous neoplasm in the first portion of the body was caused by cancer in a

second portion of the body; providing, by the system, the generated likelihood to another device for display on the other device.

In an aspect, provided herein is a system for training an analysis model for identifying at least one attribute of a cancer sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising: one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising: generating, by the system, an analysis model, wherein generating the analysis model includes generating a plurality of model signatures, wherein each model signature is configured to differentiate between at least one attribute within each of the at least one attribute; obtaining, by the system, a set of training data items, wherein each training data item represents DNA or RNA sequencing results and includes data indicating (i) whether or not a variant was detected in the sequencing results and (ii) a number of copies of a gene or transcript in the sequencing results; and training, by the system, an analysis model using the obtained set of training data items. In some embodiments, the plurality of model signatures are generated using random forest models, wherein optionally the random forest models comprise gradient boosted forests.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1A is a block diagram of an example of a prior art system for training a machine learning model.

FIG. 1B is a block diagram of a system that generates training data structures for training a machine learning model to predict a sample origin.

FIG. 1C is a block diagram of a system for using a trained machine learning model to predict a sample origin of sample data from a subject.

FIG. 1D is a flowchart of a process for generating training data structures for training a machine learning model to predict sample origin.

FIG. 1E is a flowchart of a process for using a trained machine learning model to predict sample origin of sample data from a subject.

FIG. 1F is an example of a system for performing pairwise to predict a sample origin.

FIG. 1G is a block diagram of a system for predicting a sample origin using a voting unit to interpret output generated by multiple machine learning models that are each trained to perform pairwise analysis.

FIG. 1H is a block diagram of system components that can be used to implement systems of **FIGs. 1B, 1C, 1G, 1F, and 1G**.

FIG. 1I illustrates a block diagram of an exemplary embodiment of a system for determining individualized medical intervention for cancer that utilizes molecular profiling of a patient's biological specimen.

FIGs. 2A-C are flowcharts of exemplary embodiments of (**FIG. 2A**) a method for determining individualized medical intervention for cancer that utilizes molecular profiling of a patient's biological specimen, (**FIG. 2B**) a method for identifying signatures or molecular profiles that can be used to predict benefit from therapy, and (**FIG. 2C**) an alternate version of (**FIG. 2B**).

FIGs. 3A-B use of biosignatures to predict a primary tumor lineage from a cancer sample.

FIGs. 4A-B show schemes for classifying a tissue sample using RNA transcript analysis (**FIG. 4A**) or combined RNA and DNA analysis (**FIG. 4B**). **FIG. 4C** is flowchart of an example of a process 400C for training a dynamic voting engine.

FIGs. 5A-E illustrate performance of the MDC/GPS to classify cancers using analysis of genomic DNA.

FIGs. 6A-AL show further development of GPS using combined RNA and DNA analysis.

FIGs. 7A-Q show an exemplary molecular profiling report that incorporates the Genomic Prevalence Score (GPS; also Genomic Profiling Similarity) information according to the systems and methods provided herein.

FIGs. 8A-M show another exemplary molecular profiling report that incorporates the Genomic Prevalence Score information according to the systems and methods provided herein.

DETAILED DESCRIPTION

Described herein are methods and systems for characterizing various phenotypes of biological systems, organisms, cells, samples, or the like, by using molecular profiling, including systems, methods, apparatuses, and computer programs for training a machine learning model and then using the trained machine learning model to characterize such phenotypes. The term "phenotype" as used herein can mean any trait or characteristic that can be identified in part or in whole by using the systems and/or methods provided herein. In some implementations, the systems can include one or

more computer programs on one or more computers in one or more locations, e.g., configured for use in a method described herein.

Phenotypes to be characterized can be any phenotype of interest, including without limitation a tissue of origin, anatomical origin, histology, organ, medical condition, ailment, disease, disorder, or useful combinations thereof. A phenotype can be any observable characteristic or trait of, such as a disease or condition, a stage of a disease or condition, susceptibility to a disease or condition, prognosis of a disease stage or condition, a physiological state, or response / potential response (or lack thereof) to interventions such as therapeutics. A phenotype can result from a subject's genetic makeup as well as the influence of environmental factors and the interactions between the two, as well as from epigenetic modifications to nucleic acid sequences.

In various embodiments, a phenotype in a subject is characterized by obtaining a biological sample from a subject and analyzing the sample using the systems and/or methods provided herein. For example, characterizing a phenotype for a subject or individual can include detecting a disease or condition (including pre-symptomatic early stage detection), determining a prognosis, diagnosis, or theranosis of a disease or condition, or determining the stage or progression of a disease or condition. Characterizing a phenotype can include identifying appropriate treatments or treatment efficacy for specific diseases, conditions, disease stages and condition stages, predictions and likelihood analysis of disease progression, particularly disease recurrence, metastatic spread or disease relapse. A phenotype can also be a clinically distinct type or subtype of a condition or disease, such as a cancer or tumor. Phenotype determination can also be a determination of a physiological condition, or an assessment of organ distress or organ rejection, such as post-transplantation. The compositions and methods described herein allow assessment of a subject on an individual basis, which can provide benefits of more efficient and economical decisions in treatment.

Theranostics includes diagnostic testing that provides the ability to affect therapy or treatment of a medical condition such as a disease or disease state. Theranostics testing provides a theranosis in a similar manner that diagnostics or prognostic testing provides a diagnosis or prognosis, respectively. As used herein, theranostics encompasses any desired form of therapy related testing, including predictive medicine, personalized medicine, precision medicine, integrated medicine, pharmacodiagnostics and Dx/Rx partnering. Therapy related tests can be used to predict and assess drug response in individual subjects, thereby providing personalized medical recommendations. Predicting a likelihood of response can be determining whether a subject is a likely responder or a likely non-responder to a candidate therapeutic agent, e.g., before the subject has been exposed or otherwise treated with the treatment. Assessing a therapeutic response can be monitoring a response to a treatment, e.g., monitoring the subject's improvement or lack thereof over a time course after initiating the treatment. Therapy related tests are useful to select a subject for treatment who is particularly likely to benefit or lack benefit from the treatment or to provide an early and objective indication of treatment efficacy in an individual subject. Characterization using the systems and

methods provided herein may indicate that treatment should be altered to select a more promising treatment, thereby avoiding the expense of delaying beneficial treatment and avoiding the financial and morbidity costs of less efficacious or ineffective treatment(s).

In various embodiments, a theranosis comprises predicting a treatment efficacy or lack thereof, classifying a patient as a responder or non-responder to treatment. A predicted "responder" can refer to a patient likely to receive a benefit from a treatment whereas a predicted "non-responder" can be a patient unlikely to receive a benefit from the treatment. Unless specified otherwise, a benefit can be any clinical benefit of interest, including without limitation cure in whole or in part, remission, or any improvement, reduction or decline in progression of the condition or symptoms. The theranosis can be directed to any appropriate treatment, e.g., the treatment may comprise at least one of chemotherapy, immunotherapy, targeted cancer therapy, a monoclonal antibody, small molecule, or any useful combinations thereof.

The phenotype can comprise detecting the presence of or likelihood of developing a tumor, neoplasm, or cancer, or characterizing the tumor, neoplasm, or cancer (e.g., stage, grade, aggressiveness, likelihood of metastatis or recurrence, etc). In some embodiments, the cancer comprises an acute myeloid leukemia (AML), breast carcinoma, cholangiocarcinoma, colorectal adenocarcinoma, extrahepatic bile duct adenocarcinoma, female genital tract malignancy, gastric adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumors (GIST), glioblastoma, head and neck squamous carcinoma, leukemia, liver hepatocellular carcinoma, low grade glioma, lung bronchioloalveolar carcinoma (BAC), lung non-small cell lung cancer (NSCLC), lung small cell cancer (SCLC), lymphoma, male genital tract malignancy, malignant solitary fibrous tumor of the pleura (MSFT), melanoma, multiple myeloma, neuroendocrine tumor, nodal diffuse large B-cell lymphoma, non epithelial ovarian cancer (non-EOC), ovarian surface epithelial carcinoma, pancreatic adenocarcinoma, pituitary carcinomas, oligodendroglioma, prostatic adenocarcinoma, retroperitoneal or peritoneal carcinoma, retroperitoneal or peritoneal sarcoma, small intestinal malignancy, soft tissue tumor, thymic carcinoma, thyroid carcinoma, or uveal melanoma. The systems and methods herein can be used to characterize these and other cancers. Thus, characterizing a phenotype can be providing a diagnosis, prognosis or theranosis of one of the cancers disclosed herein.

In various embodiments, the phenotype comprises a tissue or anatomical origin. For example, the tissue can be muscle, epithelial, connective tissue, nervous tissue, or any combination thereof. For example, the anatomical origin can be the stomach, liver, small intestine, large intestine, rectum, anus, lungs, nose, bronchi, kidneys, urinary bladder, urethra, pituitary gland, pineal gland, adrenal gland, thyroid, pancreas, parathyroid, prostate, heart, blood vessels, lymph node, bone marrow, thymus, spleen, skin, tongue, nose, eyes, ears, teeth, uterus, vagina, testis, penis, ovaries, breast, mammary glands, brain, spinal cord, nerve, bone, ligament, tendon, or any combination thereof. Additional non-

limiting examples of phenotypes of interest include clinical characteristics, such as a stage or grade of a tumor, or the tumor's origin, e.g., the tissue origin.

In various embodiments, phenotypes are determined by analyzing a biological sample obtained from a subject. A subject (individual, patient, or the like) can include, but is not limited to, mammals such as bovine, avian, canine, equine, feline, ovine, porcine, or primate animals (including humans and non-human primates). In preferred embodiments, the subject is a human subject. A subject can also include a mammal of importance due to being endangered, such as a Siberian tiger; or economic importance, such as an animal raised on a farm for consumption by humans, or an animal of social importance to humans, such as an animal kept as a pet or in a zoo. Examples of such animals include, but are not limited to, carnivores such as cats and dogs; swine including pigs, hogs and wild boars; ruminants or ungulates such as cattle, oxen, sheep, giraffes, deer, goats, bison, camels or horses. Also included are birds that are endangered or kept in zoos, as well as fowl and more particularly domesticated fowl, e.g., poultry, such as turkeys and chickens, ducks, geese, guinea fowl. Also included are domesticated swine and horses (including race horses). In addition, any animal species connected to commercial activities are also included such as those animals connected to agriculture and aquaculture and other activities in which disease monitoring, diagnosis, and therapy selection are routine practice in husbandry for economic productivity and/or safety of the food chain. The subject can have a pre-existing disease or condition, including without limitation cancer. Alternatively, the subject may not have any known pre-existing condition. The subject may also be non-responsive to an existing or past treatment, such as a treatment for cancer.

Data Analysis and Machine Learning

Aspects of the present disclosure are directed towards a system that generates a set of one or more training data structures that can be used to train a machine learning model to provide various classifications, such as characterizing a phenotype of a biological sample. As described above, characterizing a phenotype can include providing a diagnosis, prognosis, theranosis or other relevant classification. For example, the classification may include a disease state, a predicted efficacy of a treatment for a disease or disorder of a subject, or the anatomical origin of a sample having a particular set of biomarkers. Once trained, the trained machine learning model can then be used to process input data provided by the system and make predictions based on the processed input data. The input data may include a set of features related to a subject such as data representing one or more subject biomarkers and data representing a phenotype of interest, e.g., a disease and/or anatomical origin. In some embodiments, the input data may further include features representing an anatomical origin and the system may make a prediction describing whether the sample is from that anatomical origin. The prediction may include data that is output by the machine learning model based on the machine learning model's processing of a specific set of features provided as an input to the machine learning model. The data may include without limitation data representing one or more subject

biomarkers, data representing a disease or anatomical origin, and data representing a proposed treatment type as desired.

As used herein, "biomarkers" or "sets of biomarkers" are used to train and test machine learning models and classify naïve samples. Such references include particular biomarkers such as particular nucleic acids or proteins, and optionally also include a state of such nucleic acids or proteins. Examples of the state of a biomarker include various aspects that can be queried such as presence, level (quantity, concentration, etc), sequence, location, activity, structure, modifications, covalent or non-covalent binding partners, and the like. As a non-limiting examples, a set of biomarkers may include a gene or gene product (i.e., mRNA or protein) having a specified sequence (e.g., KRAS mutant), and/or a gene or gene product and a level thereof (e.g., amplified ERBB2 gene or overexpressed HER2 protein). Useful biomarkers and aspects thereof are further described below.

Innovative aspects of the present disclosure include the extraction of specific data from incoming data streams for use in generating training data structures. An important aspect may be the selection of a specific set of one or more biomarkers for inclusion in the training data structure. This is because the presence, absence or other state of particular biomarkers may be indicative of the desired classification. For example, certain biomarkers may be selected to determine a desired phenotype, such as whether a treatment for a disease or disorder is of likely benefit, or a tumor origin. By way of example, in the present disclosure, the Applicant puts forth specific sets of biomarkers that, when used to train a machine learning model, result in a trained model that can more accurately predict a tumor origin than using a different set of biomarkers. See, e.g., **Examples 1-3**, **Tables 121-130**.

The system is configured to obtain output data generated by the trained machine learning model based on the machine learning model's processing of the input data. In various embodiments, the input data comprises biological data representing one or more biomarkers, data representing a disease or disorder, data representing a sample, data representing sample origins, or any combination thereof. The system may then predict an anatomical origin of a biological sample having a particular set of biomarkers. In some implementations, the disease or disorder may include a type of cancer and the anatomical origins can include various tissues and organs. In this setting, output of the trained machine learning model that is generated based on trained machine learning model processing of the input data that includes the set of biomarkers, the disease or disorder and various anatomical origins includes data representing the predicted anatomical origin of the biological sample.

In some implementations, the output data generated by the trained machine learning model includes a probability of the desired classification. By way of illustration, such probability may be a probability that the biological sample is derived from tissue from a particular organ. In other implementations, the output data may include any output data generated by the trained machine learning model based on the trained machine learning model's processing of the input data. In some embodiments, the input data comprises set of biomarkers, data representing the disease or disorder, data representing a sample, the data representing the sample origin, or any combination thereof.

In some implementations, the training data structures generated by the present disclosure may include a plurality of training data structures that each include fields representing feature vector corresponding to a particular training sample. The feature vector includes a set of features derived from, and representative of, a training sample. The training sample may include, for example, one or more biomarkers of a biological sample, a disease or disorder associated with the biological sample, and an anatomical origin from the biological sample. The training data structures are flexible because each respective training data structure may be assigned a weight representing each respective feature of the feature vector. Thus, each training data structure of the plurality of training data structures can be particularly configured to cause certain inferences to be made by a machine learning model during training.

Consider a non-limiting example wherein the model is trained to make a prediction of likely anatomical origin of a biological sample, e.g., a tumor sample. As a result, the novel training data structures that are generated in accordance with this specification are designed to improve the performance of a machine learning model because they can be used to train a machine learning model to predict an anatomical origin of a biological sample having a particular set of biomarkers. By way of example, a machine learning model that could not perform predictions regarding the anatomical origin of a biological sample having a particular set of biomarkers prior to being trained using the training data structures, system, and operations described by this disclosure can learn to make predictions regarding the anatomical origin of a biological sample having a particular set of biomarkers by being trained using the training data structures, systems and operations described by the present disclosure. Accordingly, this process takes an otherwise general purpose machine learning model and changes the general purpose machine leaning model into a specific computer for perform a specific task of performing predicting the anatomical origin of a biological sample having a particular set of biomarkers.

FIG. 1A is a block diagram of an example of a prior art system 100 for training a machine learning model 110. In some implementations, the machine learning model may be, for example, a support vector machine. Alternatively, the machine learning model may include a neural network model, a linear regression model, a random forest model, a logistic regression model, a naive Bayes model, a quadratic discriminant analysis model, a K-nearest neighbor model, a support vector machine, or the like. The machine learning model training system 100 may be implemented as computer programs on one or more computers in one or more locations, in which the systems, components, and techniques described below can be implemented. The machine learning model training system 100 trains the machine learning model 110 using training data items from a database (or data set) 120 of training data items. The training data items may include a plurality of feature vectors. Each training vector may include a plurality of values that each correspond to a particular feature of a training sample that the training vector represents. The training features may be referred

to as independent variables. In addition, the system 100 maintains a respective weight for each feature that is included in the feature vectors.

The machine learning model 110 is configured to receive an input training data item 122 and to process the input training data item 122 to generate an output 118. The input training data item may include a plurality of features (or independent variables "X") and a training label (or dependent variable "Y"). The machine learning model may be trained using the training items, and once trained, is capable of predicting $X = f(Y)$.

To enable machine learning model 110 to generate accurate outputs for received data items, the machine learning model training system 100 may train the machine learning model 110 to adjust the values of the parameters of the machine learning model 110, e.g., to determine trained values of the parameters from initial values. These parameters derived from the training steps may include weights that can be used during the prediction stage using the fully trained machine learning model 110.

In training, the machine learning model 110, the machine learning model training system 100 uses training data items stored in the database (data set) 120 of labeled training data items. The database 120 stores a set of multiple training data items, with each training data item in the set of multiple training items being associated with a respective label. Generally, the label for the training data item identifies a correct classification (or prediction) for the training data item, i.e., the classification that should be identified as the classification of the training data item by the output values generated by the machine learning model 110. With reference to **FIG. 1A**, a training data item 122 may be associated with a training label 122a.

The machine learning model training system 100 trains the machine learning model 110 to optimize an objective function. Optimizing an objective function may include, for example, minimizing a loss function 130. Generally, the loss function 130 is a function that depends on the (i) output 118 generated by the machine learning model 110 by processing a given training data item 122 and (ii) the label 122a for the training data item 122, i.e., the target output that the machine learning model 110 should have generated by processing the training data item 122.

Conventional machine learning model training system 100 can train the machine learning model 110 to minimize the (cumulative) loss function 130 by performing multiple iterations of conventional machine learning model training techniques on training data items from the database 120, e.g., hinge loss, stochastic gradient methods, stochastic gradient descent with backpropagation, or the like, to iteratively adjust the values of the parameters of the machine learning model 110. A fully trained machine learning model 110 may then be deployed as a predicting model that can be used to make predictions based on input data that is not labeled.

FIG. 1B is a block diagram of a system that generates training data structures for training a machine learning model to predict a sample origin.

The system 200 includes two or more distributed computers 210, 310, a network 230, and an application server 240. The application server 240 includes an extraction unit 242, a memory unit 244, a vector generation unit 250, and a machine learning model 270. The machine learning model 270 may include one or more of a neural network model, a linear regression model, a random forest model, a logistic regression model, a naive Bayes model, a quadratic discriminant analysis, model, a K-nearest neighbor model, a support vector machine, or the like. Each distributed computer 210, 310 may include a smartphone, a tablet computer, laptop computer, or a desktop computer, or the like. Alternatively, the distributed computers 210, 310 may include server computers that receive data input by one or more terminals 205, 305, respectively. The terminal computers 205, 305 may include any user device including a smartphone, a tablet computer, a laptop computer, a desktop computer or the like. The network 230 may include one or more networks 230 such as a LAN, a WAN, a wired Ethernet network, a wireless network, a cellular network, the Internet, or any combination thereof.

The application server 240 is configured to obtain, or otherwise receive, data records 220, 222, 224, 320 provided by one or more distributed computers such as the first distributed computer 210 and the second distributed computer 310 using the network 230. In some implementations, each respective distributed computer 210, 310 may provide different types of data records 220, 222, 224, 320. For example, the first distributed computer 210 may provide biomarker data records 220, 222, 224 representing biomarkers for a biological sample from a subject and the second distributed computer 310 may provide sample data 320 representing anatomical origin or other sample data for a subject obtained from the sample database 312. However, the present disclosure need not be limited to two computers 210, 310 providing data records 220, 222, 224, 230. Though such implementations can provide technical advantages such as load balancing, bandwidth optimization, or both, it is also contemplated that the data records 220, 222, 224, 230 can each be provided by the same computer.

The biomarker data records 220, 222, 224 may include any type of biomarker data that describes biometric attributes of a biological sample. By way of example, the example of **FIG. 1B** shows the biomarker data records as including data records representing DNA biomarkers 220, protein biomarkers 222, and RNA data biomarkers 224. These biomarker data records may each include data structures having fields that structure information 220a, 222a, 224a describing biomarkers of a subject such as a subject's DNA biomarkers 220a, protein biomarkers 222a, or RNA biomarkers 224a. However, the present disclosure need not be so limited and any useful biomarkers can be assessed. In some embodiments, the biomarker data records 220, 222, 224 include next generation sequencing data from DNA and/or RNA, including without limitation single variants, insertions and deletions, substitution, translocation, fusion, break, duplication, amplification, loss, copy number, repeat, total mutational burden, microsatellite instability, or the like. Alternatively, or in addition, the biomarker data records 220, 222, 224 may also include in situ hybridization data. Such in situ hybridization data may include DNA copy numbers, translocations, or the like. Alternatively, or in addition, the biomarker data records 220, 222, 224 may include RNA data such as gene expression or

gene fusion, including without limitation data derived from whole transcriptome sequencing. Alternatively, or in addition, the biomarker data records 220, 222, 224 may include protein expression data such as obtained using immunohistochemistry (IHC). Alternatively, or in addition, the biomarker data records 220, 222, 224 may include ADAPT data such as complexes.

In some implementations, the biomarker data records 220, 222, 224 include one or more biomarkers and attributes listed in any one of **Tables 2-116**, **Tables 117-120**, ISNM1, **Tables 121-130**. However, the present disclosure need not be so limited, and other types of biomarkers may be used as desired. For example, the biomarker data may be obtained by whole exome sequencing, whole transcriptome sequencing, whole genome sequencing, or a combination thereof.

The sample data records 320 may describe various aspects of a biological sample, e.g., a tissue and/or organ from which the sample is derived. For example, the sample data records 320 obtained from the sample database 312 may include one or more data structures having fields that structure data attributes of a biological sample such as a disease or disorder 320a-1 ("ailment"), a tissue or organ 320a-2 where the sample was obtained, a sample type 320a-3, a verified sample origin label 320a-4, or any combination thereof. The sample record 320 can include up to n data records describing a sample, where n is any positive integer greater than 0. For example, though the example of **FIG. 1B** trains the machine learning model using patient sample data describing disease / disorder, tissue / organ where sample was obtained, and sample type, the present disclosure is not so limited. For example, in some implementations, the machine learning model 370 can be trained to predict the origin of sample using patient sample information that includes the tissue or organ 320a-2 where the sample was obtained and sample type 320a-3 without including the ailment or disorder 320a-1.

Alternatively, or in addition, the sample data records 320 may also include fields that structure data attributes describing details of the biological sample, including attributes of a subject from which the sample is derived. An example of a disease or disorder may include, for example, a type of cancer. A tissue or organ may include, for example, a type of tissue (e.g., muscle tissue, epithelial tissue, connective tissue, nervous tissue, etc.) or organ (e.g., colon, lung, brain, etc.). A sample type may include data representing the type of sample, such as tumor sample, bodily fluid, fresh or frozen, biopsy, FFPE, or the like. In some implementations, attributes of a subject from which the sample is derived include clinical attributes such as pathology details of the sample, subject age and/or sex, prior subject treatments, or the like. If the sample is a metastatic sample of unknown primary origin (i.e., a cancer of unknown primary (CUPS)), the attributes may include the location from which the sample was taken. As a non-limiting example, a metastatic lesion of unknown primary origin may be found in the liver or brain. Accordingly, though the example of **FIG. 1B** shows that sample data may include a disease or disorder, a tissue or organ, and a sample type, the sample data may include other types of information, as described herein. Moreover, there is no requirements that the sample data be limited to human "patients." Instead, the sample data records 220, 222, 224 and biometric data records 320 may be associated with any desired subject including any non-human organism.

In some implementations, each of the data records 220, 222, 224, 320 may include keyed data that enables the data records from each respective distributed computer to be correlated by application server 240. The keyed data may include, for example, data representing a subject identifier. The subject identifier may include any form of data that identifies a subject and that can associate biomarker for the subject with sample data for the subject.

The first distributed computer 210 may provide 208 the biomarker data records 220, 222, 224 to the application server 240. The second distributed computer 310 may provide 210 the sample data records 320 to the application server 240. The application server 240 can provide the biomarker data records 220 and the sample data records 220, 222, 224 to the extraction unit 242.

The extraction unit 242 can process the received biomarker data 220, 222, 224 and sample data records 320 in order to extract data 220a-1, 222a-1, 224a-1, 320a-1, 320a-2, 320a-3 that can be used to train the machine learning model. For example, the extraction unit 242 can obtain data structured by fields of the data structures of the biometric data records 220, 222, 224, obtain data structured by fields of the data structures of the outcome data records 320, or a combination thereof. The extraction unit 242 may perform one or more information extraction algorithms such as keyed data extraction, pattern matching, natural language processing, or the like to identify and obtain data 220a-1, 222a-1, 224a-1, 320a-1, 320a-2, 320a-3 from the biometric data records 220, 222, 224 and sample data records 320, respectively. The extraction unit 242 may provide the extracted data to the memory unit 244. The extracted data unit may be stored in the memory unit 244 such as flash memory (as opposed to a hard disk) to improve data access times and reduce latency in accessing the extracted data to improve system performance. In some implementations, the extracted data may be stored in the memory unit 244 as an in-memory data grid.

In more detail, the extraction unit 242 may be configured to filter a portion of the biomarker data records 220, 222, 224 and the sample data records 320 such as 220a-1, 222a-1, 224a-1, 320a-1, 320a-2, 320a-3 that will be used to generate an input data structure 260 for processing by the machine learning model 270 from the portion of the sample data records 320a-4 that will be used as a label for the generated input data structure 260. Such filtering includes the extraction unit 242 separating the biomarker data and a first portion of the sample data that includes a disease or disorder 320a-1, tissue / organ 320a-1 where sample was obtained (e.g., biopsied), sample type 320a-3 details, or any combination thereof, from the verified origin of the sample 320a-4. The verified sample origin of the sample may be a different tissue / organ or the same tissue / organ than the sample was obtained from. An example of who the tissue / organ that the sample was obtained from can be different than the verified origin can include instances where the disease or disorder has spread from a first tissue / organ to a second tissue / organ from which the sample was then obtained. The application server 240 can then use the biomarker data 220a-1, 222a-1, 224a-1, and the first portion of the sample data that includes the disease or disorder 320a-1, tissue or organ 320a-2, sample type details (not shown in **FIG. 1B**), or a combination thereof, to generate the input data structure 260. In addition, the

application server 240 can use the second portion of the sample data describing the verified origin of the sample 320a-4 as the label for the generated data structure.

The application server 240 may process the extracted data stored in the memory unit 244 correlate the biomarker data 220a-1, 222a-1, 224a-1 extracted from biomarker data records 220, 222, 224 with the first portion of the sample data 320a-1, 320a-2, 320a-3. The purpose of this correlation is to cluster biomarker data with sample data so that the sample data for the biological sample is clustered with the biomarker data for the same biological sample. In some implementations, the correlation of the biomarker data and the first portion of the sample data may be based on keyed data associated with each of the biomarker data records 220, 222, 224 and the sample data records 320. For example, the keyed data may include a sample identifier or a subject identifier, e.g., a subject from which the sample is derived.

The application server 240 provides the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3 as an input to a vector generation unit 250. The vector generation unit 250 is used to generate a data structure based on the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3. The generated data structure is a feature vector 260 that includes a plurality of values that numerical represents the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3. The feature vector 260 may include a field for each type of biomarker and each type of sample data. For example, the feature vector 260 may include one or more fields corresponding to (i) one or more types of next generation sequencing data such as single variants, insertions and deletions, substitution, translocation, fusion, break, duplication, amplification, loss, copy number, repeat, total mutational burden, microsatellite instability, (ii) one or more types of in situ hybridization data such as DNA copy number, gene copies, gene translocations, (iii) one or more types of RNA data such as gene expression or gene fusion, (iv) one or more types of protein data such as presence, level or cellular location obtained using immunohistochemistry, (v) one or more types of ADAPT data such as complexes, and (vi) one or more types of sample data such as disease or disorder, sample type, each sample details, or the like.

The vector generation unit 250 is configured to assign a weight to each field of the feature vector 260 that indicates an extent to which the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3 includes the data represented by each field. In one implementation, for example, the vector generation unit 250 may assign a '1' to each field of the feature vector that corresponds to a feature found in the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3. In such implementations, the vector generation unit 250 may, for example, also assign a '0' to each field of the feature vector that corresponds to a feature not found in the extracted biomarker data 220a-1, 222a-1, 224a-1 and the extracted first portion of the sample data 320a-1, 320a-2, 320a-3. The output

of the vector generation unit 250 may include a data structures such as a feature vector 260 that can be used to train the machine learning model 270.

The application server 240 can label the training feature vector 260. Specifically, the application server can use the extracted second portion of the sample data 320a-4 to label the generated feature vector 260 with a verified sample origin 320a-4. The label of the training feature vector 260 generated based on the verified sample origin 320a-4 can be used to predict the tissue or organ that was the origin for a biological sample represented by the sample record 320 and having disease or disorder 320a-1 defined by the specific set of biomarkers 220a-1, 222a-1, 224a-1, each of which is described by described in the training data structure 260.

The application server 240 can train the machine learning model 270 by providing the feature vector 260 as an input to the machine learning model 270. The machine learning model 270 may process the generated feature vector 260 and generate an output 272. The application server 240 can use a loss function 280 to determine the amount of error between the output 272 of the machine learning model 280 and the value specified by the training label, which is generated based on the second portion of the extracted sample data describing the verified sample origin 320a-4. The output 282 of the loss function 280 can be used to adjust the parameters of the machine learning model 282.

In some implementations, adjusting the parameters of the machine learning model 270 may include manually tuning of the machine learning model parameters model parameters. Alternatively, in some implementations, the parameters of the machine learning model 270 may be automatically tuned by one or more algorithms of executed by the application server 242.

The application server 240 may perform multiple iterations of the process described above with reference to **FIG. 1B** for each sample data record 320 stored in the sample database that correspond to a set of biomarker data for a biological sample. This may include hundreds of iterations, thousands of iterations, tens of thousands of iterations, hundreds of thousands of iterations, millions of iterations, or more, until each of the sample data records 320 stored in the sample database 312 and having a corresponding set of biomarker data for a biological sample are exhausted, until the machine learning model 270 is trained to within a particular margin of error, or a combination thereof. A machine learning model 270 is trained within a particular margin of error when, for example, the machine learning model 270 is able to predict, based upon a set of unlabeled biomarker data, disease or disorder data, and sample type data, an origin of an sample having the biomarker data. The origin may include, for example, a probability, a general indication of the confidence in the origin classification, or the like.

FIG. 1C is a block diagram of a system for using a trained machine learning model 370 to predict a sample origin of sample data from a subject.

The machine learning model 370 includes a machine learning model that has been trained using the process described with reference to the system of **FIG. 1B** above. For example, **FIG. 1B** is an example of a machine learning model 370 that has been trained to predict sample origin using

patient sample data that comprises data representing a tissue / organ 422a where the sample was obtained and a sample type 420a. In the example of **FIG. 1B**, a disease, disorder, or ailment was not used to train the model – though there may be implementations of the present disclosure where the machine learning model 370 can be trained using an ailment or disorder in addition to a tissue / organ 422a where the sample was obtained and a sample type 420a. The trained machine learning model 370 is capable of predicting, based on an input feature vector representative of a set of one or more biomarkers, a disease or disorder, and other relevant sample data such as sample type, a origin of a biological sample having the biomarkers. In some implementations, the "origin" may include an anatomical system, location, organ, tissue type, and the like.

The application server 240 hosting the machine learning model 370 is configured to receive unlabeled biomarker data records 320, 322, 324. The biomarker data records 320, 322, 324 include one or more data structures that have fields structuring data that represents one or more particular biomarkers such as DNA biomarkers 320a, protein biomarkers 322a, RNA biomarkers 324a, or any combination thereof. As discussed above, the received biomarker data records may include various types of biomarkers not explicitly depicted by **FIG. 1C** such as (i) next generation sequencing data from DNA and/or RNA, including without limitation single variants, insertions and deletions, substitution, translocation, fusion, break, duplication, amplification, loss, copy number, repeat, total mutational burden, microsatellite instability, or the like, (ii) one or more types of in situ hybridization data such as DNA copies, gene copies, gene translocations, (iii) one or more types of RNA data such as gene expression or gene fusion, (iv) one or more types of protein data such as presence, level or location obtained using immunohistochemistry, or (v) one or more types of ADAPT data such as complexes. In some implementations, the biomarker data records 320, 322, 324 include one or more biomarkers and attributes listed in any one of **Tables 2-116**, **Tables 117-120**, ISNM1, and/or **Tables 121-130**. However, the present disclosure need not be so limited, and other biomarkers may be used as desired. For example, the biomarker data may be obtained by whole exome sequencing, whole transcriptome sequencing, or a combination thereof.

The application server 240 hosting the machine learning model 370 is also configured to receive sample data 420 representing a proposed origin data 422a for a biological sample described by the sample data 420a of the biological sample having biomarkers represented by the received biomarker data records 320, 322, 324. The proposed origin data 422a for the biological sample 420a are also unlabeled and merely a suggestion for the origin of a biological sample having biomarkers representing by biomarker data records 320, 322, 324. However, as discussed elsewhere herein, due to the potential for disease (e.g., cancer) to spread from, e.g., organ to organ, the tissue / organ 422a where a sample was obtained may not be the actual sample origin.

In some implementations, the sample data 420 is received or provided 305 by a terminal 405 over the network 230 and the biomarker data is obtained from a second distributed computer 310. The biomarker data may be derived from laboratory machinery used to perform various assays. See, e.g.,

Example 1 herein. The sample data 420 can include data representing a tissue / organ 422a where the sample was obtained and a sample type 420a. The tissue / organ 422a from where the sample was obtained may be referred to as the proposed origin of the sample. In other implementations, the sample data 420a, the proposed origin 422a, and the biomarker data 320, 322, 324 may each be received from the terminal 405. For example, the terminal 405 may be user device of a doctor, an employee or agent of the doctor working at the doctor's office, or other human entity that inputs data representing a sample, data representing a proposed origin, and a data representing patient attributes for a the biological sample. In some implementations, the sample data 420 may include data structures structuring fields of data representing a proposed origin described by a tissue or organ name. In other implementations, the sample data 420 may include data structures structuring fields of data representing more complex sample data such as sample type, age and/or sex of the patient from which the sample is derived, or the like.

The application server 240 receives the biomarker data records 320, 322, 324, the sample data 420, and the proposed origin data 422. The application server 240 provides the biomarker data records 320, 322, 324, the sample data 420, and the origin data 422 to an extraction unit 242 that is configured to extract (i) particular biomarker data such as DNA biomarker data 320a-1, protein expression data 322a-1, 324a-1, (ii) sample data 420a-1, and (iii) proposed origin data 422a-1 from the fields of the biomarker data records 320, 322, 324 and the sample data records 420, 422. In some implementations, the extracted data is stored in the memory unit 244 as a buffer, cache or the like, and then provided as an input to the vector generation unit 250 when the vector generation unit 250 has bandwidth to receive an input for processing. In other implementations, the extracted data is provided directly to a vector generation unit 250 for processing. For example, in some implementations, multiple vector generation units 250 may be employed to enable parallel processing of inputs to reduce latency.

The vector generation unit 250 can generate a data structure such as a feature vector 360 that includes a plurality of fields and includes one or more fields for each type of biomarker data and one or more fields for each type of origin data. For example, each field of the feature vector 360 may correspond to (i) each type of extracted biomarker data that can be extracted from the biomarker data records 320, 322, 324 such as each type of next generation sequencing data, each type of in situ hybridization data, each type of RNA or DNA data, each type of protein (e.g., immunohistochemistry) data, and each type of ADAPT data and (ii) each type of sample data that can be extracted from the sample data records 420, 422 such as each type of disease or disorder, each type of sample, and each type of origin details.

The vector generation unit 250 is configured to assign a weight to each field of the feature vector 360 that indicates an extent to which the extracted biomarker data 320a-1, 322a-1, 324a-1, the extracted sample 420a-1, and the extracted origin 422a-1 includes the data represented by each field. In one implementation, for example, the vector generation unit 250 may assign a '1' to each field of the feature vector 360 that corresponds to a feature found in the extracted biomarker data 320a-1,

322a-1, 324a-1, the extracted sample 420a-1, and the extracted origin 422a-1. In such implementations, the vector generation unit 250 may, for example, also assign a '0' to each field of the feature vector that corresponds to a feature not found in the extracted biomarker data 320a-1, 322a-1, 324a-1, the extracted sample 420a-1, and the extracted origin 422a-1. The output of the vector generation unit 250 may include a data structure such as a feature vector 360 that can be provided as an input to the trained machine learning model 370.

The trained machine learning model 370 process the generated feature vector 360 based on the adjusted parameters that were determining during the training stage and described with reference to **FIG. 1B**. The output 272 of the trained machine learning model provides an indication of the origin 422a-1 of the sample 420a-1 for the biological sample having biomarkers 320a-1, 322a-1, 324a-1. In some implementations, the output 272 may include a probability that is indicative of the origin 422a-1 of the sample 420a-1 for the biological sample having biomarkers 320a-1, 322a-1, 324a-1. In such implementations, the output 272 may be provided 311 to the terminal 405 using the network 230. The terminal 405 may then generate output on a user interface 420 that indicates a predicted origin for the biological sample having the biomarkers represented by the feature vector 360.

In other implementations, the output 272 may be provided to a prediction unit 380 that is configured to decipher the meaning of the output 272. For example, the prediction unit 380 can be configured to map the output 272 to one or more categories of effectiveness. Then, the output of the prediction unit 328 can be used as part of message 390 that is provided 311 to the terminal 305 using the network 230 for review by laboratory staff, a healthcare provider, a subject, a guardian of the subject, a nurse, a doctor, or the like.

FIG. 1D is a flowchart of a process 400 for generating training data structures for training a machine learning model to predict sample origin. In one aspect, the process 400 may include obtaining, from a first distributed data source, a first data structure that includes fields structuring data representing a set of one or more biomarkers associated with a biological sample (410), storing the first data structure in one or more memory devices (420), obtaining from a second distributed data source, a second data structure that includes fields structuring data representing the biological sample and origin data for the biological sample having the one or more biomarkers (430), storing the second data structure in the one or more memory devices (440), generating a labeled training data structure that structures data representing (i) the one or more biomarkers, (ii) a biological sample, (iii) an origin, and (iv) a predicted origin for the biological sample based on the first data structure and the second data structure (450), and training a machine learning model using the generated labeled training data (460).

FIG. 1E is a flowchart of a process 500 for using a trained machine learning model to predict sample origin of sample data from a subject. In one aspect, the process 500 may include obtaining a data structure representing a set of one or more biomarkers associated with a biological sample (510), obtaining data representing sample data for the biological sample (520), obtaining data representing a

origin type for the biological sample (530), generating a data structure for input to a machine learning model that structures data representing (i) the one or more biomarkers, (ii) the biological sample, and (iii) the origin type (540), providing the generated data structure as an input to the machine learning model that has been trained to predict sample origins using labeled training data structures structuring data representing one or more obtained biomarkers, one or more sample types, and one or more origins (550), and obtaining an output generated by the machine learning model based on the machine learning model processing of the provided data structure (560), and determining a predicted origin for the biological sample having the one or more biomarkers based on the obtained output generated by the machine learning model (570).

Provided herein are methods of employing multiple machine learning models to improve classification performance. Conventionally, a single model is chosen to perform a desired prediction/classification. For example, one may compare different model parameters or types of models, e.g., random forests, support vector machines, logistic regression, k-nearest neighbors, artificial neural network, naïve Bayes, quadratic discriminant analysis, or Gaussian processes models, during the training stage in order to identify the model having the optimal desired performance. Applicant realized that selection of a single model may not provide optimal performance in all settings. Instead, multiple models can be trained to perform the prediction/classification and the joint predictions can be used to make the classification. In this scenario, each model is allowed to "vote" and the classification receiving the majority of the votes is deemed the winner.

This voting scheme disclosed herein can be applied to any machine learning classification, including both model building (e.g., using training data) and application to classify naïve samples. Such settings include without limitation data in the fields of biology, finance, communications, media and entertainment. In some preferred embodiments, the data is highly dimensional "big data." In some embodiments, the data comprises biological data, including without limitation biological data obtained via molecular profiling such as described herein. See, e.g., **Example 1**. The molecular profiling data can include without limitation highly dimensional next-generation sequencing data, e.g., for particular biomarker panels (see, e.g., **Example 1**) or whole exome and/or whole transcriptome data. The classification can be any useful classification, e.g., to characterize a phenotype. For example, the classification may provide a diagnosis (e.g., disease or healthy), prognosis (e.g., predict a better or worse outcome), theranosis (e.g., predict or monitor therapeutic efficacy or lack thereof), or other phenotypic characterization (e.g., origin of a CUPs tumor sample).

FIG. 1F is an example of a system for performing pairwise analysis to predict a sample origin. A disease type can include, for example, an origin of a subject sample processed by the system. An origin of a subject sample can include, for example location of a subject's body where a disease, such as cancer, originated. With reference to a practical example, a biopsy of a subject tumor may be obtained from a subject's liver. Then, input data can be generated based on the biopsied tumor and provided as an input to the pairwise analysis model 340. The model can compare the

generated input data to a corresponding biological signature of each known type of disease (e.g., different cancer types). Based on the output generated by the pairwise analysis model 340, the computer 310 can determine whether biopsied tumor represented by the input data originated in the liver or in some other portion of the subject's body such as the pancreas. One or more treatments can then be determined based on the ***origin*** of the disease as opposed to the treatments being based on the biopsied tumor, alone.

In more detail, the system 300 can include one or more processors and one or more memory units 320 storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations. In some implementations, the one or more processors and the one or memories 320 may be implemented in a computer such as a computer 310.

The system 300 can obtain first biological signature data 322, 324 as an input. The first biological signature 322, 324 data can include one or more biomarkers 322, sample data 324, or both. Sample data 324 can include data representing the sample that was obtained from the body, e.g., a tissue sample, tumor sample, malignant fluid, or other sample such as described herein. In some implementations, the biological signature 322, 324 represents features of a disease, e.g., a cancer. In some implementations, the features may represent molecular data obtained using next generation sequencing (NGS). In some implementations, the features may be present in the DNA of a disease sample, including without limitation mutations, polymorphisms, deletions, insertions, substitutions, translocations, fusions, breaks, duplications, loss, amplification, repeats, or gene copy numbers. In some implementations, the features may be present in the RNA of a disease.

The system can generate input data for input to a machine learning model 340 that has been trained to perform pairwise analysis. The machine learning model can include a neural network model, a linear regression model, a random forest model, a logistic regression model, a naive Bayes model, a quadratic discriminant analysis model, a K-nearest neighbor model, a support vector machine, or the like. The machine learning model 340 can be implemented as one or more computer programs on one or more computers in one or more locations.

In some implementations, the generated input data may include data representing the biological signature 322, 324. In other implementations, the generated data that represents the biological signature can include a vector 332 generated using a vector generation unit 330. For example, the vector generation unit 330 can obtain biological signature data 322, 324 from the memory unit 320 and generate an input vector 333, based on the biological signature data 322, 324 that represents the biological signature data 322, 324 in a vector space. The generated vector 332 can be provided, as an input, to the pairwise analysis model 340.

The pairwise analysis model 340 can be configured to perform pairwise analysis of the input vector 352 representing the biological signature 322, 324 with each biological signature 341-1, 341-2, 341-n, where n is any positive, non-zero integer. Each of the multiple different biological signatures correspond to a different type of disease, e.g., a different type of cancer. In some implementations,

the model 340 can be a single model that is trained to determine a source of a sample based on in input sample by determining a level of similarity of features of an input sample to each of a plurality of biological signature classifications represented by biological signatures 341-1, 341-2, 341-n. In other implementations, the model 340 can include multiple different models that each perform a pairwise comparison between an input vector 332 and one biological signature such as 341-1. In such instances, output data generated by each of the models can be evaluated by a voting unit to determine a source of a sample represented by the processed input vector 332.

The pairwise analysis model 340 can generate an output 342 that can be obtained by the system such as computer 310. The output 342 can indicate a likely disease type of the sample based on the pairwise analysis. In some implementations, the output 342 can include a matrix such as the matrix described in **FIG. 5B**. The system can determine, based on the generated matrix and using the prediction unit 350, data 360 indicating a likely disease type.

Example 2 herein provides an implementation of such a system. In the **Example**, the models are trained to distinguish 115 disease types, where each disease type comprises a primary tumor origin and histology. In some embodiments, the data 360 provides a list of disease types ranked by probability. If desired, the data 360 can be presented as an aggregate of various disease types. In the Example, such aggregation of Organ Groups is presented, wherein each Organ Group comprises appropriate disease types. As an example, the Organ Group "colon" comprises the disease types "colon adenocarcinoma, NOS; colon carcinoma, NOS; colon mucinous adenocarcinoma" and the like.

FIG. 1G is a block diagram of a system for predicting a sample origin using a voting unit to interpret output generated by multiple machine learning models that are each trained to perform pairwise analysis. The system 600 is similar to the system 300 of **FIG. 1F**. However, instead of a single machine learning model 340 trained to perform pairwise analysis, the system 600 includes multiple machine learning models 340-0, 340-1 ... 340-x, where x is any non-zero integer greater than 1, that have been trained to perform pairwise analysis. The system 600 also include a voting unit 480. As a non-limiting example, system 600 can be used for predicting origin and related attributes of a biological sample having a particular set of biomarkers. See, e.g., **Examples 2-3**.

Each machine learning model 370-0, 370-1, 370-x can include a machine learning model that has been trained to classify a particular type of input data 320-0, 320-1 ... 320-x, wherein x is any non-zero integer greater than 1 and equal to the number x of machine learning models. In some implementations, each machine learning models 340-0, 340-1, 340-x (labeled PW Compare Models in **FIG. 1G**) can be trained, or otherwise configured, to perform a particular pairwise comparison between (i) an input vector including data representing the sample data and (ii) another vector representing a particular biological signature including data representing a known disease type, portion of a subject body, or a both. Accordingly, in such implementations, the classification operation can include classifying (i) an input data vector including data representing sample data (e.g., sample origin, sample type, or the like) and (ii) one or more biomarkers associated with the sample as

being sufficiently similar to a biological signature associated with the particular machine learning model or not sufficiently similar to the biological signature associated with the particular machine learning model. In some implementations, an input vector may be sufficiently similar to a biological signature if a similarity between the input vector and biological signature satisfies a predetermined threshold.

In some implementations, each of the machine learning models 340-0, 340-1, 340-x can be of the same type. For example, each of the machine learning models 340-0, 340-1, 340-x can be a random forest classification algorithm, e.g., trained using differing parameters. In other implementations, the machine learning models 340-0, 340-1, 340-x can be of different types. For example, there can be one or more random forest classifiers, one or more neural networks, one or more K-nearest neighbor classifiers, other types of machine learning models, or any combination thereof.

Input data such as 420 representing sample data and one or more biomarkers associated with the sample can be obtained by the application server 240. The sample data can include a sample type, sample origin, or the like, as described herein. In some implementations, the input data 420 is obtained across the network 230 from one or more distributed computers 310, 405. By way of example, one or more of the input data items 420 can be generated by correlating data from multiple different data sources 210, 405. In such an implementation, (i) first data describing biomarkers for a biological sample can be obtained from the first distributed computer 310 and (ii) second data describing a biological sample and related data can be obtained from the second computer 405. The application server 240 can correlate the first data and the second data to generate an input data structure such as input data structure 420. This process is described in more detail in **FIG. 1C**. The input data 420 can be provided to the vector generation unit 250. The vector generation unit 250 can generate input vectors 360-0, 360-1, 360-x that that each represent the input data 420. While some implementations may generate vectors 360-0, 360-1, 360-x serially, the present disclosure need not be so limited.

In some implementations, each input data structure 320-0, 320-1, 320-x can include data representing biomarkers of a biological sample, data describing a biological sample and related data (e.g., a sample type, disease or disorder associated with the sample, and/or patient characteristics from which the sample is derived), or any combination thereof. The data representing the biomarkers of a biological sample can include data describing a specific subset or panel of genes or gene products. Alternatively, in some implementations, the data representing biomarkers of the biological sample can include data representing complete set of known genes or gene products, e.g., via whole exome sequencing and/or whole transcriptome sequencing. The complete set of known genes can include all of the genes of the subject from which the biological sample is derived. In some implementations, each of the machine learning models 340-0, 340-1, 340-x are the same type machine learning model such as a random forest model trained to classify the input data vectors as corresponding to a sample

origin (e.g., tissue or organ) associated by the vector processed by the machine learning model. In such implementations, though each of the machine learning models 340-0, 340-1, 340-x is the same type of machine learning model, each of the machine learning models 340-0, 340-1, 340-x may be trained in different ways. The machine learning models 340-0, 340-1, 340-x can generate output data 372-0, 372-1, 372-x, respectively, representing whether a biological sample associated with input vectors 360-0, 360-1, 360-x is likely to be derived from an anatomical origin associated with the input vectors 360-0, 360-1, 360-x. In this example, the input data sets, and their corresponding input vectors, are the same – e.g., each set of input data has the same biomarkers, same sample type, same origin, or any combination thereof. Nonetheless, given the different training methods used to train each respective machine learning model 340-0, 340-1, 340-x may generate different outputs 372-0, 372-1, 372-x, respectively, based on each machine learning model 370-0, 370-1, 370-x processing the input vector 360-0, 361-1, 361-x, as shown in **FIG. 1G**.

Alternatively, each of the machine learning models 340-0, 340-1, 340-x can be a different type of machine learning model that has been trained, or otherwise configured, to classify input data as most likely origin of a biological sample. For example, the first machine learning model 340-1 can include a neural network, the machine learning model 340-1 can include a random forest classification algorithm, and the machine learning model 340-x can include a K-nearest neighbor algorithm. In this example, each of these different types of machine learning models 340-0, 340-1, 340-x can be trained, or otherwise configured, to receive and process an input vector and determine whether the input vector is associated with to a sample origin also associated with the input vector. In this example, the input data sets, and their corresponding input vectors, can be the same – e.g., each set of input data has the same biomarkers, same sample type, same origin, or any combination thereof. Accordingly, the machine learning model 340-0 can be a neural network trained to process input vector 360-0 and generate output data 372-0 indicating whether the biological associated with the input vector 360-0 is likely to be from an origin also associated with input vector 360-0. In addition, the machine learning model 340-1 can be a random forest classification algorithm trained to process input vector 360-1, which for purposes of this example is the same as input vector 360-0, and generate output data 372-1 indicating whether the biological sample associated with the input vector 360-1 is likely to be from an origin also associated with the input vector 360-1. This method of input vector analysis can continue for each of the x inputs, x input vectors, and x machine learning models. Continuing with this example with reference to **FIG. 1G** the machine learning model 340-x can be a K-nearest neighbor algorithm trained to process input vector 360-x, which for purposes of this example is the same as input vector 360-0 and 360-1, and generate output data 372-x indicating whether the subject associated with the input vector 360-x is likely to be responsive or non-responsive to the treatment also associated with the input vector 360-x.

Alternatively, each of the machine learning models 340-0, 340-1, 340-x can be the same type of machine learning models or different type of machine learning models that are each configured to

receive different inputs. For example, the input to the first machine learning model 340-0 can include a vector 360-0 that includes data representing a first subset or first panel of biomarkers from a biological sample and then predict, based on the machine learning models 340-0 processing of vector 360-0 whether the sample is more or less likely to be from a number of origins. In addition, in this example, an input to the second machine learning model 340-1 can include a vector 360-1 that includes data representing a second subset or second panel of biomarkers from the biological sample that is different than the first subset or first panel of biomarkers. Then, the second machine learning model can generate second output data 372-1 that is indicative of whether the sample associated with the input vector 360-1 is likely to be responsive or likely to be of an origin associated with the input vector 360-2. This method of input vector analysis can continue for each of the x inputs, x input vectors, and x machine learning models. The input to the xth machine learning model 340-x can include a vector 360-x that includes data representing an xth subset or xth panel of biomarkers of a subject that is different than (i) at least one, (i) two or more, or (iii) each of the other x-1 input data vectors 340-0 to 340-x-1. In some implementations, at least one of the x input data vectors can include data representing a complete set of biomarkers from the sample, e.g., next generation sequencing data. Then, the xth machine learning model 340-x can generate second output data 372-x, the second output data 372-x being indicative of whether the sample associated with the input vector 360-x is likely of an origin associated with the input vector 360-x.

Multiple implementations of system 400 described above are not intended to be limiting, and instead, are merely examples of configurations of the multiple machine learning models 340-0, 340-1, 340-x, and their respective inputs, that can be employed using the present disclosure. With reference to these examples, the subject can be any human, non-human animal, plant, or other subject such as described herein. As described above, the input feature vectors can be generated, based on the input data, and represent the input data. Accordingly, each input vector can represent data that includes one or more biomarkers, a disease or disorder, a sample type, an origin, patient data, an origin of a sample having the biomarkers.

In the implementation of **FIG. 1G**, the output data 372-0, 372-1, 372-x can be analyzed using a voting unit 480. For example, the output data 372-0, 372-1, 372-x can be input into the vote unit 480. In some implementations, the output data 372-0, 372-1, 372-x can be data indicating whether the biological sample associated with the input vector processed by the machine learning model is likely to be from a certain origin associated with the vector processed by the machine learning model. Data indicating whether the sample associated with the input vector, and generated by each machine learning model, can include a "0" or a "1." A "0," produced by a machine learning model 340-0 based on the machine learning model's 340-0 processing of an input vector 360-0, can indicate that the sample associated with the input vector 360-0 is not likely to be from an origin associated with input vector 360-0. Similarity, as "1," produced by a machine learning model 360-0 based on the machine learning model's 370-0 processing of an input vector 360-0, can indicate that the sample

associated with the input vector 360-0 is likely to be of an origin associated with the input vector 360-0. Though the example uses "0" as not likely and "1" as likely, the present disclosure is not so limited. Instead, any value can be generated as output data to represent the output classes. For example, in some implementations "1" can be used to represent the "not likely" class and "0" to represent the "likely" class. In yet other implementations, the output data 372-0, 372-1, 372-x can include probabilities that indicate a likelihood that the sample associated with an input vector processed by a machine learning model is associated with a given origin (e.g., a given organ). In such implementations, for example, the generated probability can be applied to a threshold, and if the threshold is satisfied, then the subject associated with an input vector processed by the machine learning model can be determined to be likely to be of that origin.

In some implementations, the machine learning models output an indication whether the sample is more likely to be from one origin versus another, instead of or in addition to indicating that the sample is more of less likely to be from a certain origin. For example, the machine learning model may indicate that the sample is more or less likely to be of prostatic origin (i.e., from the prostate), or the machine learning module may indicate whether the sample is most likely derived from the prostate or from the colon. Any such origins can be so compared.

The voting unit 480 can evaluate the received output data 370-0, 372-1, 372-x and determine whether the sample associated with the processed input vectors 360-0, 360-1, 360-x is likely to be of an origin associated with the processed input vectors 360-0, 360-1, 360-x. The voting unit 480 can then determine, based on the set of received output data 370-0, 372-1, 372-x, whether the sample associated with input vectors 360-0, 360-1, 360-x is likely to be from an origin associated with the input vectors 360-0, 360-2, 360-x. In some implementations, the voting unit 480 can apply a "majority rule." Applying a majority rule, the voting unit 480 can tally the outputs 372-0, 372-1, and 372-x indicating that the sample is from a given origin and outputs 372-0, 372-1, 372-x indicating that the sample is not from that origin (or is from a different origin as described above). Then, the class – e.g., from origin A or not from origin A, or from origin A and not from origin B, etc – having the majority predictions or votes is selected as the appropriate classification for the subject associated with the input vector 360-0, 360-1, 360-x. For example, the majority may determine that the sample is from origin A or is not from origin A, or alternately the majority may determine that the sample is from origin A or is from origin B.

In some implementations, the voting unit 480 can complete a more nuanced analysis. For example, in some implementations, the voting unit 480 can store a confidence score for each machine learning model 340-0, 340-1, 340-x. This confidence score, for each machine learning model 340-0, 340-1, 340-x, can be initially set to a default value such as 0, 1, or the like. Then, with each round of processing of input vectors, the voting unit 480, or other module of the application server 240, can adjust the confidence score for the machine learning model 340-0, 340-1, 340-x based on whether the machine learning model accurately predicted the sample classification selected by the voting unit 480

during a previous iteration. Accordingly, the stored confidence score, for each machine learning model, can provide an indication of the historical accuracy for each machine learning model.

In the more nuanced approached, the voting unit 480 can adjust output data 372-0, 372-0, 372-x produced by each machine learning model 340-0, 340-1, 340-x, respectively, based on the confidence score calculated for the machine learning model. Accordingly, a confidence score indicating that a machine learning mode is historically accurate can be used to boost a value of output data generated by the machine learning model. Similarly, a confidence score indicating that a machine learning model is historically inaccurate can be used to reduce a value of output data generated by the machine learning model. Such boosting or reducing of the value of output data generated by a machine learning model can be achieved, for example, by using the confidence score as a multiplier of less than one for reduction and more than 1 for boosting. Other operations can also be used to adjust the value of output data such as subtracting a confidence score from the value of the output data to reduce the value of the output data or adding the confidence score to the value of the output data to boost the value of the output data. Use of confidence scores to boost or reduce the value of output data generated by the machine learning models is particularly useful when the machine learning models are configured to output probabilities that will be applied to one or more thresholds to determine whether a sample is or is not from an origin, or is from one of two possible origins. This is because using the confidence score to adjust the output of a machine learning model can be used to move a generated output value above or below a class threshold, thereby altering a prediction by a machine learning model based on its historical accuracy.

Use of the voting unit 480 to evaluate outputs of multiple machine learning models can lead to greater accuracy in prediction of the origin of a sample for a particular set of subject biomarkers, as the consensus amongst multiple machine learning models can be evaluated instead of the output of only a single machine learning model.

FIG. 1H is a block diagram of system components that can be used to implement systems of **FIGs. 1B**, **1C**, **1G**, **1F**, and **1G**.

Computing device 600 is intended to represent various forms of digital computers, such as laptops, desktops, workstations, personal digital assistants, servers, blade servers, mainframes, and other appropriate computers. Computing device 650 is intended to represent various forms of mobile devices, such as personal digital assistants, cellular telephones, smartphones, and other similar computing devices. Additionally, computing device 600 or 650 can include Universal Serial Bus (USB) flash drives. The USB flash drives can store operating systems and other applications. The USB flash drives can include input/output components, such as a wireless transmitter or USB connector that can be inserted into a USB port of another computing device. The components shown here, their connections and relationships, and their functions, are meant to be exemplary only, and are not meant to limit implementations of the inventions described and/or claimed in this document.

Computing device 600 includes a processor 602, memory 604, a storage device 608, a high-speed interface 608 connecting to memory 604 and high-speed expansion ports 610, and a low speed interface 612 connecting to low speed bus 614 and storage device 608. Each of the components 602, 604, 608, 608, 610, and 612, are interconnected using various busses, and can be mounted on a common motherboard or in other manners as appropriate. The processor 602 can process instructions for execution within the computing device 600, including instructions stored in the memory 604 or on the storage device 608 to display graphical information for a GUI on an external input/output device, such as display 616 coupled to high speed interface 608. In other implementations, multiple processors and/or multiple buses can be used, as appropriate, along with multiple memories and types of memory. Also, multiple computing devices 600 can be connected, with each device providing portions of the necessary operations, e.g., as a server bank, a group of blade servers, or a multi-processor system.

The memory 604 stores information within the computing device 600. In one implementation, the memory 604 is a volatile memory unit or units. In another implementation, the memory 604 is a non-volatile memory unit or units. The memory 604 can also be another form of computer-readable medium, such as a magnetic or optical disk.

The storage device 608 is capable of providing mass storage for the computing device 600. In one implementation, the storage device 608 can be or contain a computer-readable medium, such as a floppy disk device, a hard disk device, an optical disk device, or a tape device, a flash memory or other similar solid state memory device, or an array of devices, including devices in a storage area network or other configurations. A computer program product can be tangibly embodied in an information carrier. The computer program product can also contain instructions that, when executed, perform one or more methods, such as those described above. The information carrier is a computer- or machine-readable medium, such as the memory 604, the storage device 608, or memory on processor 602.

The high speed controller 608 manages bandwidth-intensive operations for the computing device 600, while the low speed controller 612 manages lower bandwidth intensive operations. Such allocation of functions is exemplary only. In one implementation, the high-speed controller 608 is coupled to memory 604, display 616, e.g., through a graphics processor or accelerator, and to high-speed expansion ports 610, which can accept various expansion cards (not shown). In the implementation, low-speed controller 612 is coupled to storage device 608 and low-speed expansion port 614. The low-speed expansion port, which can include various communication ports, e.g., USB, Bluetooth, Ethernet, wireless Ethernet can be coupled to one or more input/output devices, such as a keyboard, a pointing device, microphone/speaker pair, a scanner, or a networking device such as a switch or router, e.g., through a network adapter. The computing device 600 can be implemented in a number of different forms, as shown in the figure. For example, it can be implemented as a standard server 620, or multiple times in a group of such servers. It can also be implemented as part of a rack

server system 624. In addition, it can be implemented in a personal computer such as a laptop computer 622. Alternatively, components from computing device 600 can be combined with other components in a mobile device (not shown), such as device 650. Each of such devices can contain one or more of computing device 600, 650, and an entire system can be made up of multiple computing devices 600, 650 communicating with each other.

The computing device 600 can be implemented in a number of different forms, as shown in the figure. For example, it can be implemented as a standard server 620, or multiple times in a group of such servers. It can also be implemented as part of a rack server system 624. In addition, it can be implemented in a personal computer such as a laptop computer 622. Alternatively, components from computing device 600 can be combined with other components in a mobile device (not shown), such as device 650. Each of such devices can contain one or more of computing device 600, 650, and an entire system can be made up of multiple computing devices 600, 650 communicating with each other.

Computing device 650 includes a processor 652, memory 664, and an input/output device such as a display 654, a communication interface 666, and a transceiver 668, among other components. The device 650 can also be provided with a storage device, such as a micro-drive or other device, to provide additional storage. Each of the components 650, 652, 664, 654, 666, and 668, are interconnected using various buses, and several of the components can be mounted on a common motherboard or in other manners as appropriate.

The processor 652 can execute instructions within the computing device 650, including instructions stored in the memory 664. The processor can be implemented as a chipset of chips that include separate and multiple analog and digital processors. Additionally, the processor can be implemented using any of a number of architectures. For example, the processor 610 can be a CISC (Complex Instruction Set Computers) processor, a RISC (Reduced Instruction Set Computer) processor, or a MISC (Minimal Instruction Set Computer) processor. The processor can provide, for example, for coordination of the other components of the device 650, such as control of user interfaces, applications run by device 650, and wireless communication by device 650.

Processor 652 can communicate with a user through control interface 658 and display interface 656 coupled to a display 654. The display 654 can be, for example, a TFT (Thin-Film-Transistor Liquid Crystal Display) display or an OLED (Organic Light Emitting Diode) display, or other appropriate display technology. The display interface 656 can comprise appropriate circuitry for driving the display 654 to present graphical and other information to a user. The control interface 658 can receive commands from a user and convert them for submission to the processor 652. In addition, an external interface 662 can be provide in communication with processor 652, so as to enable near area communication of device 650 with other devices. External interface 662 can provide, for example, for wired communication in some implementations, or for wireless communication in other implementations, and multiple interfaces can also be used.

The memory 664 stores information within the computing device 650. The memory 664 can be implemented as one or more of a computer-readable medium or media, a volatile memory unit or units, or a non-volatile memory unit or units. Expansion memory 674 can also be provided and connected to device 650 through expansion interface 672, which can include, for example, a SIMM (Single In Line Memory Module) card interface. Such expansion memory 674 can provide extra storage space for device 650, or can also store applications or other information for device 650. Specifically, expansion memory 674 can include instructions to carry out or supplement the processes described above, and can include secure information also. Thus, for example, expansion memory 674 can be provide as a security module for device 650, and can be programmed with instructions that permit secure use of device 650. In addition, secure applications can be provided via the SIMM cards, along with additional information, such as placing identifying information on the SIMM card in a non-hackable manner.

The memory can include, for example, flash memory and/or NVRAM memory, as discussed below. In one implementation, a computer program product is tangibly embodied in an information carrier. The computer program product contains instructions that, when executed, perform one or more methods, such as those described above. The information carrier is a computer- or machine-readable medium, such as the memory 664, expansion memory 674, or memory on processor 652 that can be received, for example, over transceiver 668 or external interface 662.

Device 650 can communicate wirelessly through communication interface 666, which can include digital signal processing circuitry where necessary. Communication interface 666 can provide for communications under various modes or protocols, such as GSM voice calls, SMS, EMS, or MMS messaging, CDMA, TDMA, PDC, WCDMA, CDMA2000, or GPRS, among others. Such communication can occur, for example, through radio-frequency transceiver 668. In addition, short-range communication can occur, such as using a Bluetooth, Wi-Fi, or other such transceiver (not shown). In addition, GPS (Global Positioning System) receiver module 670 can provide additional navigation- and location-related wireless data to device 650, which can be used as appropriate by applications running on device 650.

Device 650 can also communicate audibly using audio codec 660, which can receive spoken information from a user and convert it to usable digital information. Audio codec 660 can likewise generate audible sound for a user, such as through a speaker, e.g., in a handset of device 650. Such sound can include sound from voice telephone calls, can include recorded sound, e.g., voice messages, music files, etc. and can also include sound generated by applications operating on device 650.

The computing device 650 can be implemented in a number of different forms, as shown in the figure. For example, it can be implemented as a cellular telephone 680. It can also be implemented as part of a smartphone 682, personal digital assistant, or other similar mobile device.

Various implementations of the systems and methods described here can be realized in digital electronic circuitry, integrated circuitry, specially designed ASICs (application specific integrated circuits), computer hardware, firmware, software, and/or combinations of such implementations. These various implementations can include implementation in one or more computer programs that are executable and/or interpretable on a programmable system including at least one programmable processor, which can be special or general purpose, coupled to receive data and instructions from, and to transmit data and instructions to, a storage system, at least one input device, and at least one output device.

These computer programs (also known as programs, software, software applications or code) include machine instructions for a programmable processor, and can be implemented in a high-level procedural and/or object-oriented programming language, and/or in assembly/machine language. As used herein, the terms "machine-readable medium" or "computer-readable medium" refers to any computer program product, apparatus and/or device, e.g., magnetic discs, optical disks, memory, Programmable Logic Devices (PLDs), used to provide machine instructions and/or data to a programmable processor, including a machine-readable medium that receives machine instructions as a machine-readable signal. The term "machine-readable signal" refers to any signal used to provide machine instructions and/or data to a programmable processor.

To provide for interaction with a user, the systems and techniques described here can be implemented on a computer having a display device, e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any form, including acoustic, speech, or tactile input.

The systems and techniques described here can be implemented in a computing system that includes a back end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front end component, e.g., a client computer having a graphical user interface or a Web browser through which a user can interact with an implementation of the systems and techniques described here, or any combination of such back end, middleware, or front end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network ("LAN"), a wide area network ("WAN"), and the Internet.

The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

Computer Systems

The practice of the present methods may also employ computer related software and systems. Computer software products as described herein typically include computer readable medium having computer-executable instructions for performing the logic steps of the method as described herein. Suitable computer readable medium include floppy disk, CD-ROM/DVD/DVD-ROM, hard-disk drive, flash memory, ROM/RAM, magnetic tapes and etc. The computer executable instructions may be written in a suitable computer language or combination of several languages. Basic computational biology methods are described in, for example Setubal and Meidanis et al., Introduction to Computational Biology Methods (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (Ed.), Computational Methods in Molecular Biology, (Elsevier, Amsterdam, 1998); Rashidi and Buehler, Bioinformatics Basics: Application in Biological Science and Medicine (CRC Press, London, 2000) and Ouelette and Bzevanis Bioinformatics: A Practical Guide for Analysis of Gene and Proteins (Wiley & Sons, Inc., 2.sup.nd ed., 2001). See U.S. Pat. No. 6,420,108.

The present methods may also make use of various computer program products and software for a variety of purposes, such as probe design, management of data, analysis, and instrument operation. See, U.S. Pat. Nos. 5,593,839, 5,795,716, 5,733,729, 5,974,164, 6,066,454, 6,090,555, 6,185,561, 6,188,783, 6,223,127, 6,229,911 and 6,308,170.

Additionally, the present methods relates to embodiments that include methods for providing genetic information over networks such as the Internet as shown in U.S. Ser. Nos. 10/197,621, 10/063,559 (U.S. Publication Number 20020183936), 10/065,856, 10/065,868, 10/328,818, 10/328,872, 10/423,403, and 60/482,389. For example, one or more molecular profiling techniques can be performed in one location, e.g., a city, state, country or continent, and the results can be transmitted to a different city, state, country or continent. Treatment selection can then be made in whole or in part in the second location. The methods as described herein comprise transmittal of information between different locations.

Conventional data networking, application development and other functional aspects of the systems (and components of the individual operating components of the systems) may not be described in detail herein but are part as described herein. Furthermore, the connecting lines shown in the various figures contained herein are intended to represent illustrative functional relationships and/or physical couplings between the various elements. It should be noted that many alternative or additional functional relationships or physical connections may be present in a practical system.

The various system components discussed herein may include one or more of the following: a host server or other computing systems including a processor for processing digital data; a memory coupled to the processor for storing digital data; an input digitizer coupled to the processor for inputting digital data; an application program stored in the memory and accessible by the processor for directing processing of digital data by the processor; a display device coupled to the processor and memory for displaying information derived from digital data processed by the processor; and a

plurality of databases. Various databases used herein may include: patient data such as family history, demography and environmental data, biological sample data, prior treatment and protocol data, patient clinical data, molecular profiling data of biological samples, data on therapeutic drug agents and/or investigative drugs, a gene library, a disease library, a drug library, patient tracking data, file management data, financial management data, billing data and/or like data useful in the operation of the system. As those skilled in the art will appreciate, user computer may include an operating system (e.g., Windows NT, 95/98/2000, OS2, UNIX, Linux, Solaris, MacOS, etc.) as well as various conventional support software and drivers typically associated with computers. The computer may include any suitable personal computer, network computer, workstation, minicomputer, mainframe or the like. User computer can be in a home or medical/business environment with access to a network. In an illustrative embodiment, access is through a network or the Internet through a commercially-available web-browser software package.

As used herein, the term "network" shall include any electronic communications means which incorporates both hardware and software components of such. Communication among the parties may be accomplished through any suitable communication channels, such as, for example, a telephone network, an extranet, an intranet, Internet, point of interaction device, personal digital assistant (e.g., Palm Pilot®, Blackberry®), cellular phone, kiosk, etc.), online communications, satellite communications, off-line communications, wireless communications, transponder communications, local area network (LAN), wide area network (WAN), networked or linked devices, keyboard, mouse and/or any suitable communication or data input modality. Moreover, although the system is frequently described herein as being implemented with TCP/IP communications protocols, the system may also be implemented using IPX, Appletalk, IP-6, NetBIOS, OSI or any number of existing or future protocols. If the network is in the nature of a public network, such as the Internet, it may be advantageous to presume the network to be insecure and open to eavesdroppers. Specific information related to the protocols, standards, and application software used in connection with the Internet is generally known to those skilled in the art and, as such, need not be detailed herein. See, for example, Dilip Naik, Internet Standards and Protocols (1998); Java 2 Complete, various authors, (Sybex 1999); Deborah Ray and Eric Ray, Mastering HTML 4.0 (1997); and Loshin, TCP/IP Clearly Explained (1997) and David Gourley and Brian Totty, HTTP, The Definitive Guide (2002), the contents of which are hereby incorporated by reference.

The various system components may be independently, separately or collectively suitably coupled to the network via data links which includes, for example, a connection to an Internet Service Provider (ISP) over the local loop as is typically used in connection with standard modem communication, cable modem, Dish networks, ISDN, Digital Subscriber Line (DSL), or various wireless communication methods, see, e.g., Gilbert Held, Understanding Data Communications (1996), which is hereby incorporated by reference. It is noted that the network may be implemented as other types of networks, such as an interactive television (ITV) network. Moreover, the system

contemplates the use, sale or distribution of any goods, services or information over any network having similar functionality described herein.

As used herein, "transmit" may include sending electronic data from one system component to another over a network connection. Additionally, as used herein, "data" may include encompassing information such as commands, queries, files, data for storage, and the like in digital or any other form.

The system contemplates uses in association with web services, utility computing, pervasive and individualized computing, security and identity solutions, autonomic computing, commodity computing, mobility and wireless solutions, open source, biometrics, grid computing and/or mesh computing.

Any databases discussed herein may include relational, hierarchical, graphical, or object-oriented structure and/or any other database configurations. Common database products that may be used to implement the databases include DB2 by IBM (White Plains, NY), various database products available from Oracle Corporation (Redwood Shores, CA), Microsoft Access or Microsoft SQL Server by Microsoft Corporation (Redmond, Washington), or any other suitable database product. Moreover, the databases may be organized in any suitable manner, for example, as data tables or lookup tables. Each record may be a single file, a series of files, a linked series of data fields or any other data structure. Association of certain data may be accomplished through any desired data association technique such as those known or practiced in the art. For example, the association may be accomplished either manually or automatically. Automatic association techniques may include, for example, a database search, a database merge, GREP, AGREP, SQL, using a key field in the tables to speed searches, sequential searches through all the tables and files, sorting records in the file according to a known order to simplify lookup, and/or the like. The association step may be accomplished by a database merge function, for example, using a "key field" in pre-selected databases or data sectors.

More particularly, a "key field" partitions the database according to the high-level class of objects defined by the key field. For example, certain types of data may be designated as a key field in a plurality of related data tables and the data tables may then be linked on the basis of the type of data in the key field. The data corresponding to the key field in each of the linked data tables is preferably the same or of the same type. However, data tables having similar, though not identical, data in the key fields may also be linked by using AGREP, for example. In accordance with one embodiment, any suitable data storage technique may be used to store data without a standard format. Data sets may be stored using any suitable technique, including, for example, storing individual files using an ISO/IEC 7816-4 file structure; implementing a domain whereby a dedicated file is selected that exposes one or more elementary files containing one or more data sets; using data sets stored in individual files using a hierarchical filing system; data sets stored as records in a single file (including compression, SQL accessible, hashed vione or more keys, numeric, alphabetical by first tuple, etc.); Binary Large Object

(BLOB); stored as ungrouped data elements encoded using ISO/IEC 7816-6 data elements; stored as ungrouped data elements encoded using ISO/IEC Abstract Syntax Notation (ASN.1) as in ISO/IEC 8824 and 8825; and/or other proprietary techniques that may include fractal compression methods, image compression methods, etc.

In one illustrative embodiment, the ability to store a wide variety of information in different formats is facilitated by storing the information as a BLOB. Thus, any binary information can be stored in a storage space associated with a data set. The BLOB method may store data sets as ungrouped data elements formatted as a block of binary via a fixed memory offset using either fixed storage allocation, circular queue techniques, or best practices with respect to memory management (*e.g.*, paged memory, least recently used, etc.). By using BLOB methods, the ability to store various data sets that have different formats facilitates the storage of data by multiple and unrelated owners of the data sets. For example, a first data set which may be stored may be provided by a first party, a second data set which may be stored may be provided by an unrelated second party, and yet a third data set which may be stored, may be provided by a third party unrelated to the first and second party. Each of these three illustrative data sets may contain different information that is stored using different data storage formats and/or techniques. Further, each data set may contain subsets of data that also may be distinct from other subsets.

As stated above, in various embodiments, the data can be stored without regard to a common format. However, in one illustrative embodiment, the data set (*e.g.*, BLOB) may be annotated in a standard manner when provided for manipulating the data. The annotation may comprise a short header, trailer, or other appropriate indicator related to each data set that is configured to convey information useful in managing the various data sets. For example, the annotation may be called a "condition header", "header", "trailer", or "status", herein, and may comprise an indication of the status of the data set or may include an identifier correlated to a specific issuer or owner of the data. Subsequent bytes of data may be used to indicate for example, the identity of the issuer or owner of the data, user, transaction/membership account identifier or the like. Each of these condition annotations are further discussed herein.

The data set annotation may also be used for other types of status information as well as various other purposes. For example, the data set annotation may include security information establishing access levels. The access levels may, for example, be configured to permit only certain individuals, levels of employees, companies, or other entities to access data sets, or to permit access to specific data sets based on the transaction, issuer or owner of data, user or the like. Furthermore, the security information may restrict/permit only certain actions such as accessing, modifying, and/or deleting data sets. In one example, the data set annotation indicates that only the data set owner or the user are permitted to delete a data set, various identified users may be permitted to access the data set for reading, and others are altogether excluded from accessing the data set. However, other access restriction parameters may also be used allowing various entities to access a data set with various

permission levels as appropriate. The data, including the header or trailer may be received by a standalone interaction device configured to add, delete, modify, or augment the data in accordance with the header or trailer.

One skilled in the art will also appreciate that, for security reasons, any databases, systems, devices, servers or other components of the system may consist of any combination thereof at a single location or at multiple locations, wherein each database or system includes any of various suitable security features, such as firewalls, access codes, encryption, decryption, compression, decompression, and/or the like.

The computing unit of the web client may be further equipped with an Internet browser connected to the Internet or an intranet using standard dial-up, cable, DSL or any other Internet protocol known in the art. Transactions originating at a web client may pass through a firewall in order to prevent unauthorized access from users of other networks. Further, additional firewalls may be deployed between the varying components of CMS to further enhance security.

Firewall may include any hardware and/or software suitably configured to protect CMS components and/or enterprise computing resources from users of other networks. Further, a firewall may be configured to limit or restrict access to various systems and components behind the firewall for web clients connecting through a web server. Firewall may reside in varying configurations including Stateful Inspection, Proxy based and Packet Filtering among others. Firewall may be integrated within an web server or any other CMS components or may further reside as a separate entity.

The computers discussed herein may provide a suitable website or other Internet-based graphical user interface which is accessible by users. In one embodiment, the Microsoft Internet Information Server (IIS), Microsoft Transaction Server (MTS), and Microsoft SQL Server, are used in conjunction with the Microsoft operating system, Microsoft NT web server software, a Microsoft SQL Server database system, and a Microsoft Commerce Server. Additionally, components such as Access or Microsoft SQL Server, Oracle, Sybase, Informix MySQL, Interbase, etc., may be used to provide an Active Data Object (ADO) compliant database management system.

Any of the communications, inputs, storage, databases or displays discussed herein may be facilitated through a website having web pages. The term "web page" as it is used herein is not meant to limit the type of documents and applications that might be used to interact with the user. For example, a typical website might include, in addition to standard HTML documents, various forms, Java applets, JavaScript, active server pages (ASP), common gateway interface scripts (CGI), extensible markup language (XML), dynamic HTML, cascading style sheets (CSS), helper applications, plug-ins, and the like. A server may include a web service that receives a request from a web server, the request including a URL (http://yahoo.com/stockquotes/ge) and an IP address (123.56.789.234). The web server retrieves the appropriate web pages and sends the data or applications for the web pages to the IP address. Web services are applications that are capable of

interacting with other applications over a communications means, such as the internet. Web services are typically based on standards or protocols such as XML, XSLT, SOAP, WSDL and UDDI. Web services methods are well known in the art, and are covered in many standard texts. See, e.g., Alex Nghiem, IT Web Services: A Roadmap for the Enterprise (2003), hereby incorporated by reference.

The web-based clinical database for the system and method of the present methods preferably has the ability to upload and store clinical data files in native formats and is searchable on any clinical parameter. The database is also scalable and may use an EAV data model (metadata) to enter clinical annotations from any study for easy integration with other studies. In addition, the web-based clinical database is flexible and may be XML and XSLT enabled to be able to add user customized questions dynamically. Further, the database includes exportability to CDISC ODM.

Practitioners will also appreciate that there are a number of methods for displaying data within a browser-based document. Data may be represented as standard text or within a fixed list, scrollable list, drop-down list, editable text field, fixed text field, pop-up window, and the like. Likewise, there are a number of methods available for modifying data in a web page such as, for example, free text entry using a keyboard, selection of menu items, check boxes, option boxes, and the like.

The system and method may be described herein in terms of functional block components, screen shots, optional selections and various processing steps. It should be appreciated that such functional blocks may be realized by any number of hardware and/or software components configured to perform the specified functions. For example, the system may employ various integrated circuit components, e.g., memory elements, processing elements, logic elements, look-up tables, and the like, which may carry out a variety of functions under the control of one or more microprocessors or other control devices. Similarly, the software elements of the system may be implemented with any programming or scripting language such as C, C++, Macromedia Cold Fusion, Microsoft Active Server Pages, Java, COBOL, assembler, PERL, Visual Basic, SQL Stored Procedures, extensible markup language (XML), with the various algorithms being implemented with any combination of data structures, objects, processes, routines or other programming elements. Further, it should be noted that the system may employ any number of conventional techniques for data transmission, signaling, data processing, network control, and the like. Still further, the system could be used to detect or prevent security issues with a client-side scripting language, such as JavaScript, VBScript or the like. For a basic introduction of cryptography and network security, see any of the following references: (1) "Applied Cryptography: Protocols, Algorithms, And Source Code In C," by Bruce Schneier, published by John Wiley & Sons (second edition, 1995); (2) "Java Cryptography" by Jonathan Knudson, published by O'Reilly & Associates (1998); (3) "Cryptography & Network Security: Principles & Practice" by William Stallings, published by Prentice Hall; all of which are hereby incorporated by reference.

As used herein, the term "end user", "consumer", "customer", "client", "treating physician", "hospital", or "business" may be used interchangeably with each other, and each shall mean any person, entity, machine, hardware, software or business. Each participant is equipped with a computing device in order to interact with the system and facilitate online data access and data input. The customer has a computing unit in the form of a personal computer, although other types of computing units may be used including laptops, notebooks, hand held computers, set-top boxes, cellular telephones, touch-tone telephones and the like. The owner/operator of the system and method of the present methods has a computing unit implemented in the form of a computer-server, although other implementations are contemplated by the system including a computing center shown as a main frame computer, a mini-computer, a PC server, a network of computers located in the same of different geographic locations, or the like. Moreover, the system contemplates the use, sale or distribution of any goods, services or information over any network having similar functionality described herein.

In one illustrative embodiment, each client customer may be issued an "account" or "account number". As used herein, the account or account number may include any device, code, number, letter, symbol, digital certificate, smart chip, digital signal, analog signal, biometric or other identifier/indicia suitably configured to allow the consumer to access, interact with or communicate with the system (e.g., one or more of an authorization/access code, personal identification number (PIN), Internet code, other identification code, and/or the like). The account number may optionally be located on or associated with a charge card, credit card, debit card, prepaid card, embossed card, smart card, magnetic stripe card, bar code card, transponder, radio frequency card or an associated account. The system may include or interface with any of the foregoing cards or devices, or a fob having a transponder and RFID reader in RF communication with the fob. Although the system may include a fob embodiment, the methods is not to be so limited. Indeed, system may include any device having a transponder which is configured to communicate with RFID reader via RF communication. Typical devices may include, for example, a key ring, tag, card, cell phone, wristwatch or any such form capable of being presented for interrogation. Moreover, the system, computing unit or device discussed herein may include a "pervasive computing device," which may include a traditionally non-computerized device that is embedded with a computing unit. The account number may be distributed and stored in any form of plastic, electronic, magnetic, radio frequency, wireless, audio and/or optical device capable of transmitting or downloading data from itself to a second device.

As will be appreciated by one of ordinary skill in the art, the system may be embodied as a customization of an existing system, an add-on product, upgraded software, a standalone system, a distributed system, a method, a data processing system, a device for data processing, and/or a computer program product. Accordingly, the system may take the form of an entirely software embodiment, an entirely hardware embodiment, or an embodiment combining aspects of both software and hardware. Furthermore, the system may take the form of a computer program product on

a computer-readable storage medium having computer-readable program code means embodied in the storage medium. Any suitable computer-readable storage medium may be used, including hard disks, CD-ROM, optical storage devices, magnetic storage devices, and/or the like.

The system and method is described herein with reference to screen shots, block diagrams and flowchart illustrations of methods, apparatus (*e.g.*, systems), and computer program products according to various embodiments. It will be understood that each functional block of the block diagrams and the flowchart illustrations, and combinations of functional blocks in the block diagrams and flowchart illustrations, respectively, can be implemented by computer program instructions.

These computer program instructions may be loaded onto a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions that execute on the computer or other programmable data processing apparatus create means for implementing the functions specified in the flowchart block or blocks. These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including instruction means which implement the function specified in the flowchart block or blocks. The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions which execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

Accordingly, functional blocks of the block diagrams and flowchart illustrations support combinations of means for performing the specified functions, combinations of steps for performing the specified functions, and program instruction means for performing the specified functions. It will also be understood that each functional block of the block diagrams and flowchart illustrations, and combinations of functional blocks in the block diagrams and flowchart illustrations, can be implemented by either special purpose hardware-based computer systems which perform the specified functions or steps, or suitable combinations of special purpose hardware and computer instructions. Further, illustrations of the process flows and the descriptions thereof may make reference to user windows, web pages, websites, web forms, prompts, etc. Practitioners will appreciate that the illustrated steps described herein may comprise in any number of configurations including the use of windows, web pages, web forms, popup windows, prompts and the like. It should be further appreciated that the multiple steps as illustrated and described may be combined into single web pages and/or windows but have been expanded for the sake of simplicity. In other cases, steps illustrated and described as single process steps may be separated into multiple web pages and/or windows but have been combined for simplicity.

Molecular Profiling

The molecular profiling approach provides a method for selecting a candidate treatment for an individual that could favorably change the clinical course for the individual with a condition or disease, such as cancer. The molecular profiling approach provides clinical benefit for individuals, such as identifying therapeutic regimens that provide a longer progression free survival (PFS), longer disease free survival (DFS), longer overall survival (OS) or extended lifespan. Methods and systems as described herein are directed to molecular profiling of cancer on an individual basis that can identify optimal therapeutic regimens. Molecular profiling provides a personalized approach to selecting candidate treatments that are likely to benefit a cancer. The molecular profiling methods described herein can be used to guide treatment in any desired setting, including without limitation the front-line / standard of care setting, or for patients with poor prognosis, such as those with metastatic disease or those whose cancer has progressed on standard front line therapies, or whose cancer has progressed on previous chemotherapeutic or hormonal regimens.

The systems and methods of the invention may be used to classify patients as more or less likely to benefit or respond to various treatments. Unless otherwise noted, the terms "response" or "non-response," as used herein, refer to any appropriate indication that a treatment provides a benefit to a patient (a "responder" or "benefiter") or has a lack of benefit to the patient (a "non-responder" or "non-benefiter"). Such an indication may be determined using accepted clinical response criteria such as the standard Response Evaluation Criteria in Solid Tumors (RECIST) criteria, or any other useful patient response criteria such as progression free survival (PFS), time to progression (TTP), disease free survival (DFS), time-to-next treatment (TNT, TTNT), time-to-treatment failure (TTF, TTTF), tumor shrinkage or disappearance, or the like. RECIST is a set of rules published by an international consortium that define when tumors improve ("respond"), stay the same ("stabilize"), or worsen ("progress") during treatment of a cancer patient. As used herein and unless otherwise noted, a patient "benefit" from a treatment may refer to any appropriate measure of improvement, including without limitation a RECIST response or longer PFS/TTP/DFS/TNT/TTNT, whereas "lack of benefit" from a treatment may refer to any appropriate measure of worsening disease during treatment. Generally disease stabilization is considered a benefit, although in certain circumstances, if so noted herein, stabilization may be considered a lack of benefit. A predicted or indicated benefit may be described as "indeterminate" if there is not an acceptable level of prediction of benefit or lack of benefit. In some cases, benefit is considered indeterminate if it cannot be calculated, e.g., due to lack of necessary data.

Personalized medicine based on pharmacogenetic insights, such as those provided by molecular profiling as described herein, is increasingly taken for granted by some practitioners and the lay press, but forms the basis of hope for improved cancer therapy. However, molecular profiling as taught herein represents a fundamental departure from the traditional approach to oncologic therapy where for the most part, patients are grouped together and treated with approaches that are based on findings from light microscopy and disease stage. Traditionally, differential response to a particular

therapeutic strategy has only been determined after the treatment was given, i.e., *a posteriori*. The "standard" approach to disease treatment relies on what is generally true about a given cancer diagnosis and treatment response has been vetted by randomized phase III clinical trials and forms the "standard of care" in medical practice. The results of these trials have been codified in consensus statements by guidelines organizations such as the National Comprehensive Cancer Network and The American Society of Clinical Oncology. The NCCN Compendium™ contains authoritative, scientifically derived information designed to support decision-making about the appropriate use of drugs and biologics in patients with cancer. The NCCN Compendium™ is recognized by the Centers for Medicare and Medicaid Services (CMS) and United Healthcare as an authoritative reference for oncology coverage policy. On-compendium treatments are those recommended by such guides. The biostatistical methods used to validate the results of clinical trials rely on minimizing differences between patients, and are based on declaring the likelihood of error that one approach is better than another for a patient group defined only by light microscopy and stage, not by individual differences in tumors. The molecular profiling methods described herein exploit such individual differences. The methods can provide candidate treatments that can be then selected by a physician for treating a patient.

Molecular profiling can be used to provide a comprehensive view of the biological state of a sample. In an embodiment, molecular profiling is used for whole tumor profiling. Accordingly, a number of molecular approaches are used to assess the state of a tumor. The whole tumor profiling can be used for selecting a candidate treatment for a tumor. Molecular profiling can be used to select candidate therapeutics on any sample for any stage of a disease. In embodiment, the methods as described herein are used to profile a newly diagnosed cancer. The candidate treatments indicated by the molecular profiling can be used to select a therapy for treating the newly diagnosed cancer. In other embodiments, the methods as described herein are used to profile a cancer that has already been treated, e.g., with one or more standard-of-care therapy. In embodiments, the cancer is refractory to the prior treatment/s. For example, the cancer may be refractory to the standard of care treatments for the cancer. The cancer can be a metastatic cancer or other recurrent cancer. The treatments can be on-compendium or off-compendium treatments.

Molecular profiling can be performed by any known means for detecting a molecule in a biological sample. Molecular profiling comprises methods that include but are not limited to, nucleic acid sequencing, such as a DNA sequencing or RNA sequencing; immunohistochemistry (IHC); in situ hybridization (ISH); fluorescent in situ hybridization (FISH); chromogenic in situ hybridization (CISH); PCR amplification (e.g., qPCR or RT-PCR); various types of microarray (mRNA expression arrays, low density arrays, protein arrays, etc); various types of sequencing (Sanger, pyrosequencing, etc); comparative genomic hybridization (CGH); high throughput or next generation sequencing (NGS); Northern blot; Southern blot; immunoassay; and any other appropriate technique to assay the presence or quantity of a biological molecule of interest. In various embodiments, any one or more of

these methods can be used concurrently or subsequent to each other for assessing target genes disclosed herein.

Molecular profiling of individual samples is used to select one or more candidate treatments for a disorder in a subject, e.g., by identifying targets for drugs that may be effective for a given cancer. For example, the candidate treatment can be a treatment known to have an effect on cells that differentially express genes as identified by molecular profiling techniques, an experimental drug, a government or regulatory approved drug or any combination of such drugs, which may have been studied and approved for a particular indication that is the same as or different from the indication of the subject from whom a biological sample is obtain and molecularly profiled.

When multiple biomarker targets are revealed by assessing target genes by molecular profiling, one or more decision rules can be put in place to prioritize the selection of certain therapeutic agent for treatment of an individual on a personalized basis. Rules as described herein aide prioritizing treatment, e.g., direct results of molecular profiling, anticipated efficacy of therapeutic agent, prior history with the same or other treatments, expected side effects, availability of therapeutic agent, cost of therapeutic agent, drug-drug interactions, and other factors considered by a treating physician. Based on the recommended and prioritized therapeutic agent targets, a physician can decide on the course of treatment for a particular individual. Accordingly, molecular profiling methods and systems as described herein can select candidate treatments based on individual characteristics of diseased cells, e.g., tumor cells, and other personalized factors in a subject in need of treatment, as opposed to relying on a traditional one-size fits all approach that is conventionally used to treat individuals suffering from a disease, especially cancer. In some cases, the recommended treatments are those not typically used to treat the disease or disorder inflicting the subject. In some cases, the recommended treatments are used after standard-of-care therapies are no longer providing adequate efficacy.

The treating physician can use the results of the molecular profiling methods to optimize a treatment regimen for a patient. The candidate treatment identified by the methods as described herein can be used to treat a patient; however, such treatment is not required of the methods. Indeed, the analysis of molecular profiling results and identification of candidate treatments based on those results can be automated and does not require physician involvement.

Biological Entities

Nucleic acids include deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, or complements thereof. Nucleic acids can contain known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl

phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs). Nucleic acid sequence can encompass conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); Rossolini et al., Mol. Cell Probes 8:91-98 (1994)). The term nucleic acid can be used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

A particular nucleic acid sequence may implicitly encompass the particular sequence and "splice variants" and nucleic acid sequences encoding truncated forms. Similarly, a particular protein encoded by a nucleic acid can encompass any protein encoded by a splice variant or truncated form of that nucleic acid. "Splice variants," as the name suggests, are products of alternative splicing of a gene. After transcription, an initial nucleic acid transcript may be spliced such that different (alternate) nucleic acid splice products encode different polypeptides. Mechanisms for the production of splice variants vary, but include alternate splicing of exons. Alternate polypeptides derived from the same nucleic acid by read-through transcription are also encompassed by this definition. Any products of a splicing reaction, including recombinant forms of the splice products, are included in this definition. Nucleic acids can be truncated at the 5' end or at the 3' end. Polypeptides can be truncated at the N-terminal end or the C-terminal end. Truncated versions of nucleic acid or polypeptide sequences can be naturally occurring or created using recombinant techniques.

The terms "genetic variant" and "nucleotide variant" are used herein interchangeably to refer to changes or alterations to the reference human gene or cDNA sequence at a particular locus, including, but not limited to, nucleotide base deletions, insertions, inversions, and substitutions in the coding and non-coding regions. Deletions may be of a single nucleotide base, a portion or a region of the nucleotide sequence of the gene, or of the entire gene sequence. Insertions may be of one or more nucleotide bases. The genetic variant or nucleotide variant may occur in transcriptional regulatory regions, untranslated regions of mRNA, exons, introns, exon/intron junctions, etc. The genetic variant or nucleotide variant can potentially result in stop codons, frame shifts, deletions of amino acids, altered gene transcript splice forms or altered amino acid sequence.

An allele or gene allele comprises generally a naturally occurring gene having a reference sequence or a gene containing a specific nucleotide variant.

A haplotype refers to a combination of genetic (nucleotide) variants in a region of an mRNA or a genomic DNA on a chromosome found in an individual. Thus, a haplotype includes a number of genetically linked polymorphic variants which are typically inherited together as a unit.

As used herein, the term "amino acid variant" is used to refer to an amino acid change to a reference human protein sequence resulting from genetic variants or nucleotide variants to the reference human gene encoding the reference protein. The term "amino acid variant" is intended to

encompass not only single amino acid substitutions, but also amino acid deletions, insertions, and other significant changes of amino acid sequence in the reference protein.

The term "genotype" as used herein means the nucleotide characters at a particular nucleotide variant marker (or locus) in either one allele or both alleles of a gene (or a particular chromosome region). With respect to a particular nucleotide position of a gene of interest, the nucleotide(s) at that locus or equivalent thereof in one or both alleles form the genotype of the gene at that locus. A genotype can be homozygous or heterozygous. Accordingly, "genotyping" means determining the genotype, that is, the nucleotide(s) at a particular gene locus. Genotyping can also be done by determining the amino acid variant at a particular position of a protein which can be used to deduce the corresponding nucleotide variant(s).

The term "locus" refers to a specific position or site in a gene sequence or protein. Thus, there may be one or more contiguous nucleotides in a particular gene locus, or one or more amino acids at a particular locus in a polypeptide. Moreover, a locus may refer to a particular position in a gene where one or more nucleotides have been deleted, inserted, or inverted.

Unless specified otherwise or understood by one of skill in art, the terms "polypeptide," "protein," and "peptide" are used interchangeably herein to refer to an amino acid chain in which the amino acid residues are linked by covalent peptide bonds. The amino acid chain can be of any length of at least two amino acids, including full-length proteins. Unless otherwise specified, polypeptide, protein, and peptide also encompass various modified forms thereof, including but not limited to glycosylated forms, phosphorylated forms, etc. A polypeptide, protein or peptide can also be referred to as a gene product.

Lists of gene and gene products that can be assayed by molecular profiling techniques are presented herein. Lists of genes may be presented in the context of molecular profiling techniques that detect a gene product (e.g., an mRNA or protein). One of skill will understand that this implies detection of the gene product of the listed genes. Similarly, lists of gene products may be presented in the context of molecular profiling techniques that detect a gene sequence or copy number. One of skill will understand that this implies detection of the gene corresponding to the gene products, including as an example DNA encoding the gene products. As will be appreciated by those skilled in the art, a "biomarker" or "marker" comprises a gene and/or gene product depending on the context.

The terms "label" and "detectable label" can refer to any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, chemical or similar methods. Such labels include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DYNABEADS™), fluorescent dyes (e.g., fluorescein, Texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and calorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc) beads. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350;

3,996,345; 4,277,437; 4,275,149; and 4,366,241. Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and calorimetric labels are detected by simply visualizing the colored label. Labels can include, e.g., ligands that bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden Introduction to Immunocytochemistry, 2nd ed., Springer Verlag, NY (1997); and in Haugland Handbook of Fluorescent Probes and Research Chemicals, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

Detectable labels include, but are not limited to, nucleotides (labeled or unlabelled), compomers, sugars, peptides, proteins, antibodies, chemical compounds, conducting polymers, binding moieties such as biotin, mass tags, calorimetric agents, light emitting agents, chemiluminescent agents, light scattering agents, fluorescent tags, radioactive tags, charge tags (electrical or magnetic charge), volatile tags and hydrophobic tags, biomolecules (e.g., members of a binding pair antibody/antigen, antibody/antibody, antibody/antibody fragment, antibody/antibody receptor, antibody/protein A or protein G, hapten/anti-hapten, biotin/avidin, biotin/streptavidin, folic acid/folate binding protein, vitamin B12/intrinsic factor, chemical reactive group/complementary chemical reactive group (e.g., sulfhydryl/maleimide, sulfhydryl/haloacetyl derivative, amine/isotriocyanate, amine/succinimidyl ester, and amine/sulfonyl halides) and the like.

The terms "primer", "probe," and "oligonucleotide" are used herein interchangeably to refer to a relatively short nucleic acid fragment or sequence. They can comprise DNA, RNA, or a hybrid thereof, or chemically modified analog or derivatives thereof. Typically, they are single-stranded. However, they can also be double-stranded having two complementing strands which can be separated by denaturation. Normally, primers, probes and oligonucleotides have a length of from about 8 nucleotides to about 200 nucleotides, preferably from about 12 nucleotides to about 100 nucleotides, and more preferably about 18 to about 50 nucleotides. They can be labeled with detectable markers or modified using conventional manners for various molecular biological applications.

The term "isolated" when used in reference to nucleic acids (e.g., genomic DNAs, cDNAs, mRNAs, or fragments thereof) is intended to mean that a nucleic acid molecule is present in a form that is substantially separated from other naturally occurring nucleic acids that are normally associated with the molecule. Because a naturally existing chromosome (or a viral equivalent thereof) includes a long nucleic acid sequence, an isolated nucleic acid can be a nucleic acid molecule having only a portion of the nucleic acid sequence in the chromosome but not one or more other portions present on the same chromosome. More specifically, an isolated nucleic acid can include naturally occurring

nucleic acid sequences that flank the nucleic acid in the naturally existing chromosome (or a viral equivalent thereof). An isolated nucleic acid can be substantially separated from other naturally occurring nucleic acids that are on a different chromosome of the same organism. An isolated nucleic acid can also be a composition in which the specified nucleic acid molecule is significantly enriched so as to constitute at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or at least 99% of the total nucleic acids in the composition.

An isolated nucleic acid can be a hybrid nucleic acid having the specified nucleic acid molecule covalently linked to one or more nucleic acid molecules that are not the nucleic acids naturally flanking the specified nucleic acid. For example, an isolated nucleic acid can be in a vector. In addition, the specified nucleic acid may have a nucleotide sequence that is identical to a naturally occurring nucleic acid or a modified form or mutein thereof having one or more mutations such as nucleotide substitution, deletion/insertion, inversion, and the like.

An isolated nucleic acid can be prepared from a recombinant host cell (in which the nucleic acids have been recombinantly amplified and/or expressed), or can be a chemically synthesized nucleic acid having a naturally occurring nucleotide sequence or an artificially modified form thereof.

The term "high stringency hybridization conditions," when used in connection with nucleic acid hybridization, includes hybridization conducted overnight at 42 °C in a solution containing 50% formamide, 5×SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate, pH 7.6, 5×Denhardt's solution, 10% dextran sulfate, and 20 microgram/ml denatured and sheared salmon sperm DNA, with hybridization filters washed in 0.1×SSC at about 65 °C. The term "moderate stringent hybridization conditions," when used in connection with nucleic acid hybridization, includes hybridization conducted overnight at 37 °C in a solution containing 50% formamide, 5×SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate, pH 7.6, 5×Denhardt's solution, 10% dextran sulfate, and 20 microgram/ml denatured and sheared salmon sperm DNA, with hybridization filters washed in 1×SSC at about 50 °C. It is noted that many other hybridization methods, solutions and temperatures can be used to achieve comparable stringent hybridization conditions as will be apparent to skilled artisans.

For the purpose of comparing two different nucleic acid or polypeptide sequences, one sequence (test sequence) may be described to be a specific percentage identical to another sequence (comparison sequence). The percentage identity can be determined by the algorithm of Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993), which is incorporated into various BLAST programs. The percentage identity can be determined by the "BLAST 2 Sequences" tool, which is available at the National Center for Biotechnology Information (NCBI) website. See Tatusova and Madden, FEMS Microbiol. Lett., 174(2):247-250 (1999). For pairwise DNA-DNA comparison, the BLASTN program is used with default parameters (e.g., Match: 1; Mismatch: -2; Open gap: 5 penalties; extension gap: 2 penalties; gap x_dropoff: 50; expect: 10; and word size: 11, with filter). For pairwise protein-protein sequence comparison, the BLASTP program can be

employed using default parameters (e.g., Matrix: BLOSUM62; gap open: 11; gap extension: 1; x_dropoff: 15; expect: 10.0; and wordsize: 3, with filter). Percent identity of two sequences is calculated by aligning a test sequence with a comparison sequence using BLAST, determining the number of amino acids or nucleotides in the aligned test sequence that are identical to amino acids or nucleotides in the same position of the comparison sequence, and dividing the number of identical amino acids or nucleotides by the number of amino acids or nucleotides in the comparison sequence. When BLAST is used to compare two sequences, it aligns the sequences and yields the percent identity over defined, aligned regions. If the two sequences are aligned across their entire length, the percent identity yielded by the BLAST is the percent identity of the two sequences. If BLAST does not align the two sequences over their entire length, then the number of identical amino acids or nucleotides in the unaligned regions of the test sequence and comparison sequence is considered to be zero and the percent identity is calculated by adding the number of identical amino acids or nucleotides in the aligned regions and dividing that number by the length of the comparison sequence. Various versions of the BLAST programs can be used to compare sequences, e.g., BLAST 2.1.2 or BLAST+ 2.2.22.

A subject or individual can be any animal which may benefit from the methods described herein, including, e.g., humans and non-human mammals, such as primates, rodents, horses, dogs and cats. Subjects include without limitation a eukaryotic organisms, most preferably a mammal such as a primate, e.g., chimpanzee or human, cow; dog; cat; a rodent, e.g., guinea pig, rat, mouse; rabbit; or a bird; reptile; or fish. Subjects specifically intended for treatment using the methods described herein include humans. A subject may also be referred to herein as an individual or a patient. In the present methods the subject has colorectal cancer, e.g., has been diagnosed with colorectal cancer. Methods for identifying subjects with colorectal cancer are known in the art, e.g., using a biopsy. See, e.g., Fleming et al., J Gastrointest Oncol. 2012 Sep; 3(3): 153–173; Chang et al., Dis Colon Rectum. 2012; 55(8):831-43.

Treatment of a disease or individual according to the methods described herein is an approach for obtaining beneficial or desired medical results, including clinical results, but not necessarily a cure. For purposes of the methods described herein, beneficial or desired clinical results include, but are not limited to, alleviation or amelioration of one or more symptoms, diminishment of extent of disease, stabilized (i.e., not worsening) state of disease, preventing spread of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, and remission (whether partial or total), whether detectable or undetectable. Treatment also includes prolonging survival as compared to expected survival if not receiving treatment or if receiving a different treatment. A treatment can include administration of various small molecule drugs or biologics such as immunotherapies, e.g., checkpoint inhibitor therapies. A biomarker refers generally to a molecule, including without limitation a gene or product thereof, nucleic acids (e.g., DNA, RNA), protein/peptide/polypeptide, carbohydrate structure, lipid, glycolipid, characteristics of which can be detected in a tissue or cell to

provide information that is predictive, diagnostic, prognostic and/or theranostic for sensitivity or resistance to candidate treatment.

Biological Samples

A sample as used herein includes any relevant biological sample that can be used for molecular profiling, e.g., sections of tissues such as biopsy or tissue removed during surgical or other procedures, bodily fluids, autopsy samples, and frozen sections taken for histological purposes. Such samples include blood and blood fractions or products (e.g., serum, buffy coat, plasma, platelets, red blood cells, and the like), sputum, malignant effusion, cheek cells tissue, cultured cells (e.g., primary cultures, explants, and transformed cells), stool, urine, other biological or bodily fluids (e.g., prostatic fluid, gastric fluid, intestinal fluid, renal fluid, lung fluid, cerebrospinal fluid, and the like), etc. The sample can comprise biological material that is a fresh frozen & formalin fixed paraffin embedded (FFPE) block, formalin-fixed paraffin embedded, or is within an RNA preservative + formalin fixative. More than one sample of more than one type can be used for each patient. In a preferred embodiment, the sample comprises a fixed tumor sample.

The sample used in the systems and methods of the invention can be a formalin fixed paraffin embedded (FFPE) sample. The FFPE sample can be one or more of fixed tissue, unstained slides, bone marrow core or clot, core needle biopsy, malignant fluids and fine needle aspirate (FNA). In an embodiment, the fixed tissue comprises a tumor containing formalin fixed paraffin embedded (FFPE) block from a surgery or biopsy. In another embodiment, the unstained slides comprise unstained, charged, unbaked slides from a paraffin block. In another embodiment, bone marrow core or clot comprises a decalcified core. A formalin fixed core and/or clot can be paraffin-embedded. In still another embodiment, the core needle biopsy comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more, e.g., 3-4, paraffin embedded biopsy samples. An 18 gauge needle biopsy can be used. The malignant fluid can comprise a sufficient volume of fresh pleural/ascitic fluid to produce a 5x5x2mm cell pellet. The fluid can be formalin fixed in a paraffin block. In an embodiment, the core needle biopsy comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more, e.g., 4-6, paraffin embedded aspirates.

A sample may be processed according to techniques understood by those in the art. A sample can be without limitation fresh, frozen or fixed cells or tissue. In some embodiments, a sample comprises formalin-fixed paraffin-embedded (FFPE) tissue, fresh tissue or fresh frozen (FF) tissue. A sample can comprise cultured cells, including primary or immortalized cell lines derived from a subject sample. A sample can also refer to an extract from a sample from a subject. For example, a sample can comprise DNA, RNA or protein extracted from a tissue or a bodily fluid. Many techniques and commercial kits are available for such purposes. The fresh sample from the individual can be treated with an agent to preserve RNA prior to further processing, e.g., cell lysis and extraction. Samples can include frozen samples collected for other purposes. Samples can be associated with relevant information such as age, gender, and clinical symptoms present in the subject; source of the

sample; and methods of collection and storage of the sample. A sample is typically obtained from a subject.

A biopsy comprises the process of removing a tissue sample for diagnostic or prognostic evaluation, and to the tissue specimen itself. Any biopsy technique known in the art can be applied to the molecular profiling methods of the present disclosure. The biopsy technique applied can depend on the tissue type to be evaluated (e.g., colon, prostate, kidney, bladder, lymph node, liver, bone marrow, blood cell, lung, breast, etc.), the size and type of the tumor (e.g., solid or suspended, blood or ascites), among other factors. Representative biopsy techniques include, but are not limited to, excisional biopsy, incisional biopsy, needle biopsy, surgical biopsy, and bone marrow biopsy. An "excisional biopsy" refers to the removal of an entire tumor mass with a small margin of normal tissue surrounding it. An "incisional biopsy" refers to the removal of a wedge of tissue that includes a cross-sectional diameter of the tumor. Molecular profiling can use a "core-needle biopsy" of the tumor mass, or a "fine-needle aspiration biopsy" which generally obtains a suspension of cells from within the tumor mass. Biopsy techniques are discussed, for example, in Harrison's Principles of Internal Medicine, Kasper, et al., eds., 16th ed., 2005, Chapter 70, and throughout Part V.

Unless otherwise noted, a "sample" as referred to herein for molecular profiling of a patient may comprise more than one physical specimen. As one non-limiting example, a "sample" may comprise multiple sections from a tumor, e.g., multiple sections of an FFPE block or multiple core-needle biopsy sections. As another non-limiting example, a "sample" may comprise multiple biopsy specimens, e.g., one or more surgical biopsy specimen, one or more core-needle biopsy specimen, one or more fine-needle aspiration biopsy specimen, or any useful combination thereof. As still another non-limiting example, a molecular profile may be generated for a subject using a "sample" comprising a solid tumor specimen and a bodily fluid specimen. In some embodiments, a sample is a unitary sample, i.e., a single physical specimen.

Standard molecular biology techniques known in the art and not specifically described are generally followed as in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York (1989), and as in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1989) and as in Perbal, *A Practical Guide to Molecular Cloning*, John Wiley & Sons, New York (1988), and as in Watson et al., *Recombinant DNA, Scientific American Books*, New York and in Birren et al (eds) *Genome Analysis: A Laboratory Manual Series, Vols. 1-4* Cold Spring Harbor Laboratory Press, New York (1998) and methodology as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057 and incorporated herein by reference. Polymerase chain reaction (PCR) can be carried out generally as in PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego, Calif. (1990).

Vesicles

The sample can comprise vesicles. Methods as described herein can include assessing one or more vesicles, including assessing vesicle populations. A vesicle, as used herein, is a membrane

vesicle that is shed from cells. Vesicles or membrane vesicles include without limitation: circulating microvesicles (cMVs), microvesicle, exosome, nanovesicle, dexosome, bleb, blebby, prostasome, microparticle, intralumenal vesicle, membrane fragment, intralumenal endosomal vesicle, endosomal-like vesicle, exocytosis vehicle, endosome vesicle, endosomal vesicle, apoptotic body, multivesicular body, secretory vesicle, phospholipid vesicle, liposomal vesicle, argosome, texasome, secresome, tolerosome, melanosome, oncosome, or exocytosed vehicle. Furthermore, although vesicles may be produced by different cellular processes, the methods as described herein are not limited to or reliant on any one mechanism, insofar as such vesicles are present in a biological sample and are capable of being characterized by the methods disclosed herein. Unless otherwise specified, methods that make use of a species of vesicle can be applied to other types of vesicles. Vesicles comprise spherical structures with a lipid bilayer similar to cell membranes which surrounds an inner compartment which can contain soluble components, sometimes referred to as the payload. In some embodiments, the methods as described herein make use of exosomes, which are small secreted vesicles of about 40–100 nm in diameter. For a review of membrane vesicles, including types and characterizations, see *Thery et al., Nat Rev Immunol. 2009 Aug;9(8):581-93*. Some properties of different types of vesicles include those in **Table 1**:

Table 1: Vesicle Properties

Feature	**Exosomes**	**Micro-vesicles**	**Ectosomes**	**Mem-brane particles**	**Exosome-like vesicles**	**Apoptotic vesicles**
Size	50-100 nm	100-1,000 nm	50-200 nm	50-80 nm	20-50 nm	50-500 nm
Density in sucrose	1.13-1.19 g/ml			1.04-1.07 g/ml	1.1 g/ml	1.16-1.28 g/ml
EM appearance	Cup shape	Irregular shape, electron dense	Bilamellar round structures	Round	Irregular shape	Hetero-geneous
Sedimen-tation	100,000 g	10,000 g	160,000-200,000 g	100,000-200,000 g	175,000 g	1,200 g, 10,000 g, 100,000 g
Lipid com-position	Enriched in cholesterol, sphingomyelin and ceramide; contains lipid rafts; expose PPS	Expose PPS	Enriched in cholesterol and diacylglycerol; expose PPS		No lipid rafts	
Major protein markers	Tetraspanins (e.g., CD63, CD9), Alix, TSG101	Integrins, selectins and CD40 ligand	CR1 and proteolytic enzymes; no CD63	CD133; no CD63	TNFRI	Histones
Intra-cellular origin	Internal compartments (endosomes)	Plasma membrane	Plasma membrane	Plasma membrane		

Abbreviations: phosphatidylserine (PPS); electron microscopy (EM)

Vesicles include shed membrane bound particles, or "microparticles," that are derived from either the plasma membrane or an internal membrane. Vesicles can be released into the extracellular environment from cells. Cells releasing vesicles include without limitation cells that originate from, or are derived from, the ectoderm, endoderm, or mesoderm. The cells may have undergone genetic, environmental, and/or any other variations or alterations. For example, the cell can be tumor cells. A vesicle can reflect any changes in the source cell, and thereby reflect changes in the originating cells, e.g., cells having various genetic mutations. In one mechanism, a vesicle is generated intracellularly when a segment of the cell membrane spontaneously invaginates and is ultimately exocytosed (see for example, *Keller et al., Immunol. Lett. 107 (2): 102–8 (2006)*). Vesicles also include cell-derived structures bounded by a lipid bilayer membrane arising from both herniated evagination (blebbing) separation and sealing of portions of the plasma membrane or from the export of any intracellular membrane-bounded vesicular structure containing various membrane-associated proteins of tumor origin, including surface-bound molecules derived from the host circulation that bind selectively to the tumor-derived proteins together with molecules contained in the vesicle lumen, including but not limited to tumor-derived microRNAs or intracellular proteins. Blebs and blebbing are further described in *Charras et al., Nature Reviews Molecular and Cell Biology, Vol. 9, No. 11, p. 730-736 (2008)*. A vesicle shed into circulation or bodily fluids from tumor cells may be referred to as a "circulating tumor-derived vesicle." When such vesicle is an exosome, it may be referred to as a circulating-tumor derived exosome (CTE). In some instances, a vesicle can be derived from a specific cell of origin. CTE, as with a cell-of-origin specific vesicle, typically have one or more unique biomarkers that permit isolation of the CTE or cell-of-origin specific vesicle, e.g., from a bodily fluid and sometimes in a specific manner. For example, a cell or tissue specific markers are used to identify the cell of origin. Examples of such cell or tissue specific markers are disclosed herein and can further be accessed in the Tissue-specific Gene Expression and Regulation (TiGER) Database, available at bioinfo.wilmer.jhu.edu/tiger/; Liu et al. (2008) TiGER: a database for tissue-specific gene expression and regulation. BMC Bioinformatics. 9:271; TissueDistributionDBs, available at genome.dkfz-heidelberg.de/menu/tissue_db/index.html.

A vesicle can have a diameter of greater than about 10 nm, 20 nm, or 30 nm. A vesicle can have a diameter of greater than 40 nm, 50 nm, 100 nm, 200 nm, 500 nm, 1000 nm or greater than 10,000 nm. A vesicle can have a diameter of about 30-1000 nm, about 30-800 nm, about 30-200 nm, or about 30-100 nm. In some embodiments, the vesicle has a diameter of less than 10,000 nm, 1000 nm, 800 nm, 500 nm, 200 nm, 100 nm, 50 nm, 40 nm, 30 nm, 20 nm or less than 10 nm. As used herein the term "about" in reference to a numerical value means that variations of 10% above or below the numerical value are within the range ascribed to the specified value. Typical sizes for various types of vesicles are shown in **Table 1**. Vesicles can be assessed to measure the diameter of a single vesicle or any number of vesicles. For example, the range of diameters of a vesicle population or an average diameter of a vesicle population can be determined. Vesicle diameter can be assessed

using methods known in the art, e.g., imaging technologies such as electron microscopy. In an embodiment, a diameter of one or more vesicles is determined using optical particle detection. See, e.g., U.S. Patent 7,751,053, entitled "Optical Detection and Analysis of Particles" and issued July 6, 2010; and U.S. Patent 7,399,600, entitled "Optical Detection and Analysis of Particles" and issued July 15, 2010.

In some embodiments, vesicles are directly assayed from a biological sample without prior isolation, purification, or concentration from the biological sample. For example, the amount of vesicles in the sample can by itself provide a biosignature that provides a diagnostic, prognostic or theranostic determination. Alternatively, the vesicle in the sample may be isolated, captured, purified, or concentrated from a sample prior to analysis. As noted, isolation, capture or purification as used herein comprises partial isolation, partial capture or partial purification apart from other components in the sample. Vesicle isolation can be performed using various techniques as described herein or known in the art, including without limitation size exclusion chromatography, density gradient centrifugation, differential centrifugation, nanomembrane ultrafiltration, immunoabsorbent capture, affinity purification, affinity capture, immunoassay, immunoprecipitation, microfluidic separation, flow cytometry or combinations thereof.

Vesicles can be assessed to provide a phenotypic characterization by comparing vesicle characteristics to a reference. In some embodiments, surface antigens on a vesicle are assessed. A vesicle or vesicle population carrying a specific marker can be referred to as a positive (biomarker+) vesicle or vesicle population. For example, a DLL4+ population refers to a vesicle population associated with DLL4. Conversely, a DLL4- population would not be associated with DLL4. The surface antigens can provide an indication of the anatomical origin and/or cellular of the vesicles and other phenotypic information, e.g., tumor status. For example, vesicles found in a patient sample can be assessed for surface antigens indicative of colorectal origin and the presence of cancer, thereby identifying vesicles associated with colorectal cancer cells. The surface antigens may comprise any informative biological entity that can be detected on the vesicle membrane surface, including without limitation surface proteins, lipids, carbohydrates, and other membrane components. For example, positive detection of colon derived vesicles expressing tumor antigens can indicate that the patient has colorectal cancer. As such, methods as described herein can be used to characterize any disease or condition associated with an anatomical or cellular origin, by assessing, for example, disease-specific and cell-specific biomarkers of one or more vesicles obtained from a subject.

In embodiments, one or more vesicle payloads are assessed to provide a phenotypic characterization. The payload with a vesicle comprises any informative biological entity that can be detected as encapsulated within the vesicle, including without limitation proteins and nucleic acids, e.g., genomic or cDNA, mRNA, or functional fragments thereof, as well as microRNAs (miRs). In addition, methods as described herein are directed to detecting vesicle surface antigens (in addition or exclusive to vesicle payload) to provide a phenotypic characterization. For example, vesicles can be

characterized by using binding agents (e.g., antibodies or aptamers) that are specific to vesicle surface antigens, and the bound vesicles can be further assessed to identify one or more payload components disclosed therein. As described herein, the levels of vesicles with surface antigens of interest or with payload of interest can be compared to a reference to characterize a phenotype. For example, overexpression in a sample of cancer-related surface antigens or vesicle payload, e.g., a tumor associated mRNA or microRNA, as compared to a reference, can indicate the presence of cancer in the sample. The biomarkers assessed can be present or absent, increased or reduced based on the selection of the desired target sample and comparison of the target sample to the desired reference sample. Non-limiting examples of target samples include: disease; treated/not-treated; different time points, such as a in a longitudinal study; and non-limiting examples of reference sample: non-disease; normal; different time points; and sensitive or resistant to candidate treatment(s).

In an embodiment, molecular profiling as described herein comprises analysis of microvesicles, such as circulating microvesicles.

MicroRNA

Various biomarker molecules can be assessed in biological samples or vesicles obtained from such biological samples. MicroRNAs comprise one class biomarkers assessed via methods as described herein. MicroRNAs, also referred to herein as miRNAs or miRs, are short RNA strands approximately 21-23 nucleotides in length. MiRNAs are encoded by genes that are transcribed from DNA but are not translated into protein and thus comprise non-coding RNA. The miRs are processed from primary transcripts known as pri-miRNA to short stem-loop structures called pre-miRNA and finally to the resulting single strand miRNA. The pre-miRNA typically forms a structure that folds back on itself in self-complementary regions. These structures are then processed by the nuclease Dicer in animals or DCL1 in plants. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules and can function to regulate translation of proteins. Identified sequences of miRNA can be accessed at publicly available databases, such as www.microRNA.org, www.mirbase.org, or www.mirz.unibas.ch/cgi/miRNA.cgi.

miRNAs are generally assigned a number according to the naming convention " mir-[number]." The number of a miRNA is assigned according to its order of discovery relative to previously identified miRNA species. For example, if the last published miRNA was mir-121, the next discovered miRNA will be named mir-122, etc. When a miRNA is discovered that is homologous to a known miRNA from a different organism, the name can be given an optional organism identifier, of the form [organism identifier]- mir-[number]. Identifiers include hsa for Homo sapiens and mmu for Mus Musculus. For example, a human homolog to mir-121 might be referred to as hsa-mir-121 whereas the mouse homolog can be referred to as mmu-mir-121.

Mature microRNA is commonly designated with the prefix "miR" whereas the gene or precursor miRNA is designated with the prefix "mir." For example, mir-121 is a precursor for miR-121. When differing miRNA genes or precursors are processed into identical mature miRNAs, the

genes/precursors can be delineated by a numbered suffix. For example, mir-121-1 and mir-121-2 can refer to distinct genes or precursors that are processed into miR-121. Lettered suffixes are used to indicate closely related mature sequences. For example, mir-121a and mir-121b can be processed to closely related miRNAs miR-121a and miR-121b, respectively. In the context of the present disclosure, any microRNA (miRNA or miR) designated herein with the prefix mir-* or miR-* is understood to encompass both the precursor and/or mature species, unless otherwise explicitly stated otherwise.

Sometimes it is observed that two mature miRNA sequences originate from the same precursor. When one of the sequences is more abundant that the other, a "*" suffix can be used to designate the less common variant. For example, miR-121 would be the predominant product whereas miR-121* is the less common variant found on the opposite arm of the precursor. If the predominant variant is not identified, the miRs can be distinguished by the suffix "5p" for the variant from the 5' arm of the precursor and the suffix "3p" for the variant from the 3' arm. For example, miR-121-5p originates from the 5' arm of the precursor whereas miR-121-3p originates from the 3' arm. Less commonly, the 5p and 3p variants are referred to as the sense ("s") and anti-sense ("as") forms, respectively. For example, miR-121-5p may be referred to as miR-121-s whereas miR-121-3p may be referred to as miR-121-as.

The above naming conventions have evolved over time and are general guidelines rather than absolute rules. For example, the let- and lin- families of miRNAs continue to be referred to by these monikers. The mir/miR convention for precursor/mature forms is also a guideline and context should be taken into account to determine which form is referred to. Further details of miR naming can be found at www.mirbase.org or Ambros et al., A uniform system for microRNA annotation, RNA **9**:277-279 (2003).

Plant miRNAs follow a different naming convention as described in Meyers et al., Plant Cell. 2008 20(12):3186-3190.

A number of miRNAs are involved in gene regulation, and miRNAs are part of a growing class of non-coding RNAs that is now recognized as a major tier of gene control. In some cases, miRNAs can interrupt translation by binding to regulatory sites embedded in the 3′-UTRs of their target mRNAs, leading to the repression of translation. Target recognition involves complementary base pairing of the target site with the miRNA's seed region (positions 2–8 at the miRNA's 5′ end), although the exact extent of seed complementarity is not precisely determined and can be modified by 3′ pairing. In other cases, miRNAs function like small interfering RNAs (siRNA) and bind to perfectly complementary mRNA sequences to destroy the target transcript.

Characterization of a number of miRNAs indicates that they influence a variety of processes, including early development, cell proliferation and cell death, apoptosis and fat metabolism. For example, some miRNAs, such as lin-4, let-7, mir-14, mir-23, and bantam, have been shown to play

critical roles in cell differentiation and tissue development. Others are believed to have similarly important roles because of their differential spatial and temporal expression patterns.

The miRNA database available at miRBase (www.mirbase.org) comprises a searchable database of published miRNA sequences and annotation. Further information about miRBase can be found in the following articles, each of which is incorporated by reference in its entirety herein: Griffiths-Jones et al., miRBase: tools for microRNA genomics. NAR 2008 36(Database Issue):D154-D158; Griffiths-Jones et al., miRBase: microRNA sequences, targets and gene nomenclature. NAR 2006 34(Database Issue):D140-D144; and Griffiths-Jones, S. The microRNA Registry. NAR 2004 32(Database Issue):D109-D111. Representative miRNAs contained in Release 16 of miRBase, made available September 2010.

As described herein, microRNAs are known to be involved in cancer and other diseases and can be assessed in order to characterize a phenotype in a sample. See, e.g., Ferracin et al., Micromarkers: miRNAs in cancer diagnosis and prognosis, Exp Rev Mol Diag, Apr 2010, Vol. 10, No. 3, Pages 297-308; Fabbri, miRNAs as molecular biomarkers of cancer, Exp Rev Mol Diag, May 2010, Vol. 10, No. 4, Pages 435-444.

In an embodiment, molecular profiling as described herein comprises analysis of microRNA.

Techniques to isolate and characterize vesicles and miRs are known to those of skill in the art. In addition to the methodology presented herein, additional methods can be found in U.S. Patent Nos. 7,888,035, entitled "METHODS FOR ASSESSING RNA PATTERNS" and issued February 15, 2011; and 7,897,356, entitled "METHODS AND SYSTEMS OF USING EXOSOMES FOR DETERMINING PHENOTYPES" and issued March 1, 2011; and International Patent Publication Nos. WO/2011/066589, entitled "METHODS AND SYSTEMS FOR ISOLATING, STORING, AND ANALYZING VESICLES" and filed November 30, 2010; WO/2011/088226, entitled "DETECTION OF GASTROINTESTINAL DISORDERS" and filed January 13, 2011; WO/2011/109440, entitled "BIOMARKERS FOR THERANOSTICS" and filed March 1, 2011; and WO/2011/127219, entitled "CIRCULATING BIOMARKERS FOR DISEASE" and filed April 6, 2011, each of which applications are incorporated by reference herein in their entirety.

Circulating Biomarkers

Circulating biomarkers include biomarkers that are detectable in body fluids, such as blood, plasma, serum. Examples of circulating cancer biomarkers include cardiac troponin T (cTnT), prostate specific antigen (PSA) for prostate cancer and CA125 for ovarian cancer. Circulating biomarkers according to the present disclosure include any appropriate biomarker that can be detected in bodily fluid, including without limitation protein, nucleic acids, e.g., DNA, mRNA and microRNA, lipids, carbohydrates and metabolites. Circulating biomarkers can include biomarkers that are not associated with cells, such as biomarkers that are membrane associated, embedded in membrane fragments, part of a biological complex, or free in solution. In one embodiment, circulating biomarkers are biomarkers that are associated with one or more vesicles present in the biological fluid of a subject.

Circulating biomarkers have been identified for use in characterization of various phenotypes, such as detection of a cancer. See, e.g., Ahmed N, et al., Proteomic-based identification of haptoglobin-1 precursor as a novel circulating biomarker of ovarian cancer. Br. J. Cancer 2004; Mathelin _et al., Circulating proteinic biomarkers and breast cancer, Gynecol Obstet Fertil. 2006 Jul-Aug;34(7-8):638-46. Epub 2006 Jul 28; Ye et al., Recent technical strategies to identify diagnostic biomarkers for ovarian cancer. Expert Rev Proteomics. 2007 Feb;4(1):121-31; Carney, Circulating oncoproteins HER2/neu, EGFR and CAIX (MN) as novel cancer biomarkers. Expert Rev Mol Diagn. 2007 May;7(3):309-19; Gagnon, Discovery and application of protein biomarkers for ovarian cancer, Curr Opin Obstet Gynecol. 2008 Feb;20(1):9-13; Pasterkamp et al., Immune regulatory cells: circulating biomarker factories in cardiovascular disease. Clin Sci (Lond). 2008 Aug;115(4):129-31; Fabbri, miRNAs as molecular biomarkers of cancer, Exp Rev Mol Diag, May 2010, Vol. 10, No. 4, Pages 435-444; PCT Patent Publication WO/2007/088537; U.S. Patents 7,745,150 and 7,655,479; U.S. Patent Publications 20110008808, 20100330683, 20100248290, 20100222230, 20100203566, 20100173788, 20090291932, 20090239246, 20090226937, 20090111121, 20090004687, 20080261258, 20080213907, 20060003465, 20050124071, and 20040096915, each of which publication is incorporated herein by reference in its entirety. In an embodiment, molecular profiling as described herein comprises analysis of circulating biomarkers.

Gene Expression Profiling

The methods and systems as described herein comprise expression profiling, which includes assessing differential expression of one or more target genes disclosed herein. Differential expression can include overexpression and/or underexpression of a biological product, e.g., a gene, mRNA or protein, compared to a control (or a reference). The control can include similar cells to the sample but without the disease (e.g., expression profiles obtained from samples from healthy individuals). A control can be a previously determined level that is indicative of a drug target efficacy associated with the particular disease and the particular drug target. The control can be derived from the same patient, e.g., a normal adjacent portion of the same organ as the diseased cells, the control can be derived from healthy tissues from other patients, or previously determined thresholds that are indicative of a disease responding or not-responding to a particular drug target. The control can also be a control found in the same sample, e.g. a housekeeping gene or a product thereof (e.g., mRNA or protein). For example, a control nucleic acid can be one which is known not to differ depending on the cancerous or non-cancerous state of the cell. The expression level of a control nucleic acid can be used to normalize signal levels in the test and reference populations. Illustrative control genes include, but are not limited to, e.g., β-actin, glyceraldehyde 3-phosphate dehydrogenase and ribosomal protein P1. Multiple controls or types of controls can be used. The source of differential expression can vary. For example, a gene copy number may be increased in a cell, thereby resulting in increased expression of the gene. Alternately, transcription of the gene may be modified, e.g., by chromatin remodeling, differential methylation, differential expression or activity of transcription factors, etc. Translation

may also be modified, e.g., by differential expression of factors that degrade mRNA, translate mRNA, or silence translation, e.g., microRNAs or siRNAs. In some embodiments, differential expression comprises differential activity. For example, a protein may carry a mutation that increases the activity of the protein, such as constitutive activation, thereby contributing to a diseased state. Molecular profiling that reveals changes in activity can be used to guide treatment selection.

Methods of gene expression profiling include methods based on hybridization analysis of polynucleotides, and methods based on sequencing of polynucleotides. Commonly used methods known in the art for the quantification of mRNA expression in a sample include northern blotting and in situ hybridization (Parker & Barnes (1999) Methods in Molecular Biology 106:247-283); RNAse protection assays (Hod (1992) Biotechniques 13:852-854); and reverse transcription polymerase chain reaction (RT-PCR) (Weis et al. (1992) Trends in Genetics 8:263-264). Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. Representative methods for sequencing-based gene expression analysis include Serial Analysis of Gene Expression (SAGE), gene expression analysis by massively parallel signature sequencing (MPSS) and/or next generation sequencing.

RT-PCR

Reverse transcription polymerase chain reaction (RT-PCR) is a variant of polymerase chain reaction (PCR). According to this technique, a RNA strand is reverse transcribed into its DNA complement (i.e., complementary DNA, or cDNA) using the enzyme reverse transcriptase, and the resulting cDNA is amplified using PCR. Real-time polymerase chain reaction is another PCR variant, which is also referred to as quantitative PCR, Q-PCR, qRT-PCR, or sometimes as RT-PCR. Either the reverse transcription PCR method or the real-time PCR method can be used for molecular profiling according to the present disclosure, and RT-PCR can refer to either unless otherwise specified or as understood by one of skill in the art.

RT-PCR can be used to determine RNA levels, e.g., mRNA or miRNA levels, of the biomarkers as described herein. RT-PCR can be used to compare such RNA levels of the biomarkers as described herein in different sample populations, in normal and tumor tissues, with or without drug treatment, to characterize patterns of gene expression, to discriminate between closely related RNAs, and to analyze RNA structure.

The first step is the isolation of RNA, e.g., mRNA, from a sample. The starting material can be total RNA isolated from human tumors or tumor cell lines, and corresponding normal tissues or cell lines, respectively. Thus RNA can be isolated from a sample, e.g., tumor cells or tumor cell lines, and compared with pooled DNA from healthy donors. If the source of mRNA is a primary tumor, mRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g. formalin-fixed) tissue samples.

General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al. (1997) Current Protocols of Molecular

Biology, John Wiley and Sons. Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp & Locker (1987) Lab Invest. 56:A67, and De Andres et al., BioTechniques 18:42044 (1995). In particular, RNA isolation can be performed using purification kit, buffer set and protease from commercial manufacturers, such as Qiagen, according to the manufacturer's instructions (QIAGEN Inc., Valencia, CA). For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Numerous RNA isolation kits are commercially available and can be used in the methods as described herein.

In the alternative, the first step is the isolation of miRNA from a target sample. The starting material is typically total RNA isolated from human tumors or tumor cell lines, and corresponding normal tissues or cell lines, respectively. Thus RNA can be isolated from a variety of primary tumors or tumor cell lines, with pooled DNA from healthy donors. If the source of miRNA is a primary tumor, miRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g. formalin-fixed) tissue samples.

General methods for miRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al. (1997) Current Protocols of Molecular Biology, John Wiley and Sons. Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp & Locker (1987) Lab Invest. 56:A67, and De Andres et al., BioTechniques 18:42044 (1995). In particular, RNA isolation can be performed using purification kit, buffer set and protease from commercial manufacturers, such as Qiagen, according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Numerous miRNA isolation kits are commercially available and can be used in the methods as described herein.

Whether the RNA comprises mRNA, miRNA or other types of RNA, gene expression profiling by RT-PCR can include reverse transcription of the RNA template into cDNA, followed by amplification in a PCR reaction. Commonly used reverse transcriptases include, but are not limited to, avilo myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukemia virus reverse transcriptase (MMLV-RT). The reverse transcription step is typically primed using specific primers, random hexamers, or oligo-dT primers, depending on the circumstances and the goal of expression profiling. For example, extracted RNA can be reverse-transcribed using a GeneAmp RNA PCR kit (Perkin Elmer, Calif., USA), following the manufacturer's instructions. The derived cDNA can then be used as a template in the subsequent PCR reaction.

Although the PCR step can use a variety of thermostable DNA-dependent DNA polymerases, it typically employs the Taq DNA polymerase, which has a 5'-3' nuclease activity but lacks a 3'-5' proofreading endonuclease activity. TaqMan PCR typically uses the 5'-nuclease activity of Taq or Tth polymerase to hydrolyze a hybridization probe bound to its target amplicon, but any enzyme with equivalent 5' nuclease activity can be used. Two oligonucleotide primers are used to generate an amplicon typical of a PCR reaction. A third oligonucleotide, or probe, is designed to detect nucleotide

sequence located between the two PCR primers. The probe is non-extendible by Taq DNA polymerase enzyme, and is labeled with a reporter fluorescent dye and a quencher fluorescent dye. Any laser-induced emission from the reporter dye is quenched by the quenching dye when the two dyes are located close together as they are on the probe. During the amplification reaction, the Taq DNA polymerase enzyme cleaves the probe in a template-dependent manner. The resultant probe fragments disassociate in solution, and signal from the released reporter dye is free from the quenching effect of the second fluorophore. One molecule of reporter dye is liberated for each new molecule synthesized, and detection of the unquenched reporter dye provides the basis for quantitative interpretation of the data.

TaqMan™ RT-PCR can be performed using commercially available equipment, such as, for example, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, Calif., USA), or LightCycler (Roche Molecular Biochemicals, Mannheim, Germany). In one specific embodiment, the 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI PRISM 7700 Sequence Detection System. The system consists of a thermocycler, laser, charge-coupled device (CCD), camera and computer. The system amplifies samples in a 96-well format on a thermocycler. During amplification, laser-induced fluorescent signal is collected in real-time through fiber optic cables for all 96 wells, and detected at the CCD. The system includes software for running the instrument and for analyzing the data.

TaqMan data are initially expressed as Ct, or the threshold cycle. As discussed above, fluorescence values are recorded during every cycle and represent the amount of product amplified to that point in the amplification reaction. The point when the fluorescent signal is first recorded as statistically significant is the threshold cycle (Ct).

To minimize errors and the effect of sample-to-sample variation, RT-PCR is usually performed using an internal standard. The ideal internal standard is expressed at a constant level among different tissues, and is unaffected by the experimental treatment. RNAs most frequently used to normalize patterns of gene expression are mRNAs for the housekeeping genes glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) and β-actin.

Real time quantitative PCR (also quantitative real time polymerase chain reaction, QRT-PCR or Q-PCR) is a more recent variation of the RT-PCR technique. Q-PCR can measure PCR product accumulation through a dual-labeled fluorigenic probe (i.e., TaqMan probe). Real time PCR is compatible both with quantitative competitive PCR, where internal competitor for each target sequence is used for normalization, and with quantitative comparative PCR using a normalization gene contained within the sample, or a housekeeping gene for RT-PCR. See, e.g. Held et al. (1996) Genome Research 6:986-994.

Protein-based detection techniques are also useful for molecular profiling, especially when the nucleotide variant causes amino acid substitutions or deletions or insertions or frame shift that affect the protein primary, secondary or tertiary structure. To detect the amino acid variations, protein

sequencing techniques may be used. For example, a protein or fragment thereof corresponding to a gene can be synthesized by recombinant expression using a DNA fragment isolated from an individual to be tested. Preferably, a cDNA fragment of no more than 100 to 150 base pairs encompassing the polymorphic locus to be determined is used. The amino acid sequence of the peptide can then be determined by conventional protein sequencing methods. Alternatively, the HPLC-microscopy tandem mass spectrometry technique can be used for determining the amino acid sequence variations. In this technique, proteolytic digestion is performed on a protein, and the resulting peptide mixture is separated by reversed-phase chromatographic separation. Tandem mass spectrometry is then performed and the data collected is analyzed. See Gatlin et al., Anal. Chem., 72:757-763 (2000).

Microarray

The biomarkers as described herein can also be identified, confirmed, and/or measured using the microarray technique. Thus, the expression profile biomarkers can be measured in cancer samples using microarray technology. In this method, polynucleotide sequences of interest are plated, or arrayed, on a microchip substrate. The arrayed sequences are then hybridized with specific DNA probes from cells or tissues of interest. The source of mRNA can be total RNA isolated from a sample, e.g., human tumors or tumor cell lines and corresponding normal tissues or cell lines. Thus RNA can be isolated from a variety of primary tumors or tumor cell lines. If the source of mRNA is a primary tumor, mRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g. formalin-fixed) tissue samples, which are routinely prepared and preserved in everyday clinical practice.

The expression profile of biomarkers can be measured in either fresh or paraffin-embedded tumor tissue, or body fluids using microarray technology. In this method, polynucleotide sequences of interest are plated, or arrayed, on a microchip substrate. The arrayed sequences are then hybridized with specific DNA probes from cells or tissues of interest. As with the RT-PCR method, the source of miRNA typically is total RNA isolated from human tumors or tumor cell lines, including body fluids, such as serum, urine, tears, and exosomes and corresponding normal tissues or cell lines. Thus RNA can be isolated from a variety of sources. If the source of miRNA is a primary tumor, miRNA can be extracted, for example, from frozen tissue samples, which are routinely prepared and preserved in everyday clinical practice.

Also known as biochip, DNA chip, or gene array, cDNA microarray technology allows for identification of gene expression levels in a biologic sample. cDNAs or oligonucleotides, each representing a given gene, are immobilized on a substrate, e.g., a small chip, bead or nylon membrane, tagged, and serve as probes that will indicate whether they are expressed in biologic samples of interest. The simultaneous expression of thousands of genes can be monitored simultaneously.

In a specific embodiment of the microarray technique, PCR amplified inserts of cDNA clones are applied to a substrate in a dense array. In one aspect, at least 100, 200, 300, 400, 500, 600, 700,

800, 900, 1,000, 1,500, 2,000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 15,000, 20,000, 25,000, 30,000, 35,000, 40,000, 45,000 or at least 50,000 nucleotide sequences are applied to the substrate. Each sequence can correspond to a different gene, or multiple sequences can be arrayed per gene. The microarrayed genes, immobilized on the microchip, are suitable for hybridization under stringent conditions. Fluorescently labeled cDNA probes may be generated through incorporation of fluorescent nucleotides by reverse transcription of RNA extracted from tissues of interest. Labeled cDNA probes applied to the chip hybridize with specificity to each spot of DNA on the array. After stringent washing to remove non-specifically bound probes, the chip is scanned by confocal laser microscopy or by another detection method, such as a CCD camera. Quantitation of hybridization of each arrayed element allows for assessment of corresponding mRNA abundance. With dual color fluorescence, separately labeled cDNA probes generated from two sources of RNA are hybridized pairwise to the array. The relative abundance of the transcripts from the two sources corresponding to each specified gene is thus determined simultaneously. The miniaturized scale of the hybridization affords a convenient and rapid evaluation of the expression pattern for large numbers of genes. Such methods have been shown to have the sensitivity required to detect rare transcripts, which are expressed at a few copies per cell, and to reproducibly detect at least approximately two-fold differences in the expression levels (Schena et al. (1996) Proc. Natl. Acad. Sci. USA 93(2):106-149). Microarray analysis can be performed by commercially available equipment following manufacturer's protocols, including without limitation the Affymetrix GeneChip technology (Affymetrix, Santa Clara, CA), Agilent (Agilent Technologies, Inc., Santa Clara, CA), or Illumina (Illumina, Inc., San Diego, CA) microarray technology.

The development of microarray methods for large-scale analysis of gene expression makes it possible to search systematically for molecular markers of cancer classification and outcome prediction in a variety of tumor types.

In some embodiments, the Agilent Whole Human Genome Microarray Kit (Agilent Technologies, Inc., Santa Clara, CA). The system can analyze more than 41,000 unique human genes and transcripts represented, all with public domain annotations. The system is used according to the manufacturer's instructions.

In some embodiments, the Illumina Whole Genome DASL assay (Illumina Inc., San Diego, CA) is used. The system offers a method to simultaneously profile over 24,000 transcripts from minimal RNA input, from both fresh frozen (FF) and formalin-fixed paraffin embedded (FFPE) tissue sources, in a high throughput fashion.

Microarray expression analysis comprises identifying whether a gene or gene product is up-regulated or down-regulated relative to a reference. The identification can be performed using a statistical test to determine statistical significance of any differential expression observed. In some embodiments, statistical significance is determined using a parametric statistical test. The parametric statistical test can comprise, for example, a fractional factorial design, analysis of variance (ANOVA),

a t-test, least squares, a Pearson correlation, simple linear regression, nonlinear regression, multiple linear regression, or multiple nonlinear regression. Alternatively, the parametric statistical test can comprise a one-way analysis of variance, two-way analysis of variance, or repeated measures analysis of variance. In other embodiments, statistical significance is determined using a nonparametric statistical test. Examples include, but are not limited to, a Wilcoxon signed-rank test, a Mann-Whitney test, a Kruskal-Wallis test, a Friedman test, a Spearman ranked order correlation coefficient, a Kendall Tau analysis, and a nonparametric regression test. In some embodiments, statistical significance is determined at a p-value of less than about 0.05, 0.01, 0.005, 0.001, 0.0005, or 0.0001. Although the microarray systems used in the methods as described herein may assay thousands of transcripts, data analysis need only be performed on the transcripts of interest, thereby reducing the problem of multiple comparisons inherent in performing multiple statistical tests. The p-values can also be corrected for multiple comparisons, e.g., using a Bonferroni correction, a modification thereof, or other technique known to those in the art, e.g., the Hochberg correction, Holm-Bonferroni correction, Šidák correction, or Dunnett's correction. The degree of differential expression can also be taken into account. For example, a gene can be considered as differentially expressed when the fold-change in expression compared to control level is at least 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.5, 2.7, 3.0, 4, 5, 6, 7, 8, 9 or 10-fold different in the sample versus the control. The differential expression takes into account both overexpression and underexpression. A gene or gene product can be considered up or down-regulated if the differential expression meets a statistical threshold, a fold-change threshold, or both. For example, the criteria for identifying differential expression can comprise both a p-value of 0.001 and fold change of at least 1.5-fold (up or down). One of skill will understand that such statistical and threshold measures can be adapted to determine differential expression by any molecular profiling technique disclosed herein.

Various methods as described herein make use of many types of microarrays that detect the presence and potentially the amount of biological entities in a sample. Arrays typically contain addressable moieties that can detect the presence of the entity in the sample, e.g., via a binding event. Microarrays include without limitation DNA microarrays, such as cDNA microarrays, oligonucleotide microarrays and SNP microarrays, microRNA arrays, protein microarrays, antibody microarrays, tissue microarrays, cellular microarrays (also called transfection microarrays), chemical compound microarrays, and carbohydrate arrays (glycoarrays). DNA arrays typically comprise addressable nucleotide sequences that can bind to sequences present in a sample. MicroRNA arrays, e.g., the MMChips array from the University of Louisville or commercial systems from Agilent, can be used to detect microRNAs. Protein microarrays can be used to identify protein–protein interactions, including without limitation identifying substrates of protein kinases, transcription factor protein-activation, or to identify the targets of biologically active small molecules. Protein arrays may comprise an array of different protein molecules, commonly antibodies, or nucleotide sequences that bind to proteins of interest. Antibody microarrays comprise antibodies spotted onto the protein chip that are used as

capture molecules to detect proteins or other biological materials from a sample, e.g., from cell or tissue lysate solutions. For example, antibody arrays can be used to detect biomarkers from bodily fluids, e.g., serum or urine, for diagnostic applications. Tissue microarrays comprise separate tissue cores assembled in array fashion to allow multiplex histological analysis. Cellular microarrays, also called transfection microarrays, comprise various capture agents, such as antibodies, proteins, or lipids, which can interact with cells to facilitate their capture on addressable locations. Chemical compound microarrays comprise arrays of chemical compounds and can be used to detect protein or other biological materials that bind the compounds. Carbohydrate arrays (glycoarrays) comprise arrays of carbohydrates and can detect, e.g., protein that bind sugar moieties. One of skill will appreciate that similar technologies or improvements can be used according to the methods as described herein.

Certain embodiments of the current methods comprise a multi-well reaction vessel, including without limitation, a multi-well plate or a multi-chambered microfluidic device, in which a multiplicity of amplification reactions and, in some embodiments, detection are performed, typically in parallel. In certain embodiments, one or more multiplex reactions for generating amplicons are performed in the same reaction vessel, including without limitation, a multi-well plate, such as a 96-well, a 384-well, a 1536-well plate, and so forth; or a microfluidic device, for example but not limited to, a TaqMan™ Low Density Array (Applied Biosystems, Foster City, CA). In some embodiments, a massively parallel amplifying step comprises a multi-well reaction vessel, including a plate comprising multiple reaction wells, for example but not limited to, a 24-well plate, a 96-well plate, a 384-well plate, or a 1536-well plate; or a multi-chamber microfluidics device, for example but not limited to a low density array wherein each chamber or well comprises an appropriate primer(s), primer set(s), and/or reporter probe(s), as appropriate. Typically such amplification steps occur in a series of parallel single-plex, two-plex, three-plex, four-plex, five-plex, or six-plex reactions, although higher levels of parallel multiplexing are also within the intended scope of the current teachings. These methods can comprise PCR methodology, such as RT-PCR, in each of the wells or chambers to amplify and/or detect nucleic acid molecules of interest.

Low density arrays can include arrays that detect 10s or 100s of molecules as opposed to 1000s of molecules. These arrays can be more sensitive than high density arrays. In embodiments, a low density array such as a TaqMan™ Low Density Array is used to detect one or more gene or gene product in any of Tables 5-12 of WO2018175501. For example, the low density array can be used to detect at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 or 100 genes or gene products selected from any of Tables 5-12 of WO2018175501.

In some embodiments, the disclosed methods comprise a microfluidics device, "lab on a chip," or micrototal analytical system (pTAS). In some embodiments, sample preparation is performed using a microfluidics device. In some embodiments, an amplification reaction is performed using a microfluidics device. In some embodiments, a sequencing or PCR reaction is performed using

a microfluidic device. In some embodiments, the nucleotide sequence of at least a part of an amplified product is obtained using a microfluidics device. In some embodiments, detecting comprises a microfluidic device, including without limitation, a low density array, such as a TaqMan™ Low Density Array. Descriptions of exemplary microfluidic devices can be found in, among other places, Published PCT Application Nos. WO/0185341 and WO 04/011666; Kartalov and Quake, Nucl. Acids Res. 32:2873-79, 2004; and Fiorini and Chiu, Bio Techniques 38:429-46, 2005.

Any appropriate microfluidic device can be used in the methods as described herein. Examples of microfluidic devices that may be used, or adapted for use with molecular profiling, include but are not limited to those described in U.S. Pat. Nos. 7,591,936, 7,581,429, 7,579,136, 7,575,722, 7,568,399, 7,552,741, 7,544,506, 7,541,578, 7,518,726, 7,488,596, 7,485,214, 7,467,928, 7,452,713, 7,452,509, 7,449,096, 7,431,887, 7,422,725, 7,422,669, 7,419,822, 7,419,639, 7,413,709, 7,411,184, 7,402,229, 7,390,463, 7,381,471, 7,357,864, 7,351,592, 7,351,380, 7,338,637, 7,329,391, 7,323,140, 7,261,824, 7,258,837, 7,253,003, 7,238,324, 7,238,255, 7,233,865, 7,229,538, 7,201,881, 7,195,986, 7,189,581, 7,189,580, 7,189,368, 7,141,978, 7,138,062, 7,135,147, 7,125,711, 7,118,910, 7,118,661, 7,640,947, 7,666,361, 7,704,735; U.S. Patent Application Publication 20060035243; and International Patent Publication WO 2010/072410; each of which patents or applications are incorporated herein by reference in their entirety. Another example for use with methods disclosed herein is described in *Chen et al., "Microfluidic isolation and transcriptome analysis of serum vesicles," Lab on a Chip, Dec. 8, 2009 DOI: 10.1039/b916199f.*

Gene Expression Analysis by Massively Parallel Signature Sequencing (MPSS)

This method, described by Brenner et al. (2000) Nature Biotechnology 18:630-634, is a sequencing approach that combines non-gel-based signature sequencing with in vitro cloning of millions of templates on separate microbeads. First, a microbead library of DNA templates is constructed by in vitro cloning. This is followed by the assembly of a planar array of the template-containing microbeads in a flow cell at a high density. The free ends of the cloned templates on each microbead are analyzed simultaneously, using a fluorescence-based signature sequencing method that does not require DNA fragment separation. This method has been shown to simultaneously and accurately provide, in a single operation, hundreds of thousands of gene signature sequences from a cDNA library.

MPSS data has many uses. The expression levels of nearly all transcripts can be quantitatively determined; the abundance of signatures is representative of the expression level of the gene in the analyzed tissue. Quantitative methods for the analysis of tag frequencies and detection of differences among libraries have been published and incorporated into public databases for SAGE™ data and are applicable to MPSS data. The availability of complete genome sequences permits the direct comparison of signatures to genomic sequences and further extends the utility of MPSS data. Because the targets for MPSS analysis are not pre-selected (like on a microarray), MPSS data can characterize

the full complexity of transcriptomes. This is analogous to sequencing millions of ESTs at once, and genomic sequence data can be used so that the source of the MPSS signature can be readily identified by computational means.

Serial Analysis of Gene Expression (SAGE)

Serial analysis of gene expression (SAGE) is a method that allows the simultaneous and quantitative analysis of a large number of gene transcripts, without the need of providing an individual hybridization probe for each transcript. First, a short sequence tag (e.g., about 10-14 bp) is generated that contains sufficient information to uniquely identify a transcript, provided that the tag is obtained from a unique position within each transcript. Then, many transcripts are linked together to form long serial molecules, that can be sequenced, revealing the identity of the multiple tags simultaneously. The expression pattern of any population of transcripts can be quantitatively evaluated by determining the abundance of individual tags, and identifying the gene corresponding to each tag. See, e.g. Velculescu et al. (1995) Science 270:484-487; and Velculescu et al. (1997) Cell 88:243-51.

DNA Copy Number Profiling

Any method capable of determining a DNA copy number profile of a particular sample can be used for molecular profiling according to the methods described herein as long as the resolution is sufficient to identify a copy number variation in the biomarkers as described herein. The skilled artisan is aware of and capable of using a number of different platforms for assessing whole genome copy number changes at a resolution sufficient to identify the copy number of the one or more biomarkers of the methods described herein. Some of the platforms and techniques are described in the embodiments below. In some embodiments as described herein, next generation sequencing or ISH techniques as described herein or known in the art are used for determining copy number / gene amplification.

In some embodiments, the copy number profile analysis involves amplification of whole genome DNA by a whole genome amplification method. The whole genome amplification method can use a strand displacing polymerase and random primers.

In some aspects of these embodiments, the copy number profile analysis involves hybridization of whole genome amplified DNA with a high density array. In a more specific aspect, the high density array has 5,000 or more different probes. In another specific aspect, the high density array has 5,000, 10,000, 20,000, 50,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, or 1,000,000 or more different probes. In another specific aspect, each of the different probes on the array is an oligonucleotide having from about 15 to 200 bases in length. In another specific aspect, each of the different probes on the array is an oligonucleotide having from about 15 to 200, 15 to 150, 15 to 100, 15 to 75, 15 to 60, or 20 to 55 bases in length.

In some embodiments, a microarray is employed to aid in determining the copy number profile for a sample, e.g., cells from a tumor. Microarrays typically comprise a plurality of oligomers (e.g., DNA or RNA polynucleotides or oligonucleotides, or other polymers), synthesized or deposited on a substrate (e.g., glass support) in an array pattern. The support-bound oligomers are "probes", which function to hybridize or bind with a sample material (e.g., nucleic acids prepared or obtained from the tumor samples), in hybridization experiments. The reverse situation can also be applied: the sample can be bound to the microarray substrate and the oligomer probes are in solution for the hybridization. In use, the array surface is contacted with one or more targets under conditions that promote specific, high-affinity binding of the target to one or more of the probes. In some configurations, the sample nucleic acid is labeled with a detectable label, such as a fluorescent tag, so that the hybridized sample and probes are detectable with scanning equipment. DNA array technology offers the potential of using a multitude (e.g., hundreds of thousands) of different oligonucleotides to analyze DNA copy number profiles. In some embodiments, the substrates used for arrays are surface-derivatized glass or silica, or polymer membrane surfaces (see e.g., in Z. Guo, et al., Nucleic Acids Res, 22, 5456-65 (1994); U. Maskos, E. M. Southern, Nucleic Acids Res, 20, 1679-84 (1992), and E. M. Southern, et al., Nucleic Acids Res, 22, 1368-73 (1994), each incorporated by reference herein). Modification of surfaces of array substrates can be accomplished by many techniques. For example, siliceous or metal oxide surfaces can be derivatized with bifunctional silanes, i.e., silanes having a first functional group enabling covalent binding to the surface (e.g., Si-halogen or Si-alkoxy group, as in $--SiCl_3$ or $--Si(OCH_3)_3$, respectively) and a second functional group that can impart the desired chemical and/or physical modifications to the surface to covalently or non-covalently attach ligands and/or the polymers or monomers for the biological probe array. Silylated derivatizations and other surface derivatizations that are known in the art (see for example U.S. Pat. No. 5,624,711 to Sundberg, U.S. Pat. No. 5,266,222 to Willis, and U.S. Pat. No. 5,137,765 to Farnsworth, each incorporated by reference herein). Other processes for preparing arrays are described in U.S. Pat. No. 6,649,348, to Bass et. al., assigned to Agilent Corp., which disclose DNA arrays created by in situ synthesis methods.

Polymer array synthesis is also described extensively in the literature including in the following: WO 00/58516, U.S. Pat. Nos. 5,143,854, 5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,405,783, 5,424,186, 5,451,683, 5,482,867, 5,491,074, 5,527,681, 5,550,215, 5,571,639, 5,578,832, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,795,716, 5,831,070, 5,837,832, 5,856,101, 5,858,659, 5,936,324, 5,968,740, 5,974,164, 5,981,185, 5,981,956, 6,025,601, 6,033,860, 6,040,193, 6,090,555, 6,136,269, 6,269,846 and 6,428,752, 5,412,087, 6,147,205, 6,262,216, 6,310,189, 5,889,165, and 5,959,098 in PCT Applications Nos. PCT/US99/00730 (International Publication No. WO 99/36760) and PCT/US01/04285 (International Publication No. WO 01/58593), which are all incorporated herein by reference in their entirety for all purposes.

Nucleic acid arrays that are useful in the present disclosure include, but are not limited to, those that are commercially available from Affymetrix (Santa Clara, Calif.) under the brand name GeneChip™. Example arrays are shown on the website at affymetrix.com. Another microarray supplier is *Illumina*, Inc., of San Diego, Calif. with example arrays shown on their website at illumina.com.

In some embodiments, the inventive methods provide for sample preparation. Depending on the microarray and experiment to be performed, sample nucleic acid can be prepared in a number of ways by methods known to the skilled artisan. In some aspects as described herein, prior to or concurrent with genotyping (analysis of copy number profiles), the sample may be amplified any number of mechanisms. The most common amplification procedure used involves PCR. See, for example, PCR Technology: Principles and Applications for DNA Amplification (Ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992); PCR Protocols: A Guide to Methods and Applications (Eds. Innis, et al., Academic Press, San Diego, Calif., 1990); Mattila et al., Nucleic Acids Res. 19, 4967 (1991); Eckert et al., PCR Methods and Applications 1, 17 (1991); PCR (Eds. McPherson et al., IRL Press, Oxford); and U.S. Pat. Nos. 4,683,202, 4,683,195, 4,800,159 4,965,188, and 5,333,675, and each of which is incorporated herein by reference in their entireties for all purposes. In some embodiments, the sample may be amplified on the array (e.g., U.S. Pat. No. 6,300,070 which is incorporated herein by reference).

Other suitable amplification methods include the ligase chain reaction (LCR) (for example, Wu and Wallace, Genomics 4, 560 (1989), Landegren et al., Science 241, 1077 (1988) and Barringer et al. Gene 89:117 (1990)), transcription amplification (Kwoh et al., Proc. Natl. Acad. Sci. USA 86, 1173 (1989) and WO88/10315), self-sustained sequence replication (Guatelli et al., Proc. Nat. Acad. Sci. USA, 87, 1874 (1990) and WO90/06995), selective amplification of target polynucleotide sequences (U.S. Pat. No. 6,410,276), consensus sequence primed polymerase chain reaction (CP-PCR) (U.S. Pat. No. 4,437,975), arbitrarily primed polymerase chain reaction (AP-PCR) (U.S. Pat. Nos. 5,413,909, 5,861,245) and nucleic acid based sequence amplification (NABSA). (See, U.S. Pat. Nos. 5,409,818, 5,554,517, and 6,063,603, each of which is incorporated herein by reference). Other amplification methods that may be used are described in, U.S. Pat. Nos. 5,242,794, 5,494,810, 4,988,617 and in U.S. Ser. No. 09/854,317, each of which is incorporated herein by reference.

Additional methods of sample preparation and techniques for reducing the complexity of a nucleic sample are described in Dong et al., Genome Research 11, 1418 (2001), in U.S. Pat. Nos. 6,361,947, 6,391,592 and U.S. Ser. Nos. 09/916,135, 09/920,491 (U.S. Patent Application Publication 20030096235), 09/910,292 (U.S. Patent Application Publication 20030082543), and 10/013,598.

Methods for conducting polynucleotide hybridization assays are well developed in the art. Hybridization assay procedures and conditions used in the methods as described herein will vary depending on the application and are selected in accordance with the general binding methods known including those referred to in: Maniatis et al. Molecular Cloning: A Laboratory Manual (2.sup.nd Ed.

Cold Spring Harbor, N.Y., 1989); Berger and Kimmel Methods in Enzymology, Vol. 152, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young and Davism, P.N.A.S, 80: 1194 (1983). Methods and apparatus for carrying out repeated and controlled hybridization reactions have been described in U.S. Pat. Nos. 5,871,928, 5,874,219, 6,045,996 and 6,386,749, 6,391,623 each of which are incorporated herein by reference.

The methods as described herein may also involve signal detection of hybridization between ligands in after (and/or during) hybridization. See U.S. Pat. Nos. 5,143,854, 5,578,832; 5,631,734; 5,834,758; 5,936,324; 5,981,956; 6,025,601; 6,141,096; 6,185,030; 6,201,639; 6,218,803; and 6,225,625, in U.S. Ser. No. 10/389,194 and in PCT Application PCT/US99/06097 (published as WO99/47964), each of which also is hereby incorporated by reference in its entirety for all purposes.

Methods and apparatus for signal detection and processing of intensity data are disclosed in, for example, U.S. Pat. Nos. 5,143,854, 5,547,839, 5,578,832, 5,631,734, 5,800,992, 5,834,758; 5,856,092, 5,902,723, 5,936,324, 5,981,956, 6,025,601, 6,090,555, 6,141,096, 6,185,030, 6,201,639; 6,218,803; and 6,225,625, in U.S. Ser. Nos. 10/389,194, 60/493,495 and in PCT Application PCT/US99/06097 (published as WO99/47964), each of which also is hereby incorporated by reference in its entirety for all purposes.

Immuno-based Assays

Protein-based detection molecular profiling techniques include immunoaffinity assays based on antibodies selectively immunoreactive with mutant gene encoded protein according to the present methods. These techniques include without limitation immunoprecipitation, Western blot analysis, molecular binding assays, enzyme-linked immunosorbent assay (ELISA), enzyme-linked immunofiltration assay (ELIFA), fluorescence activated cell sorting (FACS) and the like. For example, an optional method of detecting the expression of a biomarker in a sample comprises contacting the sample with an antibody against the biomarker, or an immunoreactive fragment of the antibody thereof, or a recombinant protein containing an antigen binding region of an antibody against the biomarker; and then detecting the binding of the biomarker in the sample. Methods for producing such antibodies are known in the art. Antibodies can be used to immunoprecipitate specific proteins from solution samples or to immunoblot proteins separated by, e.g., polyacrylamide gels. Immunocytochemical methods can also be used in detecting specific protein polymorphisms in tissues or cells. Other well-known antibody-based techniques can also be used including, e.g., ELISA, radioimmunoassay (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal or polyclonal antibodies. See, e.g., U.S. Pat. Nos. 4,376,110 and 4,486,530, both of which are incorporated herein by reference.

In alternative methods, the sample may be contacted with an antibody specific for a biomarker under conditions sufficient for an antibody-biomarker complex to form, and then detecting said complex. The presence of the biomarker may be detected in a number of ways, such as by

Western blotting and ELISA procedures for assaying a wide variety of tissues and samples, including plasma or serum. A wide range of immunoassay techniques using such an assay format are available, see, e.g., U.S. Pat. Nos. 4,016,043, 4,424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in the traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target biomarker.

A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present methods. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate, and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of biomarker.

Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In a typical forward sandwich assay, a first antibody having specificity for the biomarker is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 40°C such as between 25°C and 32°C inclusive) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the biomarker. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the molecular marker.

An alternative method involves immobilizing the target biomarkers in the sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled

antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule. By "reporter molecule", as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, β-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable color change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody-molecular marker complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of biomarker which was present in the sample. Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic color visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-molecular marker complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength, the fluorescence observed indicates the presence of the molecular marker of interest. Immunofluorescence and EIA techniques are both very well established in the art. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

Immunohistochemistry (IHC)

IHC is a process of localizing antigens (e.g., proteins) in cells of a tissue binding antibodies specifically to antigens in the tissues. The antigen-binding antibody can be conjugated or fused to a tag that allows its detection, e.g., via visualization. In some embodiments, the tag is an enzyme that can catalyze a color-producing reaction, such as alkaline phosphatase or horseradish peroxidase. The

enzyme can be fused to the antibody or non-covalently bound, e.g., using a biotin-avadin system. Alternatively, the antibody can be tagged with a fluorophore, such as fluorescein, rhodamine, DyLight Fluor or Alexa Fluor. The antigen-binding antibody can be directly tagged or it can itself be recognized by a detection antibody that carries the tag. Using IHC, one or more proteins may be detected. The expression of a gene product can be related to its staining intensity compared to control levels. In some embodiments, the gene product is considered differentially expressed if its staining varies at least 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.5, 2.7, 3.0, 4, 5, 6, 7, 8, 9 or 10-fold in the sample versus the control.

IHC comprises the application of antigen-antibody interactions to histochemical techniques. In an illustrative example, a tissue section is mounted on a slide and is incubated with antibodies (polyclonal or monoclonal) specific to the antigen (primary reaction). The antigen-antibody signal is then amplified using a second antibody conjugated to a complex of peroxidase antiperoxidase (PAP), avidin-biotin-peroxidase (ABC) or avidin-biotin alkaline phosphatase. In the presence of substrate and chromogen, the enzyme forms a colored deposit at the sites of antibody-antigen binding. Immunofluorescence is an alternate approach to visualize antigens. In this technique, the primary antigen-antibody signal is amplified using a second antibody conjugated to a fluorochrome. On UV light absorption, the fluorochrome emits its own light at a longer wavelength (fluorescence), thus allowing localization of antibody-antigen complexes.

Epigenetic Status

Molecular profiling methods according to the present disclosure also comprise measuring epigenetic change, i.e., modification in a gene caused by an epigenetic mechanism, such as a change in methylation status or histone acetylation. Frequently, the epigenetic change will result in an alteration in the levels of expression of the gene which may be detected (at the RNA or protein level as appropriate) as an indication of the epigenetic change. Often the epigenetic change results in silencing or down regulation of the gene, referred to as "epigenetic silencing." The most frequently investigated epigenetic change in the methods as described herein involves determining the DNA methylation status of a gene, where an increased level of methylation is typically associated with the relevant cancer (since it may cause down regulation of gene expression). Aberrant methylation, which may be referred to as hypermethylation, of the gene or genes can be detected. Typically, the methylation status is determined in suitable CpG islands which are often found in the promoter region of the gene(s). The term "methylation," "methylation state" or "methylation status" may refers to the presence or absence of 5-methylcytosine at one or a plurality of CpG dinucleotides within a DNA sequence. CpG dinucleotides are typically concentrated in the promoter regions and exons of human genes.

Diminished gene expression can be assessed in terms of DNA methylation status or in terms of expression levels as determined by the methylation status of the gene. One method to detect

epigenetic silencing is to determine that a gene which is expressed in normal cells is less expressed or not expressed in tumor cells. Accordingly, the present disclosure provides for a method of molecular profiling comprising detecting epigenetic silencing.

Various assay procedures to directly detect methylation are known in the art, and can be used in conjunction with the present methods. These assays rely onto two distinct approaches: bisulphite conversion based approaches and non-bisulphite based approaches. Non-bisulphite based methods for analysis of DNA methylation rely on the inability of methylation-sensitive enzymes to cleave methylation cytosines in their restriction. The bisulphite conversion relies on treatment of DNA samples with sodium bisulphite which converts unmethylated cytosine to uracil, while methylated cytosines are maintained (Furuichi Y, Wataya Y, Hayatsu H, Ukita T. Biochem Biophys Res Commun. 1970 Dec 9;41(5):1185-91). This conversion results in a change in the sequence of the original DNA. Methods to detect such changes include MS AP-PCR (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction), a technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., Cancer Research 57:594-599, 1997; MethyLight™, which refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., Cancer Res. 59:2302-2306, 1999; the HeavyMethyl™assay, in the embodiment thereof implemented herein, is an assay, wherein methylation specific blocking probes (also referred to herein as blockers) covering CpG positions between, or covered by the amplification primers enable methylation-specific selective amplification of a nucleic acid sample; HeavyMethyl™MethyLight™ is a variation of the MethyLight™ assay wherein the MethyLight™ assay is combined with methylation specific blocking probes covering CpG positions between the amplification primers; Ms-SNuPE (Methylation-sensitive Single Nucleotide Primer Extension) is an assay described by Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997; MSP (Methylation-specific PCR) is a methylation assay described by Herman et al. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996, and by U.S. Pat. No. 5,786,146; COBRA (Combined Bisulfite Restriction Analysis) is a methylation assay described by Xiong & Laird, Nucleic Acids Res. 25:2532-2534, 1997; MCA (Methylated CpG Island Amplification) is a methylation assay described by Toyota et al., Cancer Res. 59:2307-12, 1999, and in WO 00/26401A1.

Other techniques for DNA methylation analysis include sequencing, methylation-specific PCR (MS-PCR), melting curve methylation-specific PCR (McMS-PCR), MLPA with or without bisulfite treatment, QAMA, MSRE-PCR, MethyLight, ConLight-MSP, bisulfite conversion-specific methylation-specific PCR (BS-MSP), COBRA (which relies upon use of restriction enzymes to reveal methylation dependent sequence differences in PCR products of sodium bisulfite-treated DNA), methylation-sensitive single-nucleotide primer extension conformation (MS-SNuPE), methylation-sensitive single-strand conformation analysis (MS-SSCA), Melting curve combined bisulfite restriction analysis (McCOBRA), PyroMethA, HeavyMethyl, MALDI-TOF, MassARRAY, Quantitative analysis of methylated alleles (QAMA), enzymatic regional methylation assay (ERMA),

QBSUPT, MethylQuant, Quantitative PCR sequencing and oligonucleotide-based microarray systems, Pyrosequencing, Meth-DOP-PCR. A review of some useful techniques is provided in Nucleic acids research, 1998, Vol. 26, No. 10, 2255-2264; Nature Reviews, 2003, Vol.3, 253-266; Oral Oncology, 2006, Vol. 42, 5-13, which references are incorporated herein in their entirety. Any of these techniques may be used in accordance with the present methods, as appropriate. Other techniques are described in U.S. Patent Publications 20100144836; and 20100184027, which applications are incorporated herein by reference in their entirety.

Through the activity of various acetylases and deacetylylases the DNA binding function of histone proteins is tightly regulated. Furthermore, histone acetylation and histone deactelyation have been linked with malignant progression. See Nature, 429: 457-63, 2004. Methods to analyze histone acetylation are described in U.S. Patent Publications 20100144543 and 20100151468, which applications are incorporated herein by reference in their entirety.

Sequence Analysis

Molecular profiling according to the present disclosure comprises methods for genotyping one or more biomarkers by determining whether an individual has one or more nucleotide variants (or amino acid variants) in one or more of the genes or gene products. Genotyping one or more genes according to the methods as described herein in some embodiments, can provide more evidence for selecting a treatment.

The biomarkers as described herein can be analyzed by any method useful for determining alterations in nucleic acids or the proteins they encode. According to one embodiment, the ordinary skilled artisan can analyze the one or more genes for mutations including deletion mutants, insertion mutants, frame shift mutants, nonsense mutants, missense mutant, and splice mutants.

Nucleic acid used for analysis of the one or more genes can be isolated from cells in the sample according to standard methodologies (Sambrook et al., 1989). The nucleic acid, for example, may be genomic DNA or fractionated or whole cell RNA, or miRNA acquired from exosomes or cell surfaces. Where RNA is used, it may be desired to convert the RNA to a complementary DNA. In one embodiment, the RNA is whole cell RNA; in another, it is poly-A RNA; in another, it is exosomal RNA. Normally, the nucleic acid is amplified. Depending on the format of the assay for analyzing the one or more genes, the specific nucleic acid of interest is identified in the sample directly using amplification or with a second, known nucleic acid following amplification. Next, the identified product is detected. In certain applications, the detection may be performed by visual means (e.g., ethidium bromide staining of a gel). Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals (Affymax Technology; Bellus, 1994).

Various types of defects are known to occur in the biomarkers as described herein. Alterations include without limitation deletions, insertions, point mutations, and duplications. Point mutations can

be silent or can result in stop codons, frame shift mutations or amino acid substitutions. Mutations in and outside the coding region of the one or more genes may occur and can be analyzed according to the methods as described herein. The target site of a nucleic acid of interest can include the region wherein the sequence varies. Examples include, but are not limited to, polymorphisms which exist in different forms such as single nucleotide variations, nucleotide repeats, multibase deletion (more than one nucleotide deleted from the consensus sequence), multibase insertion (more than one nucleotide inserted from the consensus sequence), microsatellite repeats (small numbers of nucleotide repeats with a typical 5-1000 repeat units), di-nucleotide repeats, tri-nucleotide repeats, sequence rearrangements (including translocation and duplication), chimeric sequence (two sequences from different gene origins are fused together), and the like. Among sequence polymorphisms, the most frequent polymorphisms in the human genome are single-base variations, also called single-nucleotide polymorphisms (SNPs). SNPs are abundant, stable and widely distributed across the genome.

Molecular profiling includes methods for haplotyping one or more genes. The haplotype is a set of genetic determinants located on a single chromosome and it typically contains a particular combination of alleles (all the alternative sequences of a gene) in a region of a chromosome. In other words, the haplotype is phased sequence information on individual chromosomes. Very often, phased SNPs on a chromosome define a haplotype. A combination of haplotypes on chromosomes can determine a genetic profile of a cell. It is the haplotype that determines a linkage between a specific genetic marker and a disease mutation. Haplotyping can be done by any methods known in the art. Common methods of scoring SNPs include hybridization microarray or direct gel sequencing, reviewed in Landgren et al., Genome Research, 8:769-776, 1998. For example, only one copy of one or more genes can be isolated from an individual and the nucleotide at each of the variant positions is determined. Alternatively, an allele specific PCR or a similar method can be used to amplify only one copy of the one or more genes in an individual, and SNPs at the variant positions of the present disclosure are determined. The Clark method known in the art can also be employed for haplotyping. A high throughput molecular haplotyping method is also disclosed in Tost et al., Nucleic Acids Res., 30(19):e96 (2002), which is incorporated herein by reference.

Thus, additional variant(s) that are in linkage disequilibrium with the variants and/or haplotypes of the present disclosure can be identified by a haplotyping method known in the art, as will be apparent to a skilled artisan in the field of genetics and haplotyping. The additional variants that are in linkage disequilibrium with a variant or haplotype of the present disclosure can also be useful in the various applications as described below.

For purposes of genotyping and haplotyping, both genomic DNA and mRNA/cDNA can be used, and both are herein referred to generically as "gene."

Numerous techniques for detecting nucleotide variants are known in the art and can all be used for the method of this disclosure. The techniques can be protein-based or nucleic acid-based. In either case, the techniques used must be sufficiently sensitive so as to accurately detect the small

nucleotide or amino acid variations. Very often, a probe is used which is labeled with a detectable marker. Unless otherwise specified in a particular technique described below, any suitable marker known in the art can be used, including but not limited to, radioactive isotopes, fluorescent compounds, biotin which is detectable using streptavidin, enzymes (e.g., alkaline phosphatase), substrates of an enzyme, ligands and antibodies, etc. See Jablonski et al., Nucleic Acids Res., 14:6115-6128 (1986); Nguyen et al., Biotechniques, 13:116-123 (1992); Rigby et al., J. Mol. Biol., 113:237-251 (1977).

In a nucleic acid-based detection method, target DNA sample, i.e., a sample containing genomic DNA, cDNA, mRNA and/or miRNA, corresponding to the one or more genes must be obtained from the individual to be tested. Any tissue or cell sample containing the genomic DNA, miRNA, mRNA, and/or cDNA (or a portion thereof) corresponding to the one or more genes can be used. For this purpose, a tissue sample containing cell nucleus and thus genomic DNA can be obtained from the individual. Blood samples can also be useful except that only white blood cells and other lymphocytes have cell nucleus, while red blood cells are without a nucleus and contain only mRNA or miRNA. Nevertheless, miRNA and mRNA are also useful as either can be analyzed for the presence of nucleotide variants in its sequence or serve as template for cDNA synthesis. The tissue or cell samples can be analyzed directly without much processing. Alternatively, nucleic acids including the target sequence can be extracted, purified, and/or amplified before they are subject to the various detecting procedures discussed below. Other than tissue or cell samples, cDNAs or genomic DNAs from a cDNA or genomic DNA library constructed using a tissue or cell sample obtained from the individual to be tested are also useful.

To determine the presence or absence of a particular nucleotide variant, sequencing of the target genomic DNA or cDNA, particularly the region encompassing the nucleotide variant locus to be detected. Various sequencing techniques are generally known and widely used in the art including the Sanger method and Gilbert chemical method. The pyrosequencing method monitors DNA synthesis in real time using a luminometric detection system. Pyrosequencing has been shown to be effective in analyzing genetic polymorphisms such as single-nucleotide polymorphisms and can also be used in the present methods. See Nordstrom et al., Biotechnol. Appl. Biochem., 31(2):107-112 (2000); Ahmadian et al., Anal. Biochem., 280:103-110 (2000).

Nucleic acid variants can be detected by a suitable detection process. Non limiting examples of methods of detection, quantification, sequencing and the like are; mass detection of mass modified amplicons (e.g., matrix-assisted laser desorption ionization (MALDI) mass spectrometry and electrospray (ES) mass spectrometry), a primer extension method (e.g., iPLEX™; Sequenom, Inc.), microsequencing methods (e.g., a modification of primer extension methodology), ligase sequence determination methods (e.g., U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (e.g., U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), direct DNA sequencing, fragment analysis (FA), restriction fragment length

polymorphism (RFLP analysis), allele specific oligonucleotide (ASO) analysis, methylation-specific PCR (MSPCR), pyrosequencing analysis, acycloprime analysis, Reverse dot blot, GeneChip microarrays, Dynamic allele-specific hybridization (DASH), Peptide nucleic acid (PNA) and locked nucleic acids (LNA) probes, TaqMan, Molecular Beacons, Intercalating dye, FRET primers, AlphaScreen, SNPstream, genetic bit analysis (GBA), Multiplex minisequencing, SNaPshot, GOOD assay, Microarray miniseq, arrayed primer extension (APEX), Microarray primer extension (e.g., microarray sequence determination methods), Tag arrays, Coded microspheres, Template-directed incorporation (TDI), fluorescence polarization, Colorimetric oligonucleotide ligation assay (OLA), Sequence-coded OLA, Microarray ligation, Ligase chain reaction, Padlock probes, Invader assay, hybridization methods (e.g., hybridization using at least one probe, hybridization using at least one fluorescently labeled probe, and the like), conventional dot blot analyses, single strand conformational polymorphism analysis (SSCP, e.g., U.S. Pat. Nos. 5,891,625 and 6,013,499; Orita et al., Proc. Natl. Acad. Sci. U.S.A. 86: 27776-2770 (1989)), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and techniques described in Sheffield et al., Proc. Natl. Acad. Sci. USA 49: 699-706 (1991), White et al., Genomics 12: 301-306 (1992), Grompe et al., Proc. Natl. Acad. Sci. USA 86: 5855-5892 (1989), and Grompe, Nature Genetics 5: 111-117 (1993), cloning and sequencing, electrophoresis, the use of hybridization probes and quantitative real time polymerase chain reaction (QRT-PCR), digital PCR, nanopore sequencing, chips and combinations thereof. The detection and quantification of alleles or paralogs can be carried out using the "closed-tube" methods described in U.S. patent application Ser. No. 11/950,395, filed on Dec. 4, 2007. In some embodiments the amount of a nucleic acid species is determined by mass spectrometry, primer extension, sequencing (e.g., any suitable method, for example nanopore or pyrosequencing), Quantitative PCR (Q-PCR or QRT-PCR), digital PCR, combinations thereof, and the like.

The term "sequence analysis" as used herein refers to determining a nucleotide sequence, e.g., that of an amplification product. The entire sequence or a partial sequence of a polynucleotide, e.g., DNA or mRNA, can be determined, and the determined nucleotide sequence can be referred to as a "read" or "sequence read." For example, linear amplification products may be analyzed directly without further amplification in some embodiments (e.g., by using single-molecule sequencing methodology). In certain embodiments, linear amplification products may be subject to further amplification and then analyzed (e.g., using sequencing by ligation or pyrosequencing methodology). Reads may be subject to different types of sequence analysis. Any suitable sequencing method can be used to detect, and determine the amount of, nucleotide sequence species, amplified nucleic acid species, or detectable products generated from the foregoing. Examples of certain sequencing methods are described hereafter.

A sequence analysis apparatus or sequence analysis component(s) includes an apparatus, and one or more components used in conjunction with such apparatus, that can be used by a person of ordinary skill to determine a nucleotide sequence resulting from processes described herein (e.g.,

linear and/or exponential amplification products). Examples of sequencing platforms include, without limitation, the 454 platform (Roche) (Margulies, M. et al. 2005 Nature 437, 376-380), Illumina Genomic Analyzer (or Solexa platform) or SOLID System (Applied Biosystems; see PCT patent application publications WO 06/084132 entitled "Reagents, Methods, and Libraries For Bead-Based Sequencing" and WO07/121,489 entitled "Reagents, Methods, and Libraries for Gel-Free Bead-Based Sequencing"), the Helicos True Single Molecule DNA sequencing technology (Harris TD et al. 2008 Science, 320, 106-109), the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and nanopore sequencing (Soni G V and Meller A. 2007 Clin Chem 53: 1996-2001), Ion semiconductor sequencing (Ion Torrent Systems, Inc, San Francisco, CA), or DNA nanoball sequencing (Complete Genomics, Mountain View, CA), VisiGen Biotechnologies approach (Invitrogen) and polony sequencing. Such platforms allow sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel manner (Dear Brief Funct Genomic Proteomic 2003; 1: 397-416; Haimovich, Methods, challenges, and promise of next-generation sequencing in cancer biology. Yale J Biol Med. 2011 Dec;84(4):439-46). These non-Sanger-based sequencing technologies are sometimes referred to as NextGen sequencing, NGS, next-generation sequencing, next generation sequencing, and variations thereof. Typically they allow much higher throughput than the traditional Sanger approach. See Schuster, Next-generation sequencing transforms today's biology, *Nature Methods* **5**:16-18 (2008); Metzker, Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31-46; Levy and Myers, Advancements in Next-Generation Sequencing. Annu Rev Genomics Hum Genet. 2016 Aug 31;17:95-115. These platforms can allow sequencing of clonally expanded or non-amplified single molecules of nucleic acid fragments. Certain platforms involve, for example, sequencing by ligation of dye-modified probes (including cyclic ligation and cleavage), pyrosequencing, and single-molecule sequencing. Nucleotide sequence species, amplification nucleic acid species and detectable products generated there from can be analyzed by such sequence analysis platforms. Next-generation sequencing can be used in the methods as described herein, e.g., to determine mutations, copy number, or expression levels, as appropriate. The methods can be used to perform whole genome sequencing or sequencing of specific sequences of interest, such as a gene of interest or a fragment thereof.

Sequencing by ligation is a nucleic acid sequencing method that relies on the sensitivity of DNA ligase to base-pairing mismatch. DNA ligase joins together ends of DNA that are correctly base paired. Combining the ability of DNA ligase to join together only correctly base paired DNA ends, with mixed pools of fluorescently labeled oligonucleotides or primers, enables sequence determination by fluorescence detection. Longer sequence reads may be obtained by including primers containing cleavable linkages that can be cleaved after label identification. Cleavage at the linker removes the label and regenerates the 5' phosphate on the end of the ligated primer, preparing the primer for another round of ligation. In some embodiments primers may be labeled with more than one fluorescent label, e.g., at least 1, 2, 3, 4, or 5 fluorescent labels.

Sequencing by ligation generally involves the following steps. Clonal bead populations can be prepared in emulsion microreactors containing target nucleic acid template sequences, amplification reaction components, beads and primers. After amplification, templates are denatured and bead enrichment is performed to separate beads with extended templates from undesired beads (e.g., beads with no extended templates). The template on the selected beads undergoes a 3' modification to allow covalent bonding to the slide, and modified beads can be deposited onto a glass slide. Deposition chambers offer the ability to segment a slide into one, four or eight chambers during the bead loading process. For sequence analysis, primers hybridize to the adapter sequence. A set of four color dye-labeled probes competes for ligation to the sequencing primer. Specificity of probe ligation is achieved by interrogating every 4th and 5th base during the ligation series. Five to seven rounds of ligation, detection and cleavage record the color at every 5th position with the number of rounds determined by the type of library used. Following each round of ligation, a new complimentary primer offset by one base in the 5' direction is laid down for another series of ligations. Primer reset and ligation rounds (5-7 ligation cycles per round) are repeated sequentially five times to generate 25-35 base pairs of sequence for a single tag. With mate-paired sequencing, this process is repeated for a second tag.

Pyrosequencing is a nucleic acid sequencing method based on sequencing by synthesis, which relies on detection of a pyrophosphate released on nucleotide incorporation. Generally, sequencing by synthesis involves synthesizing, one nucleotide at a time, a DNA strand complimentary to the strand whose sequence is being sought. Target nucleic acids may be immobilized to a solid support, hybridized with a sequencing primer, incubated with DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5' phosphosulfate and luciferin. Nucleotide solutions are sequentially added and removed. Correct incorporation of a nucleotide releases a pyrophosphate, which interacts with ATP sulfurylase and produces ATP in the presence of adenosine 5' phosphosulfate, fueling the luciferin reaction, which produces a chemiluminescent signal allowing sequence determination. The amount of light generated is proportional to the number of bases added. Accordingly, the sequence downstream of the sequencing primer can be determined. An illustrative system for pyrosequencing involves the following steps: ligating an adaptor nucleic acid to a nucleic acid under investigation and hybridizing the resulting nucleic acid to a bead; amplifying a nucleotide sequence in an emulsion; sorting beads using a picoliter multiwell solid support; and sequencing amplified nucleotide sequences by pyrosequencing methodology (e.g., Nakano et al., "Single-molecule PCR using water-in-oil emulsion;" Journal of Biotechnology 102: 117-124 (2003)).

Certain single-molecule sequencing embodiments are based on the principal of sequencing by synthesis, and use single-pair Fluorescence Resonance Energy Transfer (single pair FRET) as a mechanism by which photons are emitted as a result of successful nucleotide incorporation. The emitted photons often are detected using intensified or high sensitivity cooled charge-couple-devices in conjunction with total internal reflection microscopy (TIRM). Photons are only emitted when the

introduced reaction solution contains the correct nucleotide for incorporation into the growing nucleic acid chain that is synthesized as a result of the sequencing process. In FRET based single-molecule sequencing, energy is transferred between two fluorescent dyes, sometimes polymethine cyanine dyes Cy3 and Cy5, through long-range dipole interactions. The donor is excited at its specific excitation wavelength and the excited state energy is transferred, non-radiatively to the acceptor dye, which in turn becomes excited. The acceptor dye eventually returns to the ground state by radiative emission of a photon. The two dyes used in the energy transfer process represent the "single pair" in single pair FRET. Cy3 often is used as the donor fluorophore and often is incorporated as the first labeled nucleotide. Cy5 often is used as the acceptor fluorophore and is used as the nucleotide label for successive nucleotide additions after incorporation of a first Cy3 labeled nucleotide. The fluorophores generally are within 10 nanometers of each for energy transfer to occur successfully.

An example of a system that can be used based on single-molecule sequencing generally involves hybridizing a primer to a target nucleic acid sequence to generate a complex; associating the complex with a solid phase; iteratively extending the primer by a nucleotide tagged with a fluorescent molecule; and capturing an image of fluorescence resonance energy transfer signals after each iteration (e.g., U.S. Pat. No. 7,169,314; Braslavsky et al., PNAS 100(7): 3960-3964 (2003)). Such a system can be used to directly sequence amplification products (linearly or exponentially amplified products) generated by processes described herein. In some embodiments the amplification products can be hybridized to a primer that contains sequences complementary to immobilized capture sequences present on a solid support, a bead or glass slide for example. Hybridization of the primer-amplification product complexes with the immobilized capture sequences, immobilizes amplification products to solid supports for single pair FRET based sequencing by synthesis. The primer often is fluorescent, so that an initial reference image of the surface of the slide with immobilized nucleic acids can be generated. The initial reference image is useful for determining locations at which true nucleotide incorporation is occurring. Fluorescence signals detected in array locations not initially identified in the "primer only" reference image are discarded as non-specific fluorescence. Following immobilization of the primer-amplification product complexes, the bound nucleic acids often are sequenced in parallel by the iterative steps of, a) polymerase extension in the presence of one fluorescently labeled nucleotide, b) detection of fluorescence using appropriate microscopy, TIRM for example, c) removal of fluorescent nucleotide, and d) return to step a with a different fluorescently labeled nucleotide.

In some embodiments, nucleotide sequencing may be by solid phase single nucleotide sequencing methods and processes. Solid phase single nucleotide sequencing methods involve contacting target nucleic acid and solid support under conditions in which a single molecule of sample nucleic acid hybridizes to a single molecule of a solid support. Such conditions can include providing the solid support molecules and a single molecule of target nucleic acid in a "microreactor." Such conditions also can include providing a mixture in which the target nucleic acid molecule can

hybridize to solid phase nucleic acid on the solid support. Single nucleotide sequencing methods useful in the embodiments described herein are described in U.S. Provisional Patent Application Ser. No. 61/021,871 filed Jan. 17, 2008.

In certain embodiments, nanopore sequencing detection methods include (a) contacting a target nucleic acid for sequencing ("base nucleic acid," e.g., linked probe molecule) with sequence-specific detectors, under conditions in which the detectors specifically hybridize to substantially complementary subsequences of the base nucleic acid; (b) detecting signals from the detectors and (c) determining the sequence of the base nucleic acid according to the signals detected. In certain embodiments, the detectors hybridized to the base nucleic acid are disassociated from the base nucleic acid (e.g., sequentially dissociated) when the detectors interfere with a nanopore structure as the base nucleic acid passes through a pore, and the detectors disassociated from the base sequence are detected. In some embodiments, a detector disassociated from a base nucleic acid emits a detectable signal, and the detector hybridized to the base nucleic acid emits a different detectable signal or no detectable signal. In certain embodiments, nucleotides in a nucleic acid (e.g., linked probe molecule) are substituted with specific nucleotide sequences corresponding to specific nucleotides ("nucleotide representatives"), thereby giving rise to an expanded nucleic acid (e.g., U.S. Pat. No. 6,723,513), and the detectors hybridize to the nucleotide representatives in the expanded nucleic acid, which serves as a base nucleic acid. In such embodiments, nucleotide representatives may be arranged in a binary or higher order arrangement (e.g., Soni and Meller, Clinical Chemistry 53(11): 1996-2001 (2007)). In some embodiments, a nucleic acid is not expanded, does not give rise to an expanded nucleic acid, and directly serves a base nucleic acid (e.g., a linked probe molecule serves as a non-expanded base nucleic acid), and detectors are directly contacted with the base nucleic acid. For example, a first detector may hybridize to a first subsequence and a second detector may hybridize to a second subsequence, where the first detector and second detector each have detectable labels that can be distinguished from one another, and where the signals from the first detector and second detector can be distinguished from one another when the detectors are disassociated from the base nucleic acid. In certain embodiments, detectors include a region that hybridizes to the base nucleic acid (e.g., two regions), which can be about 3 to about 100 nucleotides in length (e.g., about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 85, 90, or 95 nucleotides in length). A detector also may include one or more regions of nucleotides that do not hybridize to the base nucleic acid. In some embodiments, a detector is a molecular beacon. A detector often comprises one or more detectable labels independently selected from those described herein. Each detectable label can be detected by any convenient detection process capable of detecting a signal generated by each label (e.g., magnetic, electric, chemical, optical and the like). For example, a CD camera can be used to detect signals from one or more distinguishable quantum dots linked to a detector.

In certain sequence analysis embodiments, reads may be used to construct a larger nucleotide sequence, which can be facilitated by identifying overlapping sequences in different reads and by

using identification sequences in the reads. Such sequence analysis methods and software for constructing larger sequences from reads are known to the person of ordinary skill (e.g., Venter et al., Science 291: 1304-1351 (2001)). Specific reads, partial nucleotide sequence constructs, and full nucleotide sequence constructs may be compared between nucleotide sequences within a sample nucleic acid (i.e., internal comparison) or may be compared with a reference sequence (i.e., reference comparison) in certain sequence analysis embodiments. Internal comparisons can be performed in situations where a sample nucleic acid is prepared from multiple samples or from a single sample source that contains sequence variations. Reference comparisons sometimes are performed when a reference nucleotide sequence is known and an objective is to determine whether a sample nucleic acid contains a nucleotide sequence that is substantially similar or the same, or different, than a reference nucleotide sequence. Sequence analysis can be facilitated by the use of sequence analysis apparatus and components described above.

Primer extension polymorphism detection methods, also referred to herein as "microsequencing" methods, typically are carried out by hybridizing a complementary oligonucleotide to a nucleic acid carrying the polymorphic site. In these methods, the oligonucleotide typically hybridizes adjacent to the polymorphic site. The term "adjacent" as used in reference to "microsequencing" methods, refers to the 3' end of the extension oligonucleotide being sometimes 1 nucleotide from the 5' end of the polymorphic site, often 2 or 3, and at times 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, often 1, 2, or 3 nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine which polymorphic variant or variants are present. Oligonucleotide extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. The extension products can be detected in any manner, such as by fluorescence methods (see, e.g., Chen & Kwok, Nucleic Acids Research 25: 347-353 (1997) and Chen et al., Proc. Natl. Acad. Sci. USA 94/20: 10756-10761 (1997)) or by mass spectrometric methods (e.g., MALDI-TOF mass spectrometry) and other methods described herein. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; 6,194,144; and 6,258,538.

Microsequencing detection methods often incorporate an amplification process that proceeds the extension step. The amplification process typically amplifies a region from a nucleic acid sample that comprises the polymorphic site. Amplification can be carried out using methods described above, or for example using a pair of oligonucleotide primers in a polymerase chain reaction (PCR), in which one oligonucleotide primer typically is complementary to a region 3' of the polymorphism and the other typically is complementary to a region 5' of the polymorphism. A PCR primer pair may be used

in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GeneAmp™ Systems available from Applied Biosystems.

Other appropriate sequencing methods include multiplex polony sequencing (as described in Shendure et al., Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome, Sciencexpress, Aug. 4, 2005, pg 1 available at www.sciencexpress.org/4 Aug. 2005/Page1/10.1126/science.1117389, incorporated herein by reference), which employs immobilized microbeads, and sequencing in microfabricated picoliter reactors (as described in Margulies et al., Genome Sequencing in Microfabricated High-Density Picolitre Reactors, Nature, August 2005, available at www.nature.com/nature (published online 31 Jul. 2005, doi:10.1038/nature03959, incorporated herein by reference).

Whole genome sequencing may also be used for discriminating alleles of RNA transcripts, in some embodiments. Examples of whole genome sequencing methods include, but are not limited to, nanopore-based sequencing methods, sequencing by synthesis and sequencing by ligation, as described above.

Nucleic acid variants can also be detected using standard electrophoretic techniques. Although the detection step can sometimes be preceded by an amplification step, amplification is not required in the embodiments described herein. Examples of methods for detection and quantification of a nucleic acid using electrophoretic techniques can be found in the art. A non-limiting example comprises running a sample (e.g., mixed nucleic acid sample isolated from maternal serum, or amplification nucleic acid species, for example) in an agarose or polyacrylamide gel. The gel may be labeled (e.g., stained) with ethidium bromide (see, Sambrook and Russell, Molecular Cloning: A Laboratory Manual 3d ed., 2001). The presence of a band of the same size as the standard control is an indication of the presence of a target nucleic acid sequence, the amount of which may then be compared to the control based on the intensity of the band, thus detecting and quantifying the target sequence of interest. In some embodiments, restriction enzymes capable of distinguishing between maternal and paternal alleles may be used to detect and quantify target nucleic acid species. In certain embodiments, oligonucleotide probes specific to a sequence of interest are used to detect the presence of the target sequence of interest. The oligonucleotides can also be used to indicate the amount of the target nucleic acid molecules in comparison to the standard control, based on the intensity of signal imparted by the probe.

Sequence-specific probe hybridization can be used to detect a particular nucleic acid in a mixture or mixed population comprising other species of nucleic acids. Under sufficiently stringent hybridization conditions, the probes hybridize specifically only to substantially complementary sequences. The stringency of the hybridization conditions can be relaxed to tolerate varying amounts of sequence mismatch. A number of hybridization formats are known in the art, which include but are

not limited to, solution phase, solid phase, or mixed phase hybridization assays. The following articles provide an overview of the various hybridization assay formats: Singer et al., Biotechniques 4:230, 1986; Haase et al., Methods in Virology, pp. 189-226, 1984; Wilkinson, In situ Hybridization, Wilkinson ed., IRL Press, Oxford University Press, Oxford; and Hames and Higgins eds., Nucleic Acid Hybridization: A Practical Approach, IRL Press, 1987.

Hybridization complexes can be detected by techniques known in the art. Nucleic acid probes capable of specifically hybridizing to a target nucleic acid (e.g., mRNA or DNA) can be labeled by any suitable method, and the labeled probe used to detect the presence of hybridized nucleic acids. One commonly used method of detection is autoradiography, using probes labeled with ^{3}H, ^{125}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, or the like. The choice of radioactive isotope depends on research preferences due to ease of synthesis, stability, and half-lives of the selected isotopes. Other labels include compounds (e.g., biotin and digoxigenin), which bind to antiligands or antibodies labeled with fluorophores, chemiluminescent agents, and enzymes. In some embodiments, probes can be conjugated directly with labels such as fluorophores, chemiluminescent agents or enzymes. The choice of label depends on sensitivity required, ease of conjugation with the probe, stability requirements, and available instrumentation.

In embodiments, fragment analysis (referred to herein as "FA") methods are used for molecular profiling. Fragment analysis (FA) includes techniques such as restriction fragment length polymorphism (RFLP) and/or (amplified fragment length polymorphism). If a nucleotide variant in the target DNA corresponding to the one or more genes results in the elimination or creation of a restriction enzyme recognition site, then digestion of the target DNA with that particular restriction enzyme will generate an altered restriction fragment length pattern. Thus, a detected RFLP or AFLP will indicate the presence of a particular nucleotide variant.

Terminal restriction fragment length polymorphism (TRFLP) works by PCR amplification of DNA using primer pairs that have been labeled with fluorescent tags. The PCR products are digested using RFLP enzymes and the resulting patterns are visualized using a DNA sequencer. The results are analyzed either by counting and comparing bands or peaks in the TRFLP profile, or by comparing bands from one or more TRFLP runs in a database.

The sequence changes directly involved with an RFLP can also be analyzed more quickly by PCR. Amplification can be directed across the altered restriction site, and the products digested with the restriction enzyme. This method has been called Cleaved Amplified Polymorphic Sequence (CAPS). Alternatively, the amplified segment can be analyzed by Allele specific oligonucleotide (ASO) probes, a process that is sometimes assessed using a Dot blot.

A variation on AFLP is cDNA-AFLP, which can be used to quantify differences in gene expression levels.

Another useful approach is the single-stranded conformation polymorphism assay (SSCA), which is based on the altered mobility of a single-stranded target DNA spanning the nucleotide variant

of interest. A single nucleotide change in the target sequence can result in different intramolecular base pairing pattern, and thus different secondary structure of the single-stranded DNA, which can be detected in a non-denaturing gel. See Orita et al., Proc. Natl. Acad. Sci. USA, 86:2776-2770 (1989). Denaturing gel-based techniques such as clamped denaturing gel electrophoresis (CDGE) and denaturing gradient gel electrophoresis (DGGE) detect differences in migration rates of mutant sequences as compared to wild-type sequences in denaturing gel. See Miller et al., Biotechniques, 5:1016-24 (1999); Sheffield et al., Am. J. Hum, Genet., 49:699-706 (1991); Wartell et al., Nucleic Acids Res., 18:2699-2705 (1990); and Sheffield et al., Proc. Natl. Acad. Sci. USA, 86:232-236 (1989). In addition, the double-strand conformation analysis (DSCA) can also be useful in the present methods. See Arguello et al., Nat. Genet., 18:192-194 (1998).

The presence or absence of a nucleotide variant at a particular locus in the one or more genes of an individual can also be detected using the amplification refractory mutation system (ARMS) technique. See e.g., European Patent No. 0,332,435; Newton et al., Nucleic Acids Res., 17:2503-2515 (1989); Fox et al., Br. J. Cancer, 77:1267-1274 (1998); Robertson et al., Eur. Respir. J., 12:477-482 (1998). In the ARMS method, a primer is synthesized matching the nucleotide sequence immediately 5' upstream from the locus being tested except that the 3'-end nucleotide which corresponds to the nucleotide at the locus is a predetermined nucleotide. For example, the 3'-end nucleotide can be the same as that in the mutated locus. The primer can be of any suitable length so long as it hybridizes to the target DNA under stringent conditions only when its 3'-end nucleotide matches the nucleotide at the locus being tested. Preferably the primer has at least 12 nucleotides, more preferably from about 18 to 50 nucleotides. If the individual tested has a mutation at the locus and the nucleotide therein matches the 3'-end nucleotide of the primer, then the primer can be further extended upon hybridizing to the target DNA template, and the primer can initiate a PCR amplification reaction in conjunction with another suitable PCR primer. In contrast, if the nucleotide at the locus is of wild type, then primer extension cannot be achieved. Various forms of ARMS techniques developed in the past few years can be used. See e.g., Gibson et al., Clin. Chem. 43:1336-1341 (1997).

Similar to the ARMS technique is the mini sequencing or single nucleotide primer extension method, which is based on the incorporation of a single nucleotide. An oligonucleotide primer matching the nucleotide sequence immediately 5' to the locus being tested is hybridized to the target DNA, mRNA or miRNA in the presence of labeled dideoxyribonucleotides. A labeled nucleotide is incorporated or linked to the primer only when the dideoxyribonucleotides matches the nucleotide at the variant locus being detected. Thus, the identity of the nucleotide at the variant locus can be revealed based on the detection label attached to the incorporated dideoxyribonucleotides. See Syvanen et al., Genomics, 8:684-692 (1990); Shumaker et al., Hum. Mutat., 7:346-354 (1996); Chen et al., Genome Res., 10:549-547 (2000).

Another set of techniques useful in the present methods is the so-called "oligonucleotide ligation assay" (OLA) in which differentiation between a wild-type locus and a mutation is based on

the ability of two oligonucleotides to anneal adjacent to each other on the target DNA molecule allowing the two oligonucleotides joined together by a DNA ligase. See Landergren et al., Science, 241:1077-1080 (1988); Chen et al, Genome Res., 8:549-556 (1998); Iannone et al., Cytometry, 39:131-140 (2000). Thus, for example, to detect a single-nucleotide mutation at a particular locus in the one or more genes, two oligonucleotides can be synthesized, one having the sequence just 5' upstream from the locus with its 3' end nucleotide being identical to the nucleotide in the variant locus of the particular gene, the other having a nucleotide sequence matching the sequence immediately 3' downstream from the locus in the gene. The oligonucleotides can be labeled for the purpose of detection. Upon hybridizing to the target gene under a stringent condition, the two oligonucleotides are subject to ligation in the presence of a suitable ligase. The ligation of the two oligonucleotides would indicate that the target DNA has a nucleotide variant at the locus being detected.

Detection of small genetic variations can also be accomplished by a variety of hybridization-based approaches. Allele-specific oligonucleotides are most useful. See Conner et al., Proc. Natl. Acad. Sci. USA, 80:278-282 (1983); Saiki et al, Proc. Natl. Acad. Sci. USA, 86:6230-6234 (1989). Oligonucleotide probes (allele-specific) hybridizing specifically to a gene allele having a particular gene variant at a particular locus but not to other alleles can be designed by methods known in the art. The probes can have a length of, e.g., from 10 to about 50 nucleotide bases. The target DNA and the oligonucleotide probe can be contacted with each other under conditions sufficiently stringent such that the nucleotide variant can be distinguished from the wild-type gene based on the presence or absence of hybridization. The probe can be labeled to provide detection signals. Alternatively, the allele-specific oligonucleotide probe can be used as a PCR amplification primer in an "allele-specific PCR" and the presence or absence of a PCR product of the expected length would indicate the presence or absence of a particular nucleotide variant.

Other useful hybridization-based techniques allow two single-stranded nucleic acids annealed together even in the presence of mismatch due to nucleotide substitution, insertion or deletion. The mismatch can then be detected using various techniques. For example, the annealed duplexes can be subject to electrophoresis. The mismatched duplexes can be detected based on their electrophoretic mobility that is different from the perfectly matched duplexes. See Cariello, Human Genetics, 42:726 (1988). Alternatively, in an RNase protection assay, a RNA probe can be prepared spanning the nucleotide variant site to be detected and having a detection marker. See Giunta et al., Diagn. Mol. Path., 5:265-270 (1996); Finkelstein et al., Genomics, 7:167-172 (1990); Kinszler et al., Science 251:1366-1370 (1991). The RNA probe can be hybridized to the target DNA or mRNA forming a heteroduplex that is then subject to the ribonuclease RNase A digestion. RNase A digests the RNA probe in the heteroduplex only at the site of mismatch. The digestion can be determined on a denaturing electrophoresis gel based on size variations. In addition, mismatches can also be detected

by chemical cleavage methods known in the art. See e.g., Roberts et al., Nucleic Acids Res., 25:3377-3378 (1997).

In the mutS assay, a probe can be prepared matching the gene sequence surrounding the locus at which the presence or absence of a mutation is to be detected, except that a predetermined nucleotide is used at the variant locus. Upon annealing the probe to the target DNA to form a duplex, the E. coli mutS protein is contacted with the duplex. Since the mutS protein binds only to heteroduplex sequences containing a nucleotide mismatch, the binding of the mutS protein will be indicative of the presence of a mutation. See Modrich et al., Ann. Rev. Genet., 25:229-253 (1991).

A great variety of improvements and variations have been developed in the art on the basis of the above-described basic techniques which can be useful in detecting mutations or nucleotide variants in the present methods. For example, the "sunrise probes" or "molecular beacons" use the fluorescence resonance energy transfer (FRET) property and give rise to high sensitivity. See Wolf et al., Proc. Nat. Acad. Sci. USA, 85:8790-8794 (1988). Typically, a probe spanning the nucleotide locus to be detected are designed into a hairpin-shaped structure and labeled with a quenching fluorophore at one end and a reporter fluorophore at the other end. In its natural state, the fluorescence from the reporter fluorophore is quenched by the quenching fluorophore due to the proximity of one fluorophore to the other. Upon hybridization of the probe to the target DNA, the 5' end is separated apart from the 3'-end and thus fluorescence signal is regenerated. See Nazarenko et al., Nucleic Acids Res., 25:2516-2521 (1997); Rychlik et al., Nucleic Acids Res., 17:8543-8551 (1989); Sharkey et al., Bio/Technology 12:506-509 (1994); Tyagi et al., Nat. Biotechnol., 14:303-308 (1996); Tyagi et al., Nat. Biotechnol., 16:49-53 (1998). The homo-tag assisted non-dimer system (HANDS) can be used in combination with the molecular beacon methods to suppress primer-dimer accumulation. See Brownie et al., Nucleic Acids Res., 25:3235-3241 (1997).

Dye-labeled oligonucleotide ligation assay is a FRET-based method, which combines the OLA assay and PCR. See Chen et al., Genome Res. 8:549-556 (1998). TaqMan is another FRET-based method for detecting nucleotide variants. A TaqMan probe can be oligonucleotides designed to have the nucleotide sequence of the gene spanning the variant locus of interest and to differentially hybridize with different alleles. The two ends of the probe are labeled with a quenching fluorophore and a reporter fluorophore, respectively. The TaqMan probe is incorporated into a PCR reaction for the amplification of a target gene region containing the locus of interest using Taq polymerase. As Taq polymerase exhibits 5'-3' exonuclease activity but has no 3'-5' exonuclease activity, if the TaqMan probe is annealed to the target DNA template, the 5'-end of the TaqMan probe will be degraded by Taq polymerase during the PCR reaction thus separating the reporting fluorophore from the quenching fluorophore and releasing fluorescence signals. See Holland et al., Proc. Natl. Acad. Sci. USA, 88:7276-7280 (1991); Kalinina et al., Nucleic Acids Res., 25:1999-2004 (1997); Whitcombe et al., Clin. Chem., 44:918-923 (1998).

In addition, the detection in the present methods can also employ a chemiluminescence-based technique. For example, an oligonucleotide probe can be designed to hybridize to either the wild-type or a variant gene locus but not both. The probe is labeled with a highly chemiluminescent acridinium ester. Hydrolysis of the acridinium ester destroys chemiluminescence. The hybridization of the probe to the target DNA prevents the hydrolysis of the acridinium ester. Therefore, the presence or absence of a particular mutation in the target DNA is determined by measuring chemiluminescence changes. See Nelson et al., Nucleic Acids Res., 24:4998-5003 (1996).

The detection of genetic variation in the gene in accordance with the present methods can also be based on the "base excision sequence scanning" (BESS) technique. The BESS method is a PCR-based mutation scanning method. BESS T-Scan and BESS G-Tracker are generated which are analogous to T and G ladders of dideoxy sequencing. Mutations are detected by comparing the sequence of normal and mutant DNA. See, e.g., Hawkins et al., Electrophoresis, 20:1171-1176 (1999).

Mass spectrometry can be used for molecular profiling according to the present methods. See Graber et al., Curr. Opin. Biotechnol., 9:14-18 (1998). For example, in the primer oligo base extension (PROBE™) method, a target nucleic acid is immobilized to a solid-phase support. A primer is annealed to the target immediately 5' upstream from the locus to be analyzed. Primer extension is carried out in the presence of a selected mixture of deoxyribonucleotides and dideoxyribonucleotides. The resulting mixture of newly extended primers is then analyzed by MALDI-TOF. See e.g., Monforte et al., Nat. Med., 3:360-362 (1997).

In addition, the microchip or microarray technologies are also applicable to the detection method of the present methods. Essentially, in microchips, a large number of different oligonucleotide probes are immobilized in an array on a substrate or carrier, e.g., a silicon chip or glass slide. Target nucleic acid sequences to be analyzed can be contacted with the immobilized oligonucleotide probes on the microchip. See Lipshutz et al., Biotechniques, 19:442-447 (1995); Chee et al., Science, 274:610-614 (1996); Kozal et al., Nat. Med. 2:753-759 (1996); Hacia et al., Nat. Genet., 14:441-447 (1996); Saiki et al., Proc. Natl. Acad. Sci. USA, 86:6230-6234 (1989); Gingeras et al., Genome Res., 8:435-448 (1998). Alternatively, the multiple target nucleic acid sequences to be studied are fixed onto a substrate and an array of probes is contacted with the immobilized target sequences. See Drmanac et al., Nat. Biotechnol., 16:54-58 (1998). Numerous microchip technologies have been developed incorporating one or more of the above described techniques for detecting mutations. The microchip technologies combined with computerized analysis tools allow fast screening in a large scale. The adaptation of the microchip technologies to the present methods will be apparent to a person of skill in the art apprised of the present disclosure. See, e.g., U.S. Pat. No. 5,925,525 to Fodor et al; Wilgenbus et al., J. Mol. Med., 77:761-786 (1999); Graber et al., Curr. Opin. Biotechnol., 9:14-18 (1998); Hacia et al., Nat. Genet., 14:441-447 (1996); Shoemaker et al., Nat. Genet., 14:450-456 (1996); DeRisi et al., Nat. Genet., 14:457-460 (1996); Chee et al., Nat. Genet., 14:610-614 (1996); Lockhart et al., Nat. Genet., 14:675-680 (1996); Drobyshev et al., Gene, 188:45-52 (1997).

As is apparent from the above survey of the suitable detection techniques, it may or may not be necessary to amplify the target DNA, i.e., the gene, cDNA, mRNA, miRNA, or a portion thereof to increase the number of target DNA molecule, depending on the detection techniques used. For example, most PCR-based techniques combine the amplification of a portion of the target and the detection of the mutations. PCR amplification is well known in the art and is disclosed in U.S. Pat. Nos. 4,683,195 and 4,800,159, both which are incorporated herein by reference. For non-PCR-based detection techniques, if necessary, the amplification can be achieved by, e.g., in vivo plasmid multiplication, or by purifying the target DNA from a large amount of tissue or cell samples. See generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989. However, even with scarce samples, many sensitive techniques have been developed in which small genetic variations such as single-nucleotide substitutions can be detected without having to amplify the target DNA in the sample. For example, techniques have been developed that amplify the signal as opposed to the target DNA by, e.g., employing branched DNA or dendrimers that can hybridize to the target DNA. The branched or dendrimer DNAs provide multiple hybridization sites for hybridization probes to attach thereto thus amplifying the detection signals. See Detmer et al., J. Clin. Microbiol., 34:901-907 (1996); Collins et al., Nucleic Acids Res., 25:2979-2984 (1997); Horn et al., Nucleic Acids Res., 25:4835-4841 (1997); Horn et al., Nucleic Acids Res., 25:4842-4849 (1997); Nilsen et al., J. Theor. Biol., 187:273-284 (1997).

The Invader™ assay is another technique for detecting single nucleotide variations that can be used for molecular profiling according to the methods. The Invader™ assay uses a novel linear signal amplification technology that improves upon the long turnaround times required of the typical PCR DNA sequenced-based analysis. See Cooksey et al., Antimicrobial Agents and Chemotherapy 44:1296-1301 (2000). This assay is based on cleavage of a unique secondary structure formed between two overlapping oligonucleotides that hybridize to the target sequence of interest to form a "flap." Each "flap" then generates thousands of signals per hour. Thus, the results of this technique can be easily read, and the methods do not require exponential amplification of the DNA target. The Invader™ system uses two short DNA probes, which are hybridized to a DNA target. The structure formed by the hybridization event is recognized by a special cleavase enzyme that cuts one of the probes to release a short DNA "flap." Each released "flap" then binds to a fluorescently-labeled probe to form another cleavage structure. When the cleavase enzyme cuts the labeled probe, the probe emits a detectable fluorescence signal. See e.g. Lyamichev et al., Nat. Biotechnol., 17:292-296 (1999).

The rolling circle method is another method that avoids exponential amplification. Lizardi et al., Nature Genetics, 19:225-232 (1998) (which is incorporated herein by reference). For example, Sniper™, a commercial embodiment of this method, is a sensitive, high-throughput SNP scoring system designed for the accurate fluorescent detection of specific variants. For each nucleotide variant, two linear, allele-specific probes are designed. The two allele-specific probes are identical

with the exception of the 3'-base, which is varied to complement the variant site. In the first stage of the assay, target DNA is denatured and then hybridized with a pair of single, allele-specific, open-circle oligonucleotide probes. When the 3'-base exactly complements the target DNA, ligation of the probe will preferentially occur. Subsequent detection of the circularized oligonucleotide probes is by rolling circle amplification, whereupon the amplified probe products are detected by fluorescence. See Clark and Pickering, Life Science News 6, 2000, Amersham Pharmacia Biotech (2000).

A number of other techniques that avoid amplification all together include, e.g., surface-enhanced resonance Raman scattering (SERRS), fluorescence correlation spectroscopy, and single-molecule electrophoresis. In SERRS, a chromophore-nucleic acid conjugate is absorbed onto colloidal silver and is irradiated with laser light at a resonant frequency of the chromophore. See Graham et al., Anal. Chem., 69:4703-4707 (1997). The fluorescence correlation spectroscopy is based on the spatio-temporal correlations among fluctuating light signals and trapping single molecules in an electric field. See Eigen et al., Proc. Natl. Acad. Sci. USA, 91:5740-5747 (1994). In single-molecule electrophoresis, the electrophoretic velocity of a fluorescently tagged nucleic acid is determined by measuring the time required for the molecule to travel a predetermined distance between two laser beams. See Castro et al., Anal. Chem., 67:3181-3186 (1995).

In addition, the allele-specific oligonucleotides (ASO) can also be used in in situ hybridization using tissues or cells as samples. The oligonucleotide probes which can hybridize differentially with the wild-type gene sequence or the gene sequence harboring a mutation may be labeled with radioactive isotopes, fluorescence, or other detectable markers. In situ hybridization techniques are well known in the art and their adaptation to the present methods for detecting the presence or absence of a nucleotide variant in the one or more gene of a particular individual should be apparent to a skilled artisan apprised of this disclosure.

Accordingly, the presence or absence of one or more genes nucleotide variant or amino acid variant in an individual can be determined using any of the detection methods described above.

Typically, once the presence or absence of one or more gene nucleotide variants or amino acid variants is determined, physicians or genetic counselors or patients or other researchers may be informed of the result. Specifically the result can be cast in a transmittable form that can be communicated or transmitted to other researchers or physicians or genetic counselors or patients. Such a form can vary and can be tangible or intangible. The result with regard to the presence or absence of a nucleotide variant of the present methods in the individual tested can be embodied in descriptive statements, diagrams, photographs, charts, images or any other visual forms. For example, images of gel electrophoresis of PCR products can be used in explaining the results. Diagrams showing where a variant occurs in an individual's gene are also useful in indicating the testing results. The statements and visual forms can be recorded on a tangible media such as papers, computer readable media such as floppy disks, compact disks, etc., or on an intangible media, e.g., an electronic media in the form of email or website on internet or intranet. In addition, the result with regard to the

presence or absence of a nucleotide variant or amino acid variant in the individual tested can also be recorded in a sound form and transmitted through any suitable media, e.g., analog or digital cable lines, fiber optic cables, etc., via telephone, facsimile, wireless mobile phone, internet phone and the like.

Thus, the information and data on a test result can be produced anywhere in the world and transmitted to a different location. For example, when a genotyping assay is conducted offshore, the information and data on a test result may be generated and cast in a transmittable form as described above. The test result in a transmittable form thus can be imported into the U.S. Accordingly, the present methods also encompasses a method for producing a transmittable form of information on the genotype of the two or more suspected cancer samples from an individual. The method comprises the steps of (1) determining the genotype of the DNA from the samples according to methods of the present methods; and (2) embodying the result of the determining step in a transmittable form. The transmittable form is the product of the production method.

In Situ Hybridization

In situ hybridization assays are well known and are generally described in Angerer et al., Methods Enzymol. 152:649-660 (1987). In an in situ hybridization assay, cells, e.g., from a biopsy, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled, e.g., with radioisotopes or fluorescent reporters, or enzymatically. FISH (fluorescence in situ hybridization) uses fluorescent probes that bind to only those parts of a sequence with which they show a high degree of sequence similarity. CISH (chromogenic in situ hybridization) uses conventional peroxidase or alkaline phosphatase reactions visualized under a standard bright-field microscope.

In situ hybridization can be used to detect specific gene sequences in tissue sections or cell preparations by hybridizing the complementary strand of a nucleotide probe to the sequence of interest. Fluorescent in situ hybridization (FISH) uses a fluorescent probe to increase the sensitivity of in situ hybridization.

FISH is a cytogenetic technique used to detect and localize specific polynucleotide sequences in cells. For example, FISH can be used to detect DNA sequences on chromosomes. FISH can also be used to detect and localize specific RNAs, e.g., mRNAs, within tissue samples. In FISH uses fluorescent probes that bind to specific nucleotide sequences to which they show a high degree of sequence similarity. Fluorescence microscopy can be used to find out whether and where the fluorescent probes are bound. In addition to detecting specific nucleotide sequences, e.g., translocations, fusion, breaks, duplications and other chromosomal abnormalities, FISH can help define the spatial-temporal patterns of specific gene copy number and/or gene expression within cells and tissues.

Various types of FISH probes can be used to detect chromosome translocations. Dual color, single fusion probes can be useful in detecting cells possessing a specific chromosomal translocation. The DNA probe hybridization targets are located on one side of each of the two genetic breakpoints. "Extra signal" probes can reduce the frequency of normal cells exhibiting an abnormal FISH pattern due to the random co-localization of probe signals in a normal nucleus. One large probe spans one breakpoint, while the other probe flanks the breakpoint on the other gene. Dual color, break apart probes are useful in cases where there may be multiple translocation partners associated with a known genetic breakpoint. This labeling scheme features two differently colored probes that hybridize to targets on opposite sides of a breakpoint in one gene. Dual color, dual fusion probes can reduce the number of normal nuclei exhibiting abnormal signal patterns. The probe offers advantages in detecting low levels of nuclei possessing a simple balanced translocation. Large probes span two breakpoints on different chromosomes. Such probes are available as Vysis probes from Abbott Laboratories, Abbott Park, IL.

CISH, or chromogenic in situ hybridization, is a process in which a labeled complementary DNA or RNA strand is used to localize a specific DNA or RNA sequence in a tissue specimen. CISH methodology can be used to evaluate gene amplification, gene deletion, chromosome translocation, and chromosome number. CISH can use conventional enzymatic detection methodology, e.g., horseradish peroxidase or alkaline phosphatase reactions, visualized under a standard bright-field microscope. In a common embodiment, a probe that recognizes the sequence of interest is contacted with a sample. An antibody or other binding agent that recognizes the probe, e.g., via a label carried by the probe, can be used to target an enzymatic detection system to the site of the probe. In some systems, the antibody can recognize the label of a FISH probe, thereby allowing a sample to be analyzed using both FISH and CISH detection. CISH can be used to evaluate nucleic acids in multiple settings, e.g., formalin-fixed, paraffin-embedded (FFPE) tissue, blood or bone marrow smear, metaphase chromosome spread, and/or fixed cells. In an embodiment, CISH is performed following the methodology in the SPoT-Light® HER2 CISH Kit available from Life Technologies (Carlsbad, CA) or similar CISH products available from Life Technologies. The SPoT-Light® HER2 CISH Kit itself is FDA approved for in vitro diagnostics and can be used for molecular profiling of HER2. CISH can be used in similar applications as FISH. Thus, one of skill will appreciate that reference to molecular profiling using FISH herein can be performed using CISH, unless otherwise specified.

Silver-enhanced in situ hybridization (SISH) is similar to CISH, but with SISH the signal appears as a black coloration due to silver precipitation instead of the chromogen precipitates of CISH.

Modifications of the in situ hybridization techniques can be used for molecular profiling according to the methods. Such modifications comprise simultaneous detection of multiple targets, e.g., Dual ISH, Dual color CISH, bright field double in situ hybridization (BDISH). *See e.g.,* the FDA

approved INFORM HER2 Dual ISH DNA Probe Cocktail kit from Ventana Medical Systems, Inc. (Tucson, AZ); DuoCISH™, a dual color CISH kit developed by Dako Denmark A/S (Denmark).

Comparative Genomic Hybridization (CGH) comprises a molecular cytogenetic method of screening tumor samples for genetic changes showing characteristic patterns for copy number changes at chromosomal and subchromosomal levels. Alterations in patterns can be classified as DNA gains and losses. CGH employs the kinetics of in situ hybridization to compare the copy numbers of different DNA or RNA sequences from a sample, or the copy numbers of different DNA or RNA sequences in one sample to the copy numbers of the substantially identical sequences in another sample. In many useful applications of CGH, the DNA or RNA is isolated from a subject cell or cell population. The comparisons can be qualitative or quantitative. Procedures are described that permit determination of the absolute copy numbers of DNA sequences throughout the genome of a cell or cell population if the absolute copy number is known or determined for one or several sequences. The different sequences are discriminated from each other by the different locations of their binding sites when hybridized to a reference genome, usually metaphase chromosomes but in certain cases interphase nuclei. The copy number information originates from comparisons of the intensities of the hybridization signals among the different locations on the reference genome. The methods, techniques and applications of CGH are known, such as described in U.S. Pat. No. 6,335,167, and in U.S. App. Ser. No. 60/804,818, the relevant parts of which are herein incorporated by reference.

In an embodiment, CGH used to compare nucleic acids between diseased and healthy tissues. The method comprises isolating DNA from disease tissues (e.g., tumors) and reference tissues (e.g., healthy tissue) and labeling each with a different "color" or fluor. The two samples are mixed and hybridized to normal metaphase chromosomes. In the case of array or matrix CGH, the hybridization mixing is done on a slide with thousands of DNA probes. A variety of detection system can be used that basically determine the color ratio along the chromosomes to determine DNA regions that might be gained or lost in the diseased samples as compared to the reference.

Molecular Profiling Methods

FIG. 1I illustrates a block diagram of an illustrative embodiment of a system **10** for determining individualized medical intervention for a particular disease state that uses molecular profiling of a patient's biological specimen. System **10** includes a user interface **12**, a host server **14** including a processor **16** for processing data, a memory **18** coupled to the processor, an application program **20** stored in the memory **18** and accessible by the processor **16** for directing processing of the data by the processor **16**, a plurality of internal databases **22** and external databases **24**, and an interface with a wired or wireless communications network **26** (such as the Internet, for example). System **10** may also include an input digitizer **28** coupled to the processor **16** for inputting digital data from data that is received from user interface **12**.

User interface **12** includes an input device **30** and a display **32** for inputting data into system **10** and for displaying information derived from the data processed by processor **16**. User interface **12** may also include a printer **34** for printing the information derived from the data processed by the processor **16** such as patient reports that may include test results for targets and proposed drug therapies based on the test results.

Internal databases **22** may include, but are not limited to, patient biological sample/specimen information and tracking, clinical data, patient data, patient tracking, file management, study protocols, patient test results from molecular profiling, and billing information and tracking. External databases **24** nay include, but are not limited to, drug libraries, gene libraries, disease libraries, and public and private databases such as UniGene, OMIM, GO, TIGR, GenBank, KEGG and Biocarta.

Various methods may be used in accordance with system **10**. **FIGs. 2A-C** shows a flowchart of an illustrative embodiment of a method for determining individualized medical intervention for a particular disease state that uses molecular profiling of a patient's biological specimen that is non disease specific. In order to determine a medical intervention for a particular disease state using molecular profiling that is independent of disease lineage diagnosis (i.e., not single disease restricted), at least one molecular test is performed on the biological sample of a diseased patient. Biological samples are obtained from diseased patients by taking a biopsy of a tumor, conducting minimally invasive surgery if no recent tumor is available, obtaining a sample of the patient's blood, or a sample of any other biological fluid including, but not limited to, cell extracts, nuclear extracts, cell lysates or biological products or substances of biological origin such as excretions, blood, sera, plasma, urine, sputum, tears, feces, saliva, membrane extracts, and the like.

A target can be any molecular finding that may be obtained from molecular testing. For example, a target may include one or more genes or proteins. For example, the presence of a copy number variation of a gene can be determined. As shown in **FIG. 2**, tests for finding such targets can include, but are not limited to, NGS, IHC, fluorescent in-situ hybridization (FISH), in-situ hybridization (ISH), and other molecular tests known to those skilled in the art.

Furthermore, the methods disclosed herein include profiling more than one target. As a non-limiting example, the copy number, or presence of a copy number variation (CNV), of a plurality of genes can be identified. Furthermore, identification of a plurality of targets in a sample can be by one method or by various means. For example, the presence of a CNV of a first gene can be determined by one method, e.g., NGS, and the presence of a CNV of a second gene determined by a different method, e.g., fragment analysis. Alternatively, the same method can be used to detect the presence of a CNV in both the first and second gene, e.g., using NGS.

The test results can be compiled to determine the individual characteristics of the cancer. After determining the characteristics of the cancer, a therapeutic regimen may be identified, e.g., comprising treatments of likely benefit as well as treatments of unlikely benefit.

Finally, a patient profile report may be provided which includes the patient's test results for various targets and any proposed therapies based on those results.

The systems as described herein can be used to automate the steps of identifying a molecular profile to assess a cancer. In an aspect, the present methods can be used for generating a report comprising a molecular profile. The methods can comprise: performing molecular profiling on a sample from a subject to assess characteristics of a plurality of cancer biomarkers, and compiling a report comprising the assessed characteristics into a list, thereby generating a report that identifies a molecular profile for the sample. The report can further comprise a list describing the potential benefit of the plurality of treatment options based on the assessed characteristics, thereby identifying candidate treatment options for the subject. The report can also suggest treatments of potential unlikely benefit, or indeterminate benefit, based on the assessed characteristics.

Molecular Profiling for Treatment Selection

The methods as described herein provide a candidate treatment selection for a subject in need thereof. Molecular profiling can be used to identify one or more candidate therapeutic agents for an individual suffering from a condition in which one or more of the biomarkers disclosed herein are targets for treatment. For example, the method can identify one or more chemotherapy treatments for a cancer. In an aspect, the methods provides a method comprising: performing at least one molecular profiling technique on at least one biomarker. Any relevant biomarker can be assessed using one or more of the molecular profiling techniques described herein or known in the art. The marker need only have some direct or indirect association with a treatment to be useful. Any relevant molecular profiling technique can be performed, such as those disclosed here. These can include without limitation, protein and nucleic acid analysis techniques. Protein analysis techniques include, by way of non-limiting examples, immunoassays, immunohistochemistry, and mass spectrometry. Nucleic acid analysis techniques include, by way of non-limiting examples, amplification, polymerase chain amplification, hybridization, microarrays, in situ hybridization, sequencing, dye-terminator sequencing, next generation sequencing, pyrosequencing, and restriction fragment analysis.

Molecular profiling may comprise the profiling of at least one gene (or gene product) for each assay technique that is performed. Different numbers of genes can be assayed with different techniques. Any marker disclosed herein that is associated directly or indirectly with a target therapeutic can be assessed. For example, any "druggable target" comprising a target that can be modulated with a therapeutic agent such as a small molecule or binding agent such as an antibody, is a candidate for inclusion in the molecular profiling methods as described herein. The target can also be indirectly drug associated, such as a component of a biological pathway that is affected by the associated drug. The molecular profiling can be based on either the gene, e.g., DNA sequence, and/or gene product, e.g., mRNA or protein. Such nucleic acid and/or polypeptide can be profiled as applicable as to presence or absence, level or amount, activity, mutation, sequence, haplotype,

rearrangement, copy number, or other measurable characteristic. In some embodiments, a single gene and/or one or more corresponding gene products is assayed by more than one molecular profiling technique. A gene or gene product (also referred to herein as "marker" or "biomarker"), e.g., an mRNA or protein, is assessed using applicable techniques (e.g., to assess DNA, RNA, protein), including without limitation ISH, gene expression, IHC, sequencing or immunoassay. Therefore, any of the markers disclosed herein can be assayed by a single molecular profiling technique or by multiple methods disclosed herein (e.g., a single marker is profiled by one or more of IHC, ISH, sequencing, microarray, etc.). In some embodiments, 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, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or at least about 100 genes or gene products are profiled by at least one technique, a plurality of techniques, or using any desired combination of ISH, IHC, gene expression, gene copy, and sequencing. In some embodiments, at least about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000, 30,000, 31,000, 32,000, 33,000, 34,000, 35,000, 36,000, 37,000, 38,000, 39,000, 40,000, 41,000, 42,000, 43,000, 44,000, 45,000, 46,000, 47,000, 48,000, 49,000, or at least 50,000 genes or gene products are profiled using various techniques. The number of markers assayed can depend on the technique used. For example, microarray and massively parallel sequencing lend themselves to high throughput analysis. Because molecular profiling queries molecular characteristics of the tumor itself, this approach provides information on therapies that might not otherwise be considered based on the lineage of the tumor.

In some embodiments, a sample from a subject in need thereof is profiled using methods which include but are not limited to IHC analysis, gene expression analysis, ISH analysis, and/or sequencing analysis (such as by PCR, RT-PCR, pyrosequencing, NGS) for one or more of the following: ABCC1, ABCG2, ACE2, ADA, ADH1C, ADH4, AGT, AR, AREG, ASNS, BCL2, BCRP, BDCA1, beta III tubulin, BIRC5, B-RAF, BRCA1, BRCA2, CA2, caveolin, CD20, CD25, CD33, CD52, CDA, CDKN2A, CDKN1A, CDKN1B, CDK2, CDW52, CES2, CK 14, CK 17, CK 5/6, c-KIT, c-Met, c-Myc, COX-2, Cyclin D1, DCK, DHFR, DNMT1, DNMT3A, DNMT3B, E-Cadherin, ECGF1, EGFR, EML4-ALK fusion, EPHA2, Epiregulin, ER, ERBR2, ERCC1, ERCC3, EREG, ESR1, FLT1, folate receptor, FOLR1, FOLR2, FSHB, FSHPRH1, FSHR, FYN, GART, GNA11, GNAQ, GNRH1, GNRHR1, GSTP1, HCK, HDAC1, hENT-1, Her2/Neu, HGF, HIF1A, HIG1, HSP90, HSP90AA1, HSPCA, IGF-1R, IGFRBP, IGFRBP3, IGFRBP4, IGFRBP5, IL13RA1, IL2RA, KDR, Ki67, KIT, K-RAS, LCK, LTB, Lymphotoxin Beta Receptor, LYN, MET, MGMT, MLH1, MMR, MRP1, MS4A1, MSH2, MSH5, Myc, NFKB1, NFKB2, NFKBIA, NRAS, ODC1, OGFR, p16, p21, p27, p53, p95, PARP-1, PDGFC, PDGFR, PDGFRA, PDGFRB, PGP, PGR, PI3K, POLA, POLA1, PPARG, PPARGC1, PR, PTEN, PTGS2, PTPN12, RAF1, RARA, ROS1, RRM1, RRM2, RRM2B, RXRB, RXRG, SIK2, SPARC, SRC, SSTR1, SSTR2, SSTR3, SSTR4, SSTR5, Survivin,

TK1, TLE3, TNF, TOP1, TOP2A, TOP2B, TS, TUBB3, TXN, TXNRD1, TYMS, VDR, VEGF, VEGFA, VEGFC, VHL, YES1, ZAP70, a biomarker listed in any one of **Tables 2-116**, **Tables 117-120**, ISNM1, **Tables 121-130**, and any useful combination thereof.

As understood by those of skill in the art, genes and proteins have developed a number of alternative names in the scientific literature. Listing of gene aliases and descriptions used herein can be found using a variety of online databases, including GeneCards® (www.genecards.org), HUGO Gene Nomenclature (www.genenames.org), Entrez Gene (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene), UniProtKB/Swiss-Prot (www.uniprot.org), UniProtKB/TrEMBL (www.uniprot.org), OMIM (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM), GeneLoc (genecards.weizmann.ac.il/geneloc/), and Ensembl (www.ensembl.org). For example, gene symbols and names used herein can correspond to those approved by HUGO, and protein names can be those recommended by UniProtKB/Swiss-Prot. In the specification, where a protein name indicates a precursor, the mature protein is also implied. Throughout the application, gene and protein symbols may be used interchangeably and the meaning can be derived from context, e.g., ISH or NGS can be used to analyze nucleic acids whereas IHC is used to analyze protein.

The choice of genes and gene products to be assessed to provide molecular profiles as described herein can be updated over time as new treatments and new drug targets are identified. For example, once the expression or mutation of a biomarker is correlated with a treatment option, it can be assessed by molecular profiling. One of skill will appreciate that such molecular profiling is not limited to those techniques disclosed herein but comprises any methodology conventional for assessing nucleic acid or protein levels, sequence information, or both. The methods as described herein can also take advantage of any improvements to current methods or new molecular profiling techniques developed in the future. In some embodiments, a gene or gene product is assessed by a single molecular profiling technique. In other embodiments, a gene and/or gene product is assessed by multiple molecular profiling techniques. In a non-limiting example, a gene sequence can be assayed by one or more of NGS, ISH and pyrosequencing analysis, the mRNA gene product can be assayed by one or more of NGS, RT-PCR and microarray, and the protein gene product can be assayed by one or more of IHC and immunoassay. One of skill will appreciate that any combination of biomarkers and molecular profiling techniques that will benefit disease treatment are contemplated by the present methods.

Genes and gene products that are known to play a role in cancer and can be assayed by any of the molecular profiling techniques as described herein include without limitation those listed in any of International Patent Publications WO/2007/137187 (Int'l Appl. No. PCT/US2007/069286), published November 29, 2007; WO/2010/045318 (Int'l Appl. No. PCT/US2009/060630), published April 22, 2010; WO/2010/093465 (Int'l Appl. No. PCT/US2010/000407), published August 19, 2010; WO/2012/170715 (Int'l Appl. No. PCT/US2012/041393), published December 13, 2012;

WO/2014/089241 (Int'l Appl. No. PCT/US2013/073184), published June 12, 2014; WO/2011/056688 (Int'l Appl. No. PCT/US2010/054366), published May 12, 2011; WO/2012/092336 (Int'l Appl. No. PCT/US2011/067527), published July 5, 2012; WO/2015/116868 (Int'l Appl. No. PCT/US2015/013618), published August 6, 2015; WO/2017/053915 (Int'l Appl. No. PCT/US2016/053614), published March 30, 2017; WO/2016/141169 (Int'l Appl. No. PCT/US2016/020657), published September 9, 2016; and WO2018175501 (Int'l Appl. No. PCT/US2018/023438), published September 27, 2018; each of which publications is incorporated by reference herein in its entirety.

Mutation profiling can be determined by sequencing, including Sanger sequencing, array sequencing, pyrosequencing, high-throughput or next generation (NGS, NextGen) sequencing, etc. Sequence analysis may reveal that genes harbor activating mutations so that drugs that inhibit activity are indicated for treatment. Alternately, sequence analysis may reveal that genes harbor mutations that inhibit or eliminate activity, thereby indicating treatment for compensating therapies. In some embodiments, sequence analysis comprises that of exon 9 and 11 of c-KIT. Sequencing may also be performed on EGFR-kinase domain exons 18, 19, 20, and 21. Mutations, amplifications or misregulations of EGFR or its family members are implicated in about 30% of all epithelial cancers. Sequencing can also be performed on PI3K, encoded by the PIK3CA gene. This gene is a found mutated in many cancers. Sequencing analysis can also comprise assessing mutations in one or more ABCC1, ABCG2, ADA, AR, ASNS, BCL2, BIRC5, BRCA1, BRCA2, CD33, CD52, CDA, CES2, DCK, DHFR, DNMT1, DNMT3A, DNMT3B, ECGF1, EGFR, EPHA2, ERBB2, ERCC1, ERCC3, ESR1, FLT1, FOLR2, FYN, GART, GNRH1, GSTP1, HCK, HDAC1, HIF1A, HSP90AA1, IGFBP3, IGFBP4, IGFBP5, IL2RA, KDR, KIT, LCK, LYN, MET, MGMT, MLH1, MS4A1, MSH2, NFKB1, NFKB2, NFKBIA, NRAS, OGFR, PARP1, PDGFC, PDGFRA, PDGFRB, PGP, PGR, POLA1, PTEN, PTGS2, PTPN12, RAF1, RARA, RRM1, RRM2, RRM2B, RXRB, RXRG, SIK2, SPARC, SRC, SSTR1, SSTR2, SSTR3, SSTR4, SSTR5, TK1, TNF, TOP1, TOP2A, TOP2B, TXNRD1, TYMS, VDR, VEGFA, VHL, YES1, and ZAP70. One or more of the following genes can also be assessed by sequence analysis: ALK, EML4, hENT-1, IGF-1R, HSP90AA1, MMR, p16, p21, p27, PARP-1, PI3K and TLE3. The genes and/or gene products used for mutation or sequence analysis can be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 or all of the genes and/or gene products listed in any of **Tables 4-12** of WO2018175501, e.g., in any of **Tables 5-10** of WO2018175501, or in any of **Tables 7-10** of WO2018175501.

In embodiments, the methods as described herein are used detect gene fusions, such as those listed in any of International Patent Publications WO/2007/137187 (Int'l Appl. No. PCT/US2007/069286), published November 29, 2007; WO/2010/045318 (Int'l Appl. No. PCT/US2009/060630), published April 22, 2010; WO/2010/093465 (Int'l Appl. No. PCT/US2010/000407), published August 19, 2010; WO/2012/170715 (Int'l Appl. No. PCT/US2012/041393), published December 13, 2012; WO/2014/089241 (Int'l Appl. No.

PCT/US2013/073184), published June 12, 2014; WO/2011/056688 (Int'l Appl. No. PCT/US2010/054366), published May 12, 2011; WO/2012/092336 (Int'l Appl. No. PCT/US2011/067527), published July 5, 2012; WO/2015/116868 (Int'l Appl. No. PCT/US2015/013618), published August 6, 2015; WO/2017/053915 (Int'l Appl. No. PCT/US2016/053614), published March 30, 2017; WO/2016/141169 (Int'l Appl. No. PCT/US2016/020657), published September 9, 2016; and WO/2018/175501 (Int'l Appl. No. PCT/US2018/023438), published September 27, 2018; each of which publications is incorporated by reference herein in its entirety. A fusion gene is a hybrid gene created by the juxtaposition of two previously separate genes. This can occur by chromosomal translocation or inversion, deletion or via trans-splicing. The resulting fusion gene can cause abnormal temporal and spatial expression of genes, leading to abnormal expression of cell growth factors, angiogenesis factors, tumor promoters or other factors contributing to the neoplastic transformation of the cell and the creation of a tumor. For example, such fusion genes can be oncogenic due to the juxtaposition of: 1) a strong promoter region of one gene next to the coding region of a cell growth factor, tumor promoter or other gene promoting oncogenesis leading to elevated gene expression, or 2) due to the fusion of coding regions of two different genes, giving rise to a chimeric gene and thus a chimeric protein with abnormal activity. Fusion genes are characteristic of many cancers. Once a therapeutic intervention is associated with a fusion, the presence of that fusion in any type of cancer identifies the therapeutic intervention as a candidate therapy for treating the cancer.

The presence of fusion genes can be used to guide therapeutic selection. For example, the BCR-ABL gene fusion is a characteristic molecular aberration in ~90% of chronic myelogenous leukemia (CML) and in a subset of acute leukemias (Kurzrock *et al., Annals of Internal Medicine* 2003; **138**:819-830). The BCR-ABL results from a translocation between chromosomes 9 and 22, commonly referred to as the Philadelphia chromosome or Philadelphia translocation. The translocation brings together the 5' region of the BCR gene and the 3' region of ABL1, generating a chimeric BCR-ABL1 gene, which encodes a protein with constitutively active tyrosine kinase activity (Mittleman et al., Nature *Reviews Cancer* 2007; 7:233-245). The aberrant tyrosine kinase activity leads to de-regulated cell signaling, cell growth and cell survival, apoptosis resistance and growth factor independence, all of which contribute to the pathophysiology of leukemia (Kurzrock et al., *Annals of Internal Medicine* 2003; **138**:819-830). Patients with the Philadelphia chromosome are treated with imatinib and other targeted therapies. Imatinib binds to the site of the constitutive tyrosine kinase activity of the fusion protein and prevents its activity. Imatinib treatment has led to molecular responses (disappearance of BCR-ABL+ blood cells) and improved progression-free survival in BCR-ABL+ CML patients (Kantarjian *et al., Clinical Cancer Research* 2007; **13**:1089-1097).

Another fusion gene, IGH-MYC, is a defining feature of ~80% of Burkitt's lymphoma (Ferry et al. *Oncologist* 2006; 11:375-83). The causal event for this is a translocation between chromosomes 8 and 14, bringing the c-Myc oncogene adjacent to the strong promoter of the immunoglobulin heavy

chain gene, causing *c-myc* overexpression (Mittleman et al., *Nature Reviews Cancer* 2007; 7:233-245). The c-*myc* rearrangement is a pivotal event in lymphomagenesis as it results in a perpetually proliferative state. It has wide ranging effects on progression through the cell cycle, cellular differentiation, apoptosis, and cell adhesion (Ferry et al. *Oncologist* 2006; 11:375-83).

A number of recurrent fusion genes have been catalogued in the Mittleman database (cgap.nci.nih.gov/Chromosomes/Mitelman). The gene fusions can be used to characterize neoplasms and cancers and guide therapy using the subject methods described herein. For example, TMPRSS2-ERG, TMPRSS2-ETV and SLC45A3-ELK4 fusions can be detected to characterize prostate cancer; and ETV6-NTRK3 and ODZ4-NRG1 can be used to characterize breast cancer. The EML4-ALK, RLF-MYCL1, TGF-ALK, or CD74-ROS1 fusions can be used to characterize a lung cancer. The ACSL3-ETV1, C15ORF21-ETV1, FLJ35294-ETV1, HERV-ETV1, TMPRSS2-ERG, TMPRSS2-ETV1/4/5, TMPRSS2-ETV4/5, SLC5A3-ERG, SLC5A3-ETV1, SLC5A3-ETV5 or KLK2-ETV4 fusions can be used to characterize a prostate cancer. The GOPC-ROS1 fusion can be used to characterize a brain cancer. The CHCHD7-PLAG1, CTNNB1-PLAG1, FHIT-HMGA2, HMGA2-NFIB, LIFR-PLAG1, or TCEA1-PLAG1 fusions can be used to characterize a head and neck cancer. The ALPHA-TFEB, NONO-TFE3, PRCC-TFE3, SFPQ-TFE3, CLTC-TFE3, or MALAT1-TFEB fusions can be used to characterize a renal cell carcinoma (RCC). The AKAP9-BRAF, CCDC6-RET, ERC1-RETM, GOLGA5-RET, HOOK3-RET, HRH4-RET, KTN1-RET, NCOA4-RET, PCM1-RET, PRKARA1A-RET, RFG-RET, RFG9-RET, Ria-RET, TGF-NTRK1, TPM3-NTRK1, TPM3-TPR, TPR-MET, TPR-NTRK1, TRIM24-RET, TRIM27-RET or TRIM33-RET fusions can be used to characterize a thyroid cancer and/or papillary thyroid carcinoma; and the PAX8-PPARy fusion can be analyzed to characterize a follicular thyroid cancer. Fusions that are associated with hematological malignancies include without limitation TTL-ETV6, CDK6-MLL, CDK6-TLX3, ETV6-FLT3, ETV6-RUNX1, ETV6-TTL, MLL-AFF1, MLL-AFF3, MLL-AFF4, MLL-GAS7, TCBA1-ETV6, TCF3-PBX1 or TCF3-TFPT, which are characteristic of acute lymphocytic leukemia (ALL); BCL11B-TLX3, IL2-TNFRFS17, NUP214-ABL1, NUP98-CCDC28A, TAL1-STIL, or ETV6-ABL2, which are characteristic of T-cell acute lymphocytic leukemia (T-ALL); ATIC-ALK, KIAA1618-ALK, MSN-ALK, MYH9-ALK, NPM1-ALK, TGF-ALK or TPM3-ALK, which are characteristic of anaplastic large cell lymphoma *(*ALCL); BCR-ABL1, BCR-JAK2, ETV6-EVI1, ETV6-MN1 or ETV6-TCBA1, characteristic of chronic myelogenous leukemia (CML); CBFB-MYH11, CHIC2-ETV6, ETV6-ABL1, ETV6-ABL2, ETV6-ARNT, ETV6-CDX2, ETV6-HLXB9, ETV6-PER1, MEF2D-DAZAP1, AML-AFF1, MLL-ARHGAP26, MLL-ARHGEF12, MLL-CASC5, MLL-CBL,MLL-CREBBP, MLL-DAB21P, MLL-ELL, MLL-EP300, MLL-EPS15, MLL-FNBP1, MLL-FOXO3A, MLL-GMPS, MLL-GPHN, MLL-MLLT1, MLL-MLLT11, MLL-MLLT3, MLL-MLLT6, MLL-MYO1F, MLL-PICALM, MLL-SEPT2, MLL-SEPT6, MLL-SORBS2, MYST3-SORBS2, MYST-CREBBP, NPM1-MLF1, NUP98-HOXA13, PRDM16-EVI1, RABEP1-PDGFRB, RUNX1-EVI1, RUNX1-MDS1, RUNX1-RPL22, RUNX1-RUNX1T1, RUNX1-SH3D19, RUNX1-USP42,

RUNX1-YTHDF2, RUNX1-ZNF687, or TAF15-ZNF-384, which are characteristic of acute myeloid leukemia (AML); CCND1-FSTL3, which is characteristic of chronic lymphocytic leukemia (CLL); BCL3-MYC, MYC-BTG1, BCL7A-MYC, BRWD3-ARHGAP20 or BTG1-MYC, which are characteristic of B-cell chronic lymphocytic leukemia (B-CLL); CITTA-BCL6, CLTC-ALK, IL21R-BCL6, PIM1-BCL6, TFCR-BCL6, IKZF1-BCL6 or SEC31A-ALK, which are characteristic of diffuse large B-cell lymphomas (DLBCL); FLIP1-PDGFRA, FLT3-ETV6, KIAA1509-PDGFRA, PDE4DIP-PDGFRB, NIN-PDGFRB, TP53BP1-PDGFRB, or TPM3-PDGFRB, which are characteristic of hyper eosinophilia / chronic eosinophilia; and IGH-MYC or LCP1-BCL6, which are characteristic of Burkitt's lymphoma. One of skill will understand that additional fusions, including those yet to be identified to date, can be used to guide treatment once their presence is associated with a therapeutic intervention.

The fusion genes and gene products can be detected using one or more techniques described herein. In some embodiments, the sequence of the gene or corresponding mRNA is determined, e.g., using Sanger sequencing, NGS, pyrosequencing, DNA microarrays, etc. Chromosomal abnormalities can be assessed using ISH, NGS or PCR techniques, among others. For example, a break apart probe can be used for ISH detection of ALK fusions such as *EML4-ALK*, *KIF5B-ALK* and/or *TFG-ALK*. As an alternate, PCR can be used to amplify the fusion product, wherein amplification or lack thereof indicates the presence or absence of the fusion, respectively. mRNA can be sequenced, e.g., using NGS to detect such fusions. See, e.g., **Table 9** or **Table 12** of WO2018175501 or **Tables 126-127** herein. In some embodiments, the fusion protein fusion is detected. Appropriate methods for protein analysis include without limitation mass spectroscopy, electrophoresis (e.g., 2D gel electrophoresis or SDS-PAGE) or antibody related techniques, including immunoassay, protein array or immunohistochemistry. The techniques can be combined. As a non-limiting example, indication of an ALK fusion by NGS can be confirmed by ISH or ALK expression using IHC, or vice versa.

Molecular Profiling Targets for Treatment Selection

The systems and methods described herein allow identification of one or more therapeutic regimes with projected therapeutic efficacy, based on the molecular profiling. Illustrative schemes for using molecular profiling to identify a treatment regime are provided throughout. Additional schemes are described in International Patent Publications WO/2007/137187 (Int'l Appl. No. PCT/US2007/069286), published November 29, 2007; WO/2010/045318 (Int'l Appl. No. PCT/US2009/060630), published April 22, 2010; WO/2010/093465 (Int'l Appl. No. PCT/US2010/000407), published August 19, 2010; WO/2012/170715 (Int'l Appl. No. PCT/US2012/041393), published December 13, 2012; WO/2014/089241 (Int'l Appl. No. PCT/US2013/073184), published June 12, 2014; WO/2011/056688 (Int'l Appl. No. PCT/US2010/054366), published May 12, 2011; WO/2012/092336 (Int'l Appl. No. PCT/US2011/067527), published July 5, 2012; WO/2015/116868 (Int'l Appl. No.

PCT/US2015/013618), published August 6, 2015; WO/2017/053915 (Int'l Appl. No. PCT/US2016/053614), published March 30, 2017; WO/2016/141169 (Int'l Appl. No. PCT/US2016/020657), published September 9, 2016; and WO2018175501 (Int'l Appl. No. PCT/US2018/023438), published September 27, 2018; each of which publications is incorporated by reference herein in its entirety.

The methods described herein comprise use of molecular profiling results to suggest associations with treatment benefit. In some embodiments, rules are used to provide the suggested chemotherapy treatments based on the molecular profiling test results. Rules can be constructed in a format such as "if biomarker positive then treatment option one, else treatment option two," or variations thereof. Treatment options comprise treatment with a single therapy (e.g., 5-FU) or treatment with a combination regimen (e.g., FOLFOX or FOLFIRI regimens for colorectal cancer). In some embodiments, more complex rules are constructed that involve the interaction of two or more biomarkers. Finally, a report can be generated that describes the association of the predicted benefit of a treatment and the biomarker and optionally a summary statement of the best evidence supporting the treatments selected. Ultimately, the treating physician will decide on the best course of treatment. The report may also list treatments with predicted lack of benefit. See, e.g., **Examples 4-5**.

The selection of a candidate treatment for an individual can be based on molecular profiling results from any one or more of the methods described.

In some embodiments, molecular profiling assays are performed to determine whether a copy number or copy number variation (CNV; also copy number alteration, CNA) of one or more genes is present in a sample as compared to a control, e.g., diploid level. The CNV of the gene or genes can be used to select a regimen that is predicted to be of benefit or lack of benefit for treating the patient. The methods can also include detection of mutations, indels, fusions, and the like in other genes and/or gene products, e.g., as described in **Example 1** herein, and International Patent Publications WO/2007/137187 (Int'l Appl. No. PCT/US2007/069286), published November 29, 2007; WO/2010/045318 (Int'l Appl. No. PCT/US2009/060630), published April 22, 2010; WO/2010/093465 (Int'l Appl. No. PCT/US2010/000407), published August 19, 2010; WO/2012/170715 (Int'l Appl. No. PCT/US2012/041393), published December 13, 2012; WO/2014/089241 (Int'l Appl. No. PCT/US2013/073184), published June 12, 2014; WO/2011/056688 (Int'l Appl. No. PCT/US2010/054366), published May 12, 2011; WO/2012/092336 (Int'l Appl. No. PCT/US2011/067527), published July 5, 2012; WO/2015/116868 (Int'l Appl. No. PCT/US2015/013618), published August 6, 2015; WO/2017/053915 (Int'l Appl. No. PCT/US2016/053614), published March 30, 2017; WO/2016/141169 (Int'l Appl. No. PCT/US2016/020657), published September 9, 2016; and WO2018175501 (Int'l Appl. No. PCT/US2018/023438), published September 27, 2018; each of which publications is incorporated by reference herein in its entirety.

The methods described herein are intended to prolong survival of a subject with cancer by providing personalized treatment. In some embodiments, the subject has been previously treated with one or more therapeutic agents to treat the cancer. The cancer may be refractory to one of these agents, e.g., by acquiring drug resistance mutations. In some embodiments, there is no known standard of care agent for the cancer or the cancer may be resistant to all known standard of care agent. Such standard of care agents may include "on label" agents, or those with an indication in a drug label. In some embodiments, the cancer is metastatic. In some embodiments, the subject has not previously been treated with one or more therapeutic agents identified by the method. Using molecular profiling, candidate treatments can be selected regardless of the stage, progression, anatomical location, or anatomical origin of the cancer cells.

The present disclosure provides methods and systems for analyzing diseased tissue using molecular profiling as previously described above. Because the methods rely on analysis of the characteristics of the tumor under analysis, the methods can be applied in for any tumor or any stage of disease, such an advanced stage of disease or a metastatic tumor of unknown origin. As described herein, a tumor or cancer sample is analyzed for one or more biomarkers in order to predict or identify a candidate therapeutic treatment.

The present methods can be used for selecting a treatment of primary or metastatic cancer.

The biomarker patterns and/or biomarker signature sets can comprise pluralities of biomarkers. In yet other embodiments, the biomarker patterns or signature sets can comprise at least 6, 7, 8, 9, or 10 biomarkers. In some embodiments, the biomarker signature sets or biomarker patterns can comprise at least 15, 20, 30, 40, 50, or 60 biomarkers. In some embodiments, the biomarker signature sets or biomarker patterns can comprise at least 70, 80, 90, 100, or 200, biomarkers. In some embodiments, the biomarker signature sets or biomarker patterns can comprise at least 100, 200, 300, 400, 500, 600, 700, or at least 800 biomarkers. In some embodiments, the biomarker signature sets or biomarker patterns can comprise at least 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 20,000, or at least 30,000 biomarkers. For example, the biomarkers may comprise whole exome sequencing and/or whole transcriptome sequencing and thus comprise all genes and gene products. Analysis of the one or more biomarkers can be by one or more methods, e.g., as described herein. See, e.g., **Example 1**.

As described herein, the molecular profiling of one or more targets can be used to determine or identify a therapeutic for an individual. For example, the presence, level or state of one or more biomarkers can be used to determine or identify a therapeutic for an individual. The one or more biomarkers, such as those disclosed herein, can be used to form a biomarker pattern or biomarker signature set, which is used to identify a therapeutic for an individual. In some embodiments, the therapeutic identified is one that the individual has not previously been treated with. For example, a reference biomarker pattern has been established for a particular therapeutic, such that individuals with the reference biomarker pattern will be responsive to that therapeutic. An individual with a

biomarker pattern that differs from the reference, for example the expression of a gene in the biomarker pattern is changed or different from that of the reference, would not be administered that therapeutic. In another example, an individual exhibiting a biomarker pattern that is the same or substantially the same as the reference is advised to be treated with that therapeutic. In some embodiments, the individual has not previously been treated with that therapeutic and thus a new therapeutic has been identified for the individual. The biomarker pattern may be based on a single biomarker (e.g., expression of HER2 suggests treatment with anti-HER2 therapy) or multiple biomarkers.

The genes used for molecular profiling, e.g., by IHC, ISH, sequencing (e.g., NGS), and/or PCR (e.g., qPCR), can be selected from those listed in **Example 1** herein, or as described in WO2018175501, e.g., in **Tables 5-10** therein. Assessing one or more biomarkers disclosed herein can be used for characterizing a cancer.

A cancer in a subject can be characterized by obtaining a biological sample from a subject and analyzing one or more biomarkers from the sample. For example, characterizing a cancer for a subject or individual can include identifying appropriate treatments or treatment efficacy for specific diseases, conditions, disease stages and condition stages, predictions and likelihood analysis of disease progression, particularly disease recurrence, metastatic spread or disease relapse. The products and processes described herein allow assessment of a subject on an individual basis, which can provide benefits of more efficient and economical decisions in treatment.

In an aspect, characterizing a cancer includes predicting whether a subject is likely to benefit from a treatment for the cancer. Biomarkers can be analyzed in the subject and compared to biomarker profiles of previous subjects that were known to benefit or not from a treatment. If the biomarker profile in a subject more closely aligns with that of previous subjects that were known to benefit from the treatment, the subject can be characterized, or predicted, as one who benefits from the treatment. Similarly, if the biomarker profile in the subject more closely aligns with that of previous subjects that did not benefit from the treatment, the subject can be characterized, or predicted as one who does not benefit from the treatment. The sample used for characterizing a cancer can be any useful sample, including without limitation those disclosed herein.

The methods can further include administering the selected treatment to the subject.

The treatment can be any beneficial treatment, e.g., small molecule drugs or biologics. Various immunotherapies, e.g., checkpoint inhibitor therapies such as ipilimumab, nivolumab, pembrolizumab, atezolizumab, avelumab, and durvalumab, are FDA approved and others are in clinical trials or developmental stages.

Genomic Prevalence Score (GPS)

The present disclosure provides systems, methods, and computer programs for determining attributes (phenotypes) of a biological sample, including without limitation a tissue of origin (TOO).

The present disclosure can determine such attribute for a biological sample in a number of different ways. For example, in some implementations, a first type of analysis can be performed on a biological sample to generate attributes of the DNA of the biological sample and then a trained model can be used to predict an attribute of the biological sample based on the assessment of the sample's DNA. In some embodiments, the model comprises a dynamic voting engine such as provided herein. By way of another example, a second type of analysis can be performed on a biological sample to generate attributes of the RNA of the biological sample and then a trained model can be used to predict the attributes for the biological sample based on the assessment of the sample's RNA. In some embodiments, the model may also comprise a dynamic voting engine such as provided herein. In other implementations, the first type of analysis and the second type of analysis can be performed in order to generate first biological data based on the biological sample's DNA and second biological data based on the biological sample's RNA and then use the trained model to predict an attribute for the biological sample based on the first biological data and the second biological data. In some embodiments, the model may also comprise a dynamic voting engine such as provided herein. In some implementations, the biological sample may be a cancer sample, e.g., tumor sample or bodily fluid comprising shed tumor cells or nucleic acids, and the attributed tissue of origin may be the origin where the tumor originated.

There are many technical advantages that are achieved through use of the systems, methods, and computer programs of the present disclosure. By way of example, the present disclosure provides a machine learning model in the form of a dynamic voting engine that can more accurately classify data a biological sample relative to conventional analyses. In some implementations, such accuracy increases can be achieved by training the machine learning model to dynamically vote a plurality of initial input tissue classifications and then select a target or final tissue classification indicative of an attribute (phenotype) tissue of origin for the biological sample such as the tissue of origin. The training processes employed to achieve such increases in accuracy are described in more detail herein.

The first step in treating cancer is diagnosis. Diagnosis may include physical exam (e.g., to detect an enlarged origin or suspicious skin lesion or discoloration), laboratory testing (e.g., urine or blood tests), medical imaging (e.g., computerized tomography (CT), bone scans, magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound and/or X-ray), and biopsy, which may be the preferred means to provide a definitive diagnosis. However, 3-9% of cases are misdiagnosed. *See, e.g.,* Peck, M. et al, Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. J Clin Pathol, 2018, 71: p. 995-1000, which reference is incorporated herein in its entirety. In addition, 5-10% of a Cancer of Occult/Unknown Primary (CUP). See www.mdanderson.org/cancer-types/cancer-of-unknown-primary.html; www.cancer.gov/types/unknown-primary/hp/unknown-primary-treatment-pdq#_1. Thus there is a need for improved methods of determining and/or verifying the tissue of origin (TOO) of a substantial

number of cancers. Automated verification of TOO may also identify laboratory errors in rare cases (e.g., switched samples).

The diagnosis of a malignancy is typically informed by clinical presentation and tumor tissue features including cell morphology, immunohistochemistry, cytogenetics, and molecular markers. Lack of reliable classification of a tumor poses a significant treatment dilemma for the oncologist leading to inappropriate and/or delayed treatment. Gene expression profiling has been used to try to identify the tumor type for CUP patients, but suffers from a number of inherent limitations. Specifically, tumor percentage, variation in expression, and the dynamic nature of RNA all contribute to suboptimal performance. For example, one commercial RNA-based assay has sensitivity of 83% in a test set of 187 tumors and confirmed results on only 78% of a separate 300 sample validation set. *See* Erlander MG, et al. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. J Mol Diagn. 2011 Sep;13(5):493-503; which reference is incorporated herein by reference in its entirety. Moreover, the diagnosis for any cancer may be mistaken in some cases.

Herein we provide systems and methods to predict attributes (phenotypes) of a biological sample, including primary location, histology, disease/cancer, and/or organ group. The granularity of the attribute can be chosen at a desired level such as described herein. We used molecular profiling (see, e.g., **Example 1**; **FIGs. 2B-C**) and machine learning to construct models and biosignatures for predicting such attributes. As a non-limiting example, such information can be used to identify the primary tumor site of a metastatic cancer of unknown primary (CUPS). In some embodiments, the predictions can be used to assist in planning treatment of cancer patients. In some embodiments, such information is used to verify the original diagnosis of a cancer at the same time molecular profiling is used to identify treatment options. If the information differs from the original diagnosis, additional inquiry may be performed (e.g., pathologist review) to verify the diagnosis and thus benefit patient treatment.

A general approach is as follows. First, we obtain a sample comprising cells from a cancer in a subject, e.g., a tumor sample or bodily fluid sample such as described herein. In some embodiments, the sample comprises metastatic cells. We perform molecular profiling assays on the sample to assess one or more biomarkers and thereby obtain a molecular profile, or biosignature, for the sample. See, e.g., **Example 1.** The sample biosignature can be input into a statistical model such as described herein. In some embodiments, this comprises comparing the sample biosignature to a number of biosignatures indicative of a plurality of attributes of interest. As a non-limiting example, one may compare the sample biosignature to each of a plurality of pre-determined biosignatures indicative of various attributes, e.g., various primary tumor origins. A probability or similar metric can be calculated that the sample biosignature corresponds to each of the pre-determined biosignatures. In some embodiments, the sample biosignature is used as an input into one or more machine learning models that are trained to take part in the overall prediction of the attribute/s of interest. Such models may calculate the probability or similarity metric described above. In some embodiments, one may

assign the attribute with the highest confidence, e.g., the highest probability. A threshold may be set such that the strength of assignment is determined.

The statistical models, e.g., machine learning models, are trained to the different attributes of interest. Herein, we demonstrate our approach using next-generation sequencing results for thousands of patient tumor samples. See, e.g., **Examples 2-3**. As a non-limiting example, consider that such data is used to identify a pre-determined biosignature for each of a plurality of tumor lineages, such as prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas, rectum, colon, breast, intrahepatic bile duct, cecum, gastroesophageal junction, frontal lobe, kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin. The biosignatures and models for each of the lineage predictors can comprise any number of features, here biomarkers, to achieve the desired level of performance. As will be understood by those of skill in the art, multiple features may provide a more robust prediction, but too many may lead to overfitting. Such parameters can be optimized in the training and testing phases of model development. As an non-limiting example, a biosignature for prostate may comprise DNA copy number for one or more of the genes FOXA1, PTEN, KLK2, GATA2, LCP1, ETV6, ERCC3, FANCA, MLLT3, MLH1, NCOA4, NCOA2, CCDC6, PTCH1, FOXO1, and IRF4.

FIGs. 3A and **3B** provide examples of the classification of individual tumor samples of known origin as test cases. **FIG. 3A** shows the prediction of a prostate cancer sample, correctly classified as of prostatic origin with high confidence as indicated by the tight shaded area. **FIG. 3B** shows the prediction of a tumor with a primary site as unknown but lineage as pancreatic. The predictor correctly identified the tumor as a pancreatic tumor although the site within the pancreas was indeterminate as indicated by the shaded region covering "Pancreas," "Head of pancreas," and "Tail of pancreas."

Provided herein is a method comprising obtaining a biological sample comprising cells from a cancer in a subject; performing an assay to assess one or more biomarkers in the sample to obtain a biosignature (also referred to as a molecular profile) for the sample; using the biosignature for the sample as an input into at least one statistical model, wherein the one or more statistical model may comprise at least one pre-determined biosignature; and (d) classifying or predicting an attribute of the sample based on the comparison, wherein the attribute comprises a primary origin, an organ type, a histology, and disease/cancer type, or any useful combination thereof. Similarly, provided herein is a method comprising: (a) obtaining a biological sample comprising cells from a subject; (b) performing an assay to assess one or more biomarkers in the sample to obtain a biosignature for the sample; (c) generating an input data based on the obtained sample and the one or more biomarkers; (d) providing the input data to a machine learning model that has been trained to predict an attribute of the sample using the input data, wherein the attribute is selected from the group consisting of a primary tumor

origin, cancer/disease type, organ group, histology, and any combination thereof (e) obtaining output data generated by the machine learning model based on the machine learning models processing of the input data; and (f) classifying the attribute of the sample based on the output data.

In some embodiments, the model is configured to perform pairwise analysis between the sample's biosignature and each of multiple different pre-determined (or trained) biosignatures, wherein each of the multiple different pre-determined biosignatures corresponds to a different attribute. See **Examples 2-3**, wherein performing pairwise analysis includes the machine learning model determining a level of similarity between the input data and biosignature for one or more of a plurality of disease types.

The desired attributes to be predicted may be determined at varying levels of specificity. For example, a tumor origin may be determined as a primary tumor location and a histology, which may be combined. For example, primary origin of a sample determined to be prostate and histology determined to be adenocarcinoma may combined as prostate adenocarcinoma. The models employed herein can be trained to such different specificities as desired. For example, a predictor model may be trained to recognize samples of prostatic origin, or may be trained to recognize prostate adenocarcinoma. In some embodiments, multiple models are trained at different attributes, e.g., organ or histology, and the results are combined to predict the desired level of attribute. As desired, the predictor models may be trained at a highly granular level, and the output can be identified in a less granular category of interest. See, e.g., more granular disease types and less granular organ groups in **Tables 2-116** below. In some embodiments, the predictor models are trained at such less granular level. In some embodiments, the predictor models are trained to different attributes (e.g., organ versus histology) which are then combined to provide the final predicted attribute.

In some embodiments, the systems and methods incorporate analysis of genomic DNA. Genomic abnormalities are a hallmark of cancer tissue. For example, 1p19q is indicative of certain cancers such as oligodendriogliomas. A single chromosome loss of 17 is the most frequent early occurrence in ovarian cancer, and 3p deletion in clear cell kidney and trisomy 7 and 17 in papillary renal cancer are established predictors. Chromosome 6 loss, 8 gain is a marker of eye cancers. Her2 amplification is observed in breast cancer. We hypothesized that the phenomena of genomic abnormalities such as gene copy number and mutational signatures may be predictive of many, if not all, types of cancers. DNA has certain advantages as an analyte biomarker as it can be robust to tumor percentage, metastasis, and sequencing depth, and can be analyzed efficiently using next-generation sequencing approaches. See, e.g., **Example 1**. In an aspect, we used the systems and methods provided herein to determine features of genomic DNA that are part of pre-determined biosignatures for 115 different granular disease/cancer types, including adrenal cortical carcinoma; anus squamous carcinoma; appendix adenocarcinoma, NOS; appendix mucinous adenocarcinoma; bile duct, NOS, cholangiocarcinoma; brain astrocytoma, anaplastic; brain astrocytoma, NOS; breast adenocarcinoma, NOS; breast carcinoma, NOS; breast infiltrating duct adenocarcinoma; breast infiltrating lobular

carcinoma, NOS; breast metaplastic carcinoma, NOS; cervix adenocarcinoma, NOS; cervix carcinoma, NOS; cervix squamous carcinoma; colon adenocarcinoma, NOS; colon carcinoma, NOS; colon mucinous adenocarcinoma; conjunctiva malignant melanoma, NOS; duodenum and ampulla adenocarcinoma, NOS; endometrial adenocarcinoma, NOS; endometrial carcinosarcoma; endometrial endometrioid adenocarcinoma; endometrial serous carcinoma; endometrium carcinoma, NOS; endometrium carcinoma, undifferentiated; endometrium clear cell carcinoma; esophagus adenocarcinoma, NOS; esophagus carcinoma, NOS; esophagus squamous carcinoma; extrahepatic cholangio, common bile, gallbladder adenocarcinoma, NOS; fallopian tube adenocarcinoma, NOS; fallopian tube carcinoma, NOS; fallopian tube carcinosarcoma, NOS; fallopian tube serous carcinoma; gastric adenocarcinoma; gastroesophageal junction adenocarcinoma, NOS; glioblastoma; glioma, NOS; gliosarcoma; head, face or neck, NOS squamous carcinoma; intrahepatic bile duct cholangiocarcinoma; kidney carcinoma, NOS; kidney clear cell carcinoma; kidney papillary renal cell carcinoma; kidney renal cell carcinoma, NOS; larynx, NOS squamous carcinoma; left colon adenocarcinoma, NOS; left colon mucinous adenocarcinoma; liver hepatocellular carcinoma, NOS; lung adenocarcinoma, NOS; lung adenosquamous carcinoma; lung carcinoma, NOS; lung mucinous adenocarcinoma; lung neuroendocrine carcinoma, NOS; lung non-small cell carcinoma; lung sarcomatoid carcinoma; lung small cell carcinoma, NOS; lung squamous carcinoma; meninges meningioma, NOS; nasopharynx, NOS squamous carcinoma; oligodendroglioma, anaplastic; oligodendroglioma, NOS; ovary adenocarcinoma, NOS; ovary carcinoma, NOS; ovary carcinosarcoma; ovary clear cell carcinoma; ovary endometrioid adenocarcinoma; ovary granulosa cell tumor, NOS; ovary high-grade serous carcinoma; ovary low-grade serous carcinoma; ovary mucinous adenocarcinoma; ovary serous carcinoma; pancreas adenocarcinoma, NOS; pancreas carcinoma, NOS; pancreas mucinous adenocarcinoma; pancreas neuroendocrine carcinoma, NOS; parotid gland carcinoma, NOS; peritoneum adenocarcinoma, NOS; peritoneum carcinoma, NOS; peritoneum serous carcinoma; pleural mesothelioma, NOS; prostate adenocarcinoma, NOS; rectosigmoid adenocarcinoma, NOS; rectum adenocarcinoma, NOS; rectum mucinous adenocarcinoma; retroperitoneum dedifferentiated liposarcoma; retroperitoneum leiomyosarcoma, NOS; right colon adenocarcinoma, NOS; right colon mucinous adenocarcinoma; salivary gland adenoid cystic carcinoma; skin melanoma; skin melanoma; skin merkel cell carcinoma; skin nodular melanoma; skin squamous carcinoma; skin trunk melanoma; small intestine adenocarcinoma; small intestine gastrointestinal stromal tumor, NOS; stomach gastrointestinal stromal tumor, NOS; stomach signet ring cell adenocarcinoma; thyroid carcinoma, anaplastic, NOS; thyroid carcinoma, NOS; thyroid papillary carcinoma of thyroid; tonsil, oropharynx, tongue squamous carcinoma; transverse colon adenocarcinoma, NOS; urothelial bladder adenocarcinoma, NOS; urothelial bladder carcinoma, NOS; urothelial bladder squamous carcinoma; urothelial carcinoma, NOS; uterine endometrial stromal sarcoma, NOS; uterus leiomyosarcoma, NOS; uterus sarcoma, NOS; uveal melanoma; vaginal squamous carcinoma; vulvar squamous carcinoma; and any combination thereof. Note that

NOS, or "Not Otherwise Specified," is a subcategory in systems of disease/disorder classification such as ICD-9, ICD-10, or DSM-IV, and is generally but not exclusively used where a more specific diagnosis was not made. The models for these disease types were trained using NGS data for a specified gene panel (see **Example 1**, **Tables 123-125**) obtained for tens of thousands of patient samples. Training of the models is further described in **Examples 2-3**.

Tables 2-116 list selections of features that contribute to the 115 disease type predictions, where each row in the table represents a feature ranked by Importance. In the tables, the column "GENE" is the identifier for the feature, which is a typically a gene ID; column "TECH" is the technology used to assess the biomarker, where "CNA" refers to copy number alteration as assessed by NGS, "NGS" is mutational analysis using next-generation sequencing, and "META" is a patient characteristic such as age at time of specimen collection ("Age") or gender ("Gender"); and column "IMP" is a normalized Importance score for the feature. A row in the tables where the GENE column is MSI and the TECH column is NGS refers to the feature microsatellite instability (MSI) as assessed by next-generation sequencing. The table headers indicate the more granular disease type (see above) and less granular organ group in the format "disease type – organ group". There are such 15 such organ groups indicated that each contain disease types originating in different organs or organ systems: bladder; skin; lung; head, face or neck (NOS); esophagus; female genital tract and peritoneum (FGTP); brain; colon; prostate; liver, gall bladder, ducts; breast; eye; stomach; kidney; and pancreas. A biological specimen can be grouped into one of the less granular 15 organ groups according to its more granular predicted disease type. As noted, the rows in the tables are sorted by importance. The higher the importance score the more important or relevant the feature is in making the disease type prediction. As indicated in the tables, in most cases we observed that gene copy numbers were driving the predictions.

Table 2: Adrenal Cortical Carcinoma – Adrenal Gland

GENE	TECH	IMP	CCNE1	CNA	0.492	LPP	CNA	0.357
HMGA2	CNA	1.000	c-KIT	NGS	0.486	ABL1	NGS	0.355
FOXL2	NGS	0.900	CDH11	CNA	0.480	LGR5	CNA	0.338
CTCF	CNA	0.886	TSC1	CNA	0.450	BTG1	CNA	0.338
WIF1	CNA	0.768	NR4A3	CNA	0.448	TPM3	CNA	0.335
DDIT3	CNA	0.698	CTNNA1	CNA	0.441	EP300	CNA	0.307
PTPN11	CNA	0.689	FGFR2	CNA	0.439	SRSF2	CNA	0.306
EWSR1	CNA	0.664	ATF1	CNA	0.438	KRAS	NGS	0.298
PPP2R1A	CNA	0.640	ATP1A1	CNA	0.428	RBM15	CNA	0.290
EBF1	CNA	0.637	FOXO1	CNA	0.401	ABL2	CNA	0.288
CDH1	CNA	0.633	ACSL6	CNA	0.394	VHL	NGS	0.284
CDK4	CNA	0.607	BRCA2	CNA	0.374	MYCL	CNA	0.279
Age	META	0.599	CHEK2	CNA	0.374	ITK	CNA	0.278
NUP93	CNA	0.507	SOX2	CNA	0.373	ZNF331	CNA	0.273
CRKL	CNA	0.499	FNBP1	CNA	0.361	TFPT	CNA	0.268

ARNT	CNA	0.267	BCL9	CNA	0.265	ELK4	CNA	0.263
ALDH2	CNA	0.265	MECOM	CNA	0.264	RB1	CNA	0.261

Table 3: Anus Squamous carcinoma – Colon

GENE	TECH	IMP						
LPP	CNA	1.000	CDKN2B	CNA	0.782	SRGAP3	CNA	0.652
FOXL2	NGS	0.956	Gender	META	0.781	NTRK2	CNA	0.646
CDKN2A	CNA	0.894	ARID1A	CNA	0.771	HMGN2P46	CNA	0.641
SOX2	CNA	0.872	BCL6	CNA	0.759	AFF3	CNA	0.636
CACNA1D	CNA	0.852	SDHD	CNA	0.746	IGF1R	CNA	0.631
CNBP	CNA	0.852	PAX3	CNA	0.745	MDS2	CNA	0.630
KLHL6	CNA	0.843	XPC	CNA	0.710	BARD1	CNA	0.624
TFRC	CNA	0.842	KDSR	CNA	0.707	EXT1	CNA	0.618
SPEN	CNA	0.805	TGFBR2	CNA	0.705	MECOM	CNA	0.617
TP53	NGS	0.804	WWTR1	CNA	0.701	TRIM27	CNA	0.615
Age	META	0.803	FLI1	CNA	0.697	KMT2A	CNA	0.614
VHL	CNA	0.797	PCSK7	CNA	0.693	GNAS	CNA	0.597
PPARG	CNA	0.794	BCL2	CNA	0.683	ATIC	CNA	0.594
RPN1	CNA	0.794	PAFAH1B2	CNA	0.674	MAX	CNA	0.569
ZBTB16	CNA	0.786	CBL	CNA	0.667	FHIT	CNA	0.563
FANCC	CNA	0.785	CREB3L2	CNA	0.664	SDHB	CNA	0.552
			CCNE1	CNA	0.654	PRDM1	CNA	0.550

Table 4: Appendix Adenocarcinoma NOS – Colon

GENE	TECH	IMP						
KRAS	NGS	1.000	CTCF	CNA	0.678	FANCC	CNA	0.570
FOXL2	NGS	0.948	SOX2	CNA	0.671	CHEK2	CNA	0.566
CDX2	CNA	0.916	HEY1	CNA	0.664	CCNE1	CNA	0.564
LHFPL6	CNA	0.901	NFIB	CNA	0.658	HOXA9	CNA	0.563
Age	META	0.873	ESR1	CNA	0.656	CBFB	CNA	0.557
FLT1	CNA	0.807	NUP214	CNA	0.645	BTG1	CNA	0.556
CDKN2A	CNA	0.781	LCP1	CNA	0.639	CACNA1D	CNA	0.555
SRSF2	CNA	0.772	SMAD4	CNA	0.635	FOXO3	CNA	0.554
BCL2	CNA	0.768	FGF14	CNA	0.617	PSIP1	CNA	0.554
Gender	META	0.744	IGF1R	CNA	0.615	RB1	CNA	0.554
SETBP1	CNA	0.728	TSC1	CNA	0.606	ERCC5	CNA	0.544
FLT3	CNA	0.728	MAP2K1	CNA	0.604	PTCH1	CNA	0.542
CRKL	CNA	0.722	WWTR1	CNA	0.599	CDKN1B	CNA	0.538
CDKN2B	CNA	0.698	FCRL4	CNA	0.597	BAP1	CNA	0.533
KDSR	CNA	0.688	CNBP	CNA	0.590	SS18	CNA	0.533
PDCD1LG2	CNA	0.687	CDH11	CNA	0.588	APC	NGS	0.533
			MLLT3	CNA	0.575	ARNT	CNA	0.533

5

Table 5: Appendix Mucinous adenocarcinoma – Colon

GENE	TECH	IMP						
			KRAS	NGS	1.000	GNAS	NGS	0.828

FOXL2	NGS	0.804	LHFPL6	CNA	0.472	ESR1	CNA	0.383
Age	META	0.682	NR4A3	CNA	0.471	EBF1	CNA	0.382
APC	NGS	0.657	GNA13	CNA	0.464	CDH1	CNA	0.382
CDX2	CNA	0.657	c-KIT	NGS	0.455	NF2	CNA	0.374
EPHA3	CNA	0.629	NSD1	CNA	0.449	SETBP1	CNA	0.372
PDCD1LG2	CNA	0.605	HERPUD1	CNA	0.442	WIF1	CNA	0.371
CDKN2A	CNA	0.603	Gender	META	0.439	HOXD13	CNA	0.370
CDKN2B	CNA	0.598	WWTR1	CNA	0.433	HOXA11	CNA	0.366
CDH11	CNA	0.597	RPN1	CNA	0.427	AFF4	CNA	0.365
HMGN2P46	CNA	0.514	TTL	CNA	0.412	TSC1	CNA	0.358
CACNA1D	CNA	0.506	FLT1	CNA	0.407	KLHL6	CNA	0.356
ERCC5	CNA	0.500	AFF3	CNA	0.396	VHL	CNA	0.352
TAL2	CNA	0.493	CD274	CNA	0.392	PBX1	CNA	0.350
MSI2	CNA	0.488	CREB3L2	CNA	0.391	KDSR	CNA	0.348
FANCG	CNA	0.481	NUP214	CNA	0.389	SPECC1	CNA	0.345
FNBP1	CNA	0.472	EXT1	CNA	0.385	SRSF2	CNA	0.342

Table 6: Bile duct NOS, cholangiocarcinoma – Liver, GallBladder, Ducts

GENE	**TECH**	**IMP**						
SPEN	CNA	1.000	SRGAP3	CNA	0.704	BTG1	CNA	0.618
FOXL2	NGS	0.944	CDKN2B	CNA	0.698	KDSR	CNA	0.611
C15orf65	CNA	0.923	MDS2	CNA	0.695	MAF	CNA	0.606
ARID1A	CNA	0.906	PBX1	CNA	0.681	MAML2	CNA	0.595
CAMTA1	CNA	0.884	EBF1	CNA	0.680	TSHR	CNA	0.585
FANCF	CNA	0.803	ERG	CNA	0.674	CDKN2A	CNA	0.575
Gender	META	0.802	VHL	NGS	0.669	ARHGAP26	NGS	0.570
Age	META	0.794	TP53	NGS	0.651	FLT3	CNA	0.562
CDK12	CNA	0.769	MTOR	CNA	0.650	NTRK2	CNA	0.559
CHIC2	CNA	0.761	FANCC	CNA	0.648	LHFPL6	CNA	0.546
FHIT	CNA	0.759	MCL1	CNA	0.646	CDH1	NGS	0.545
SDHB	CNA	0.753	VHL	CNA	0.643	HLF	CNA	0.544
PTPRC	NGS	0.742	LPP	CNA	0.638	BCL6	CNA	0.544
NOTCH2	CNA	0.734	FOXA1	CNA	0.634	MYD88	CNA	0.542
XPC	CNA	0.714	SUZ12	CNA	0.630	FSTL3	CNA	0.535
APC	NGS	0.706	PRDM1	CNA	0.629	PPARG	CNA	0.532
			WISP3	CNA	0.624	PDCD1LG2	CNA	0.532

Table 7: Brain Astrocytoma NOS – Brain

GENE	**TECH**	**IMP**						
IDH1	NGS	1.000	SOX2	CNA	0.722	ZNF217	CNA	0.587
Age	META	0.867	SPECC1	CNA	0.705	HIST1H3B	CNA	0.575
FOXL2	NGS	0.856	CREB3L2	CNA	0.651	PDGFRA	CNA	0.556
EGFR	CNA	0.769	NDRG1	CNA	0.647	HMGA2	CNA	0.552
FGFR2	CNA	0.755	CDK6	CNA	0.625	MSI2	CNA	0.548
MYC	CNA	0.722	ATRX	NGS	0.604	AKAP9	CNA	0.534
			KAT6B	CNA	0.598	OLIG2	CNA	0.533

Gender	META	0.528
TP53	NGS	0.514
DDX6	CNA	0.508
TRRAP	CNA	0.501
TET1	CNA	0.493
MCL1	CNA	0.480
ZBTB16	CNA	0.472
BTG1	CNA	0.458
NFKB2	CNA	0.451
CDKN2B	CNA	0.447
GID4	CNA	0.438
SRSF2	CNA	0.435
CBL	CNA	0.424
NUP93	CNA	0.424
CHIC2	CNA	0.414
SRGAP3	CNA	0.414
ECT2L	CNA	0.413
KRAS	NGS	0.410
CCDC6	CNA	0.409
ACSL6	CNA	0.405
NCOA2	CNA	0.390
STK11	CNA	0.387
PIK3CG	CNA	0.387
LPP	CNA	0.387
MECOM	CNA	0.383
CDX2	CNA	0.381
SPEN	CNA	0.378
TCL1A	CNA	0.376
RABEP1	CNA	0.375
PMS2	CNA	0.370

Table 8: Brain Astrocytoma anaplastic – Brain

GENE	TECH	IMP
Age	META	1.000
IDH1	NGS	0.864
FOXL2	NGS	0.847
HMGA2	CNA	0.709
SOX2	CNA	0.709
MYC	CNA	0.695
SPECC1	CNA	0.675
CREB3L2	CNA	0.672
MSI2	CNA	0.617
ZNF217	CNA	0.593
EXT1	CNA	0.582
TPM3	CNA	0.572
SETBP1	CNA	0.548
CACNA1D	CNA	0.536
NR4A3	CNA	0.524
Gender	META	0.523
MSI	NGS	0.519
NTRK2	CNA	0.499
SDHD	CNA	0.481
TET1	CNA	0.470
OLIG2	CNA	0.451
CLP1	CNA	0.445
VHL	NGS	0.432
CTCF	CNA	0.432
VTI1A	CNA	0.427
PMS2	CNA	0.423
CDK6	CNA	0.422
CBFB	CNA	0.420
NUP93	CNA	0.419
ELK4	CNA	0.416
FNBP1	CNA	0.409
TP53	NGS	0.409
PBX1	CNA	0.406
KRAS	NGS	0.405
MLLT11	CNA	0.403
FGFR2	CNA	0.401
EGFR	CNA	0.394
RUNX1T1	CNA	0.394
NFKBIA	CNA	0.391
c-KIT	NGS	0.382
FAM46C	CNA	0.380
BCL9	CNA	0.377
FGF10	CNA	0.376
CDKN2B	CNA	0.374
MLH1	CNA	0.374
CCDC6	CNA	0.373
PDE4DIP	CNA	0.372
H3F3A	CNA	0.370
MECOM	CNA	0.368
NUP214	CNA	0.366

Table 9: Breast Adenocarcinoma NOS – Breast

GENE	TECH	IMP
GATA3	CNA	1.000
Gender	META	0.906
Age	META	0.811
ELK4	CNA	0.773
FUS	CNA	0.739
CCND1	CNA	0.698
KRAS	NGS	0.682
FOXL2	NGS	0.646
PBX1	CNA	0.631
MCL1	CNA	0.625
APC	NGS	0.602
PAX8	CNA	0.592
GNAQ	NGS	0.588
EWSR1	CNA	0.579
BCL9	CNA	0.571
MYC	CNA	0.569
HIST1H4I	NGS	0.556
CDH1	NGS	0.556
LHFPL6	CNA	0.555
VHL	NGS	0.551
PRCC	CNA	0.550
CREBBP	CNA	0.545
PDGFRA	NGS	0.539
FLI1	CNA	0.536
CDX2	CNA	0.535
SDHD	CNA	0.535
FHIT	CNA	0.533
CACNA1D	CNA	0.528
MECOM	CNA	0.526
YWHAE	CNA	0.522
AKT3	CNA	0.522
CDKN2A	CNA	0.521
SDHC	CNA	0.518
RPL22	CNA	0.513
FOXO1	CNA	0.512
TRIM27	CNA	0.511
TNFRSF17	CNA	0.511
STAT3	CNA	0.506

RMI2	CNA	0.506
PAFAH1B2	CNA	0.504
ZNF217	CNA	0.499
CDKN2B	CNA	0.498

TPM3	CNA	0.498
MUC1	CNA	0.498
EXT1	CNA	0.498
CCND2	CNA	0.496

FH	CNA	0.494
HMGA2	CNA	0.493
RUNX1T1	CNA	0.492
POU2AF1	CNA	0.490

Table 10: Breast Carcinoma NOS – Breast

GENE	TECH	IMP
GATA3	CNA	1.000
Age	META	0.974
ELK4	CNA	0.922
Gender	META	0.908
FOXL2	NGS	0.898
MCL1	CNA	0.886
MYC	CNA	0.865
CCND1	CNA	0.845
RMI2	CNA	0.807
LHFPL6	CNA	0.790
PBX1	CNA	0.789
USP6	CNA	0.776
FOXA1	CNA	0.760
MUC1	CNA	0.757
MLLT11	CNA	0.752
COX6C	CNA	0.738

BCL9	CNA	0.734
TNFRSF17	CNA	0.734
CREBBP	CNA	0.725
CACNA1D	CNA	0.723
EXT1	CNA	0.721
MECOM	CNA	0.700
PAX8	CNA	0.699
FUS	CNA	0.698
FLI1	CNA	0.694
HMGA2	CNA	0.689
ARID1A	CNA	0.689
TP53	NGS	0.685
PRCC	CNA	0.684
STAT3	CNA	0.681
FOXO1	CNA	0.677
CDH11	CNA	0.672
ZNF217	CNA	0.672

SPECC1	CNA	0.671
H3F3A	CNA	0.670
SDHC	CNA	0.665
SETBP1	CNA	0.659
YWHAE	CNA	0.658
TGFBR2	CNA	0.656
CDKN2A	CNA	0.656
PDE4DIP	CNA	0.651
FHIT	CNA	0.650
GAS7	CNA	0.648
ARNT	CNA	0.647
CDKN2B	CNA	0.642
CDH1	CNA	0.639
MAML2	CNA	0.634
GID4	CNA	0.632
TPM3	CNA	0.630
RPN1	CNA	0.626

Table 11: Breast Infiltrating Duct Adenocarcinoma – Breast

GENE	TECH	IMP
GATA3	CNA	1.000
Age	META	0.841
FOXL2	NGS	0.833
MYC	CNA	0.797
EXT1	CNA	0.796
Gender	META	0.786
PBX1	CNA	0.778
MCL1	CNA	0.727
ELK4	CNA	0.692
COX6C	CNA	0.683
CDH1	NGS	0.671
CCND1	CNA	0.667
FUS	CNA	0.665
RUNX1T1	CNA	0.647
BCL9	CNA	0.640
LHFPL6	CNA	0.624

TNFRSF17	CNA	0.617
USP6	CNA	0.604
RAD21	CNA	0.604
STAT5B	CNA	0.603
FLI1	CNA	0.595
SNX29	CNA	0.592
FH	CNA	0.590
PIK3CA	NGS	0.584
SLC34A2	CNA	0.580
CACNA1D	CNA	0.578
PAX8	CNA	0.578
CREBBP	CNA	0.576
CDKN2A	CNA	0.574
PCM1	CNA	0.571
SPECC1	CNA	0.571
U2AF1	CNA	0.568
TP53	NGS	0.564

MSI2	CNA	0.563
GID4	CNA	0.562
ZNF217	CNA	0.561
MAML2	CNA	0.556
TPM3	CNA	0.554
BRCA1	CNA	0.554
PAFAH1B2	CNA	0.553
IKBKE	CNA	0.553
MUC1	CNA	0.552
RMI2	CNA	0.547
FOXO1	CNA	0.547
CDKN2B	CNA	0.547
HMGA2	CNA	0.546
MDM4	CNA	0.546
ESR1	NGS	0.545
HOXD13	CNA	0.544
FANCC	CNA	0.538

5

Table 12: Breast Infiltrating Lobular Carcinoma NOS – Breast

GENE	TECH	IMP
CDH1	NGS	1.000
CDH1	CNA	0.684
CTCF	CNA	0.649
CDH11	CNA	0.640
ELK4	CNA	0.600
FOXL2	NGS	0.590
CAMTA1	CNA	0.563
Gender	META	0.535
IKBKE	CNA	0.478
FLI1	CNA	0.477
CBFB	CNA	0.474
PBX1	CNA	0.450
CDC73	CNA	0.438
GATA3	CNA	0.394
BCL9	CNA	0.387
CREBBP	CNA	0.385
FANCA	CNA	0.377
YWHAE	CNA	0.361
Age	META	0.344
BCL2	CNA	0.343
TP53	NGS	0.342
MECOM	CNA	0.339
FH	CNA	0.332
USP6	CNA	0.331
PCSK7	CNA	0.330
AKT3	CNA	0.328
KCNJ5	CNA	0.323
CDKN2B	CNA	0.314
CBL	CNA	0.302
ETV5	CNA	0.302
MDM4	CNA	0.295
FUS	CNA	0.292
CDX2	CNA	0.285
NUP93	CNA	0.282
ARNT	CNA	0.282
VHL	NGS	0.281
ABL2	CNA	0.280
TRIM33	NGS	0.273
PAX8	CNA	0.271
KDM5C	NGS	0.270
PAFAH1B2	CNA	0.270
HOXD11	CNA	0.269
APC	NGS	0.269
AURKB	CNA	0.269
TFRC	CNA	0.267
KRAS	NGS	0.266
CDKN2A	CNA	0.265
KLHL6	CNA	0.262
CTNNA1	CNA	0.261
DDR2	CNA	0.261

Table 13: Breast Metaplastic Carcinoma NOS – Breast

GENE	TECH	IMP
Gender	META	1.000
MAF	CNA	0.966
FOXL2	NGS	0.919
NUTM2B	CNA	0.916
EP300	CNA	0.906
CDKN2A	CNA	0.880
Age	META	0.873
ERBB3	CNA	0.855
DDIT3	CNA	0.849
PIK3CA	NGS	0.816
MSI2	CNA	0.815
PRRX1	CNA	0.791
NTRK2	CNA	0.755
CDKN2B	CNA	0.748
HMGA2	CNA	0.744
STAT5B	CNA	0.735
EWSR1	CNA	0.733
ERCC3	CNA	0.728
TRIM27	CNA	0.723
PRKDC	CNA	0.718
MYC	CNA	0.714
COX6C	CNA	0.714
HEY1	CNA	0.701
PDCD1LG2	CNA	0.697
FGF10	CNA	0.695
ITK	CNA	0.688
NR4A3	CNA	0.687
NF2	CNA	0.684
PIK3R1	NGS	0.661
SMARCB1	CNA	0.632
EXT1	CNA	0.629
CCNE1	CNA	0.629
CLTCL1	CNA	0.626
ARHGAP26	CNA	0.595
TP53	NGS	0.592
PLAG1	CNA	0.592
ATF1	CNA	0.562
CDK4	CNA	0.561
WISP3	CNA	0.560
CDH11	CNA	0.558
FANCC	CNA	0.557
RNF43	CNA	0.555
CHEK2	CNA	0.555
HMGN2P46	CNA	0.551
ERG	CNA	0.546
CHCHD7	CNA	0.543
PMS2	CNA	0.538
TAL2	CNA	0.537
SDHD	CNA	0.531
NFIB	CNA	0.531

Table 14: Cervix Adenocarcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000
FOXL2	NGS	0.815
TP53	NGS	0.718
Gender	META	0.704
GNAS	CNA	0.695
FLI1	CNA	0.692
KRAS	NGS	0.641
SDC4	CNA	0.626
CDK6	CNA	0.601
LPP	CNA	0.599
MECOM	CNA	0.596

LHFPL6	CNA	0.593
KLHL6	CNA	0.570
KDSR	CNA	0.566
CREB3L2	CNA	0.548
RAC1	CNA	0.548
PBX1	CNA	0.538
ETV5	CNA	0.534
MLLT11	CNA	0.531
BCL6	CNA	0.526
MUC1	CNA	0.526
PLAG1	CNA	0.522
TPM3	CNA	0.521
ZNF217	CNA	0.517

MYC	CNA	0.511
HEY1	CNA	0.504
MLF1	CNA	0.498
PDGFRA	CNA	0.496
PAX8	CNA	0.493
CTNNA1	CNA	0.488
CDKN2A	CNA	0.483
TFRC	CNA	0.481
WWTR1	CNA	0.477
SETBP1	CNA	0.471
SDHAF2	CNA	0.471
EXT1	CNA	0.470
APC	NGS	0.466

CDH1	CNA	0.463
TRRAP	CNA	0.452
CBL	CNA	0.451
UBR5	CNA	0.451
PIK3CA	NGS	0.446
EWSR1	CNA	0.444
IKZF1	CNA	0.441
ARID1A	CNA	0.430
ASXL1	CNA	0.427
CCNE1	CNA	0.427
KIAA1549	CNA	0.425
PRRX1	CNA	0.425
FGFR2	CNA	0.425

Table 15: Cervix Carcinoma NOS – FGTP

GENE	TECH	IMP
MECOM	CNA	1.000
FOXL2	NGS	0.973
Gender	META	0.973
Age	META	0.972
RPN1	CNA	0.950
U2AF1	CNA	0.900
SOX2	CNA	0.856
BCL6	CNA	0.832
EXT1	CNA	0.819
HMGN2P46	CNA	0.802
ATIC	CNA	0.761
RAC1	CNA	0.750
KLHL6	CNA	0.748
ECT2L	CNA	0.747
LPP	CNA	0.741
USP6	CNA	0.740

WWTR1	CNA	0.714
CCNE1	CNA	0.692
SRSF2	CNA	0.683
PDGFRA	CNA	0.673
SEPT5	CNA	0.671
BTG1	CNA	0.668
CDK12	CNA	0.654
CDKN2B	CNA	0.647
RAD50	CNA	0.624
RNF213	NGS	0.615
TP53	NGS	0.600
DAXX	CNA	0.598
MLF1	CNA	0.596
BCL2	CNA	0.585
ETV5	CNA	0.585
ARFRP1	CNA	0.579
GMPS	CNA	0.569

NDRG1	CNA	0.568
YWHAE	CNA	0.567
ZNF217	CNA	0.558
FOXL2	CNA	0.555
EGFR	CNA	0.549
ACSL3	NGS	0.546
ERCC3	CNA	0.541
IKZF1	CNA	0.539
SDHC	CNA	0.536
SDC4	CNA	0.535
CREB3L2	CNA	0.525
TFRC	CNA	0.522
CACNA1D	CNA	0.519
CCND2	CNA	0.517
MUC1	CNA	0.510
BCL9	CNA	0.508
MYCL	CNA	0.505

Table 16: Cervix Squamous Carcinoma – FGTP

GENE	TECH	IMP
Age	META	1.000
TP53	NGS	0.863
CNBP	CNA	0.851
TFRC	CNA	0.838
FOXL2	NGS	0.828
RPN1	CNA	0.794
LPP	CNA	0.758
BCL6	CNA	0.751
KLHL6	CNA	0.740

WWTR1	CNA	0.739
ARID1A	CNA	0.736
Gender	META	0.724
SOX2	CNA	0.722
CREB3L2	CNA	0.699
CDKN2B	CNA	0.663
CDKN2A	CNA	0.614
SPEN	CNA	0.600
MECOM	CNA	0.595
ETV5	CNA	0.578

MAX	CNA	0.553
PAX3	CNA	0.548
CACNA1D	CNA	0.539
FOXP1	CNA	0.527
ERBB3	CNA	0.526
PMS2	CNA	0.513
MDS2	CNA	0.507
ATIC	CNA	0.502
RUNX1	CNA	0.500
SYK	CNA	0.498

SETBP1	CNA	0.495
IGF1R	CNA	0.494
ERBB4	CNA	0.478
KDSR	CNA	0.473
ZNF384	CNA	0.470
BCL2	CNA	0.467
FGF10	CNA	0.464

SLC34A2	CNA	0.464
SFPQ	CNA	0.463
EPHB1	CNA	0.454
NFKBIA	CNA	0.453
TRIM27	CNA	0.450
MITF	CNA	0.450
ERG	CNA	0.449

KIAA1549	CNA	0.447
GSK3B	CNA	0.444
NSD2	CNA	0.441
SPECC1	CNA	0.437
EXT1	CNA	0.430
LHFPL6	CNA	0.426
BCL11A	CNA	0.421

Table 17: Colon Adenocarcinoma NOS – Colon

GENE	TECH	IMP
CDX2	CNA	1.000
APC	NGS	0.912
FOXL2	NGS	0.801
KRAS	NGS	0.781
SETBP1	CNA	0.764
ASXL1	CNA	0.715
LHFPL6	CNA	0.713
FLT3	CNA	0.707
BCL2	CNA	0.704
FOXO1	CNA	0.703
SDC4	CNA	0.693
KDSR	CNA	0.691
ZNF217	CNA	0.686
Age	META	0.660
FLT1	CNA	0.639
EBF1	CNA	0.627

GNAS	CNA	0.620
Gender	META	0.615
ERG	CNA	0.600
CDKN2B	CNA	0.592
ERCC5	CNA	0.587
NSD2	CNA	0.580
IRS2	CNA	0.577
SMAD4	CNA	0.574
TOP1	CNA	0.574
EPHA5	CNA	0.564
HOXA9	CNA	0.552
CDH1	CNA	0.551
CDKN2A	CNA	0.548
CBFB	CNA	0.537
ZNF521	CNA	0.536
CDK8	CNA	0.533
USP6	CNA	0.529

FGFR2	CNA	0.512
WWTR1	CNA	0.512
RAC1	CNA	0.511
TP53	NGS	0.511
MYC	CNA	0.509
JAK1	CNA	0.508
SPEN	CNA	0.508
SPECC1	CNA	0.505
TP53	CNA	0.505
MSI2	CNA	0.499
EWSR1	CNA	0.497
CCNE1	CNA	0.496
ARID1A	CNA	0.494
CDK6	CNA	0.491
MAML2	CNA	0.490
RB1	CNA	0.489
U2AF1	CNA	0.485

Table 18: Colon Carcinoma NOS – Colon

GENE	TECH	IMP
APC	NGS	1.000
SDC4	CNA	0.773
VHL	NGS	0.715
CDH1	CNA	0.683
GNAS	CNA	0.676
IDH1	NGS	0.676
HMGN2P46	CNA	0.647
Gender	META	0.634
CDX2	CNA	0.616
c-KIT	NGS	0.601
Age	META	0.574
LHFPL6	CNA	0.554
CDH1	NGS	0.553
ASXL1	CNA	0.522
SMAD4	CNA	0.520

ZNF217	CNA	0.507
SETBP1	CNA	0.496
FOXL2	NGS	0.487
ARID1A	NGS	0.482
FANCF	CNA	0.480
CTCF	CNA	0.478
TOP1	CNA	0.475
KRAS	NGS	0.472
TP53	NGS	0.465
U2AF1	CNA	0.463
MYC	CNA	0.451
CDKN2C	CNA	0.438
AURKA	CNA	0.437
HOXA9	CNA	0.435
KLHL6	CNA	0.434
BCL9	CNA	0.431

PML	CNA	0.430
BCL2L11	CNA	0.428
CDK12	CNA	0.427
CYP2D6	CNA	0.424
TTL	CNA	0.423
KDM5C	NGS	0.422
BCL6	CNA	0.421
CASP8	CNA	0.416
ACKR3	NGS	0.415
KIAA1549	CNA	0.414
RPL22	CNA	0.408
FLT3	CNA	0.408
TPM3	CNA	0.407
STAT3	CNA	0.404
FOXO1	CNA	0.393
FNBP1	CNA	0.392

PTEN	NGS	0.390	PTCH1	CNA	0.383	MECOM	CNA	0.381

Table 19: Colon Mucinous Adenocarcinoma – Colon

GENE	TECH	IMP						
KRAS	NGS	1.000	TFRC	CNA	0.533	STAT3	CNA	0.474
APC	NGS	0.778	SRSF2	CNA	0.527	EPHA5	CNA	0.454
RPN1	CNA	0.745	ALDH2	CNA	0.513	SLC34A2	CNA	0.450
FOXL2	NGS	0.727	SDHAF2	CNA	0.511	HEY1	CNA	0.449
Age	META	0.686	PTEN	CNA	0.504	MSI2	CNA	0.449
CDX2	CNA	0.668	TSC1	CNA	0.501	CAMTA1	CNA	0.448
NUP214	CNA	0.638	SMAD4	CNA	0.500	FGF14	CNA	0.442
CDKN2B	CNA	0.632	WWTR1	CNA	0.492	MAX	CNA	0.441
LHFPL6	CNA	0.620	IDH1	NGS	0.492	TPM4	CNA	0.441
SETBP1	CNA	0.619	KDSR	CNA	0.491	BCL2	CNA	0.426
Gender	META	0.608	VHL	NGS	0.485	LPP	CNA	0.423
TP53	NGS	0.571	NFIB	CNA	0.485	KLF4	CNA	0.420
FGFR2	CNA	0.568	MAF	CNA	0.481	BTG1	CNA	0.420
RUNX1T1	CNA	0.558	BCL6	CNA	0.481	CDH11	CNA	0.417
PTEN	NGS	0.554	FLT3	CNA	0.479	FANCG	CNA	0.409
CDKN2A	CNA	0.553	PDCD1LG2	CNA	0.478	H3F3B	CNA	0.405
			GID4	CNA	0.475	PRKDC	CNA	0.402

Table 20: Conjunctiva Malignant melanoma NOS – Skin

GENE	TECH	IMP						
IRF4	CNA	1.000	Age	META	0.465	GNAQ	NGS	0.301
ACSL6	NGS	0.847	VHL	NGS	0.465	CCND3	CNA	0.300
FLI1	CNA	0.837	POU2AF1	CNA	0.463	LPP	CNA	0.283
WWTR1	CNA	0.810	DAXX	CNA	0.454	KRAS	NGS	0.282
TRIM27	CNA	0.763	NRAS	NGS	0.436	PDGFRA	CNA	0.279
RPN1	CNA	0.762	PMS2	CNA	0.421	SOX2	CNA	0.277
CDH1	NGS	0.738	KLHL6	CNA	0.411	EPHB1	CNA	0.275
FOXL2	NGS	0.738	ZBTB16	CNA	0.378	AFF3	CNA	0.275
TP53	NGS	0.602	APC	NGS	0.370	ESR1	CNA	0.274
KCNJ5	CNA	0.593	EBF1	CNA	0.367	CTNNB1	NGS	0.273
SOX10	CNA	0.575	PRKAR1A	CNA	0.351	KIT	CNA	0.257
DEK	CNA	0.557	ETV1	CNA	0.339	CLP1	CNA	0.251
MLF1	CNA	0.519	SRSF3	CNA	0.338	GATA2	CNA	0.246
EP300	CNA	0.491	TRIM26	CNA	0.328	SDHD	CNA	0.245
CNBP	CNA	0.484	WT1	CNA	0.328	CBL	CNA	0.244
Gender	META	0.482	BCL6	CNA	0.321	WIF1	CNA	0.233
			BRAF	NGS	0.306	KDSR	CNA	0.230

5

Table 21: Duodenum and Ampulla Adenocarcinoma NOS – Colon

GENE	TECH	IMP						
KRAS	NGS	1.000	FOXL2	NGS	0.926	CDX2	CNA	0.870
			SETBP1	CNA	0.902	Age	META	0.842

FLT3	CNA	0.837
KDSR	CNA	0.829
JAZF1	CNA	0.807
FLT1	CNA	0.804
USP6	CNA	0.769
APC	NGS	0.768
CDKN2A	CNA	0.741
LHFPL6	CNA	0.741
BCL2	CNA	0.725
SPECC1	CNA	0.704
Gender	META	0.695
GID4	CNA	0.691
TCF7L2	CNA	0.685
CDKN2B	CNA	0.681
FOXO1	CNA	0.665

CBFB	CNA	0.657
PMS2	CNA	0.648
U2AF1	CNA	0.631
CACNA1D	CNA	0.623
CDK8	CNA	0.620
CRTC3	CNA	0.620
LCP1	CNA	0.604
RB1	CNA	0.604
CDH1	CNA	0.603
ERCC5	CNA	0.602
TP53	NGS	0.600
SDHB	CNA	0.598
ETV6	CNA	0.584
CDH1	NGS	0.568
FGF6	CNA	0.565

BCL6	CNA	0.564
EXT1	CNA	0.559
PRRX1	CNA	0.557
PTPN11	CNA	0.557
CALR	CNA	0.556
VHL	NGS	0.552
CTCF	CNA	0.551
CRKL	CNA	0.548
GNAS	CNA	0.547
CHEK2	CNA	0.545
HOXA9	CNA	0.543
SDC4	CNA	0.543
ARID1A	CNA	0.542
FHIT	CNA	0.537
NF2	CNA	0.537

Table 22: Endometrial Endometroid Adenocarcinoma – FGTP

GENE	TECH	IMP
PTEN	NGS	1.000
ESR1	CNA	0.807
Gender	META	0.759
CDH1	NGS	0.696
Age	META	0.683
FOXL2	NGS	0.641
PIK3CA	NGS	0.600
APC	NGS	0.589
ARID1A	NGS	0.586
GATA2	CNA	0.575
CDX2	CNA	0.562
CBFB	CNA	0.558
CTNNB1	NGS	0.551
ZNF217	CNA	0.529
FNBP1	CNA	0.528
FANCF	CNA	0.526

IKZF1	CNA	0.520
MUC1	CNA	0.516
CDKN2A	CNA	0.513
FGFR2	CNA	0.513
NUP214	CNA	0.513
RAC1	CNA	0.512
HOXA13	CNA	0.511
TP53	NGS	0.509
PBX1	CNA	0.503
GNAS	CNA	0.503
MLLT11	CNA	0.502
CRKL	CNA	0.495
MECOM	CNA	0.493
AFF3	CNA	0.493
HMGN2P46	CNA	0.491
ELK4	CNA	0.491
U2AF1	CNA	0.488

PAX8	CNA	0.488
HMGN2P46	NGS	0.485
CCDC6	CNA	0.481
FGFR1	CNA	0.479
CDKN2B	CNA	0.472
FHIT	CNA	0.472
SOX2	CNA	0.462
MYC	CNA	0.457
SETBP1	CNA	0.456
EWSR1	CNA	0.454
LHFPL6	CNA	0.452
PIK3R1	NGS	0.451
PRRX1	CNA	0.444
CDH11	CNA	0.444
STAT3	CNA	0.439
MDM4	CNA	0.434
BCL9	CNA	0.434

Table 23: Endometrial Adenocarcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000
PTEN	NGS	0.967
Gender	META	0.852
MECOM	CNA	0.801
APC	NGS	0.779
PAX8	CNA	0.742
PIK3CA	NGS	0.737

KAT6B	CNA	0.707
CDH1	NGS	0.700
MLLT11	CNA	0.684
ESR1	CNA	0.664
CDH11	CNA	0.648
CDX2	CNA	0.647
FGFR2	CNA	0.646
HMGN2P46	CNA	0.627

ELK4	CNA	0.619
MUC1	CNA	0.602
CDH1	CNA	0.597
TP53	NGS	0.594
NR4A3	CNA	0.593
BCL9	CNA	0.589
LHFPL6	CNA	0.587
CDKN2B	CNA	0.583

CDKN2A	CNA	0.580	GNAQ	NGS	0.549	KLHL6	CNA	0.523
ARID1A	NGS	0.580	MAP2K1	CNA	0.548	CCDC6	CNA	0.523
KRAS	NGS	0.575	ETV5	CNA	0.547	MAF	CNA	0.521
CCNE1	CNA	0.571	CBFB	CNA	0.546	SETBP1	CNA	0.520
NUTM1	CNA	0.566	IKZF1	CNA	0.536	EXT1	CNA	0.519
GATA3	CNA	0.563	ARID1A	CNA	0.533	CDK6	CNA	0.517
FOXL2	NGS	0.562	EBF1	CNA	0.530	HOOK3	CNA	0.517
CTCF	CNA	0.561	RAC1	CNA	0.527	ERBB3	CNA	0.514
PRRX1	CNA	0.556	NUP214	CNA	0.526	VHL	CNA	0.505

Table 24: Endometrial Carcinosarcoma – FGTP

GENE	**TECH**	**IMP**						
CCNE1	CNA	1.000	FGFR1	CNA	0.687	IKZF1	CNA	0.609
FOXL2	NGS	0.961	XPA	CNA	0.682	NCOA2	CNA	0.607
Age	META	0.906	MAF	CNA	0.672	FSTL3	CNA	0.606
Gender	META	0.819	BCL9	CNA	0.672	NTRK2	CNA	0.603
MAP2K2	CNA	0.814	PRRX1	CNA	0.654	HOXD13	CNA	0.596
ASXL1	CNA	0.799	FNBP1	CNA	0.654	FANCF	CNA	0.595
HMGN2P46	CNA	0.792	SYK	CNA	0.647	TAL2	CNA	0.589
MLLT11	CNA	0.785	CBFB	CNA	0.646	MECOM	CNA	0.588
KLF4	CNA	0.777	PIK3CA	NGS	0.641	DDR2	CNA	0.588
PTEN	NGS	0.742	ALK	CNA	0.633	PRKDC	CNA	0.581
AFF3	CNA	0.734	TP53	NGS	0.631	FANCC	CNA	0.571
WDCP	CNA	0.723	TRIM27	CNA	0.626	CDKN2B	CNA	0.570
NR4A3	CNA	0.721	ETV6	CNA	0.623	EWSR1	CNA	0.569
RPN1	CNA	0.707	RAC1	CNA	0.622	BTG1	CNA	0.566
WISP3	CNA	0.705	CDKN2A	CNA	0.621	GATA2	CNA	0.563
CDH1	CNA	0.694	EP300	CNA	0.616	GNAQ	CNA	0.561
			ETV1	CNA	0.611	FOXA1	CNA	0.554

Table 25: Endometrial Serous Carcinoma – FGTP

GENE	**TECH**	**IMP**						
CCNE1	CNA	1.000	STAT3	CNA	0.702	KAT6B	CNA	0.633
Age	META	0.984	CBFB	CNA	0.696	ESR1	CNA	0.633
MECOM	CNA	0.959	RAC1	CNA	0.695	KLF4	CNA	0.632
TP53	NGS	0.955	CDKN2A	CNA	0.685	CREBBP	CNA	0.632
FOXL2	NGS	0.910	CREB3L2	CNA	0.683	FGFR2	CNA	0.628
PAX8	CNA	0.908	CDK6	CNA	0.674	PIK3CA	NGS	0.628
NUTM1	CNA	0.865	FSTL3	CNA	0.666	MAP2K1	CNA	0.627
Gender	META	0.854	BCL6	CNA	0.665	IKZF1	CNA	0.614
KLHL6	CNA	0.826	MAP2K2	CNA	0.663	NR4A3	CNA	0.611
CDH1	CNA	0.776	FANCF	CNA	0.661	LPP	CNA	0.611
HMGN2P46	CNA	0.765	C15orf65	CNA	0.653	CDH11	CNA	0.607
MAF	CNA	0.716	GATA2	CNA	0.648	ETV1	CNA	0.604
ETV5	CNA	0.705	SS18	CNA	0.634	TAL2	CNA	0.600
			AFF3	CNA	0.634	STK11	CNA	0.590

TPM4	CNA	0.590	INHBA	CNA	0.582	LHFPL6	CNA	0.578
NUP214	CNA	0.585	CTCF	CNA	0.581	ALK	CNA	0.578
MLLT11	CNA	0.584	GID4	CNA	0.581	CALR	CNA	0.573

Table 26: Endometrium Carcinoma NOS – FGTP

GENE	TECH	IMP	KLF4	CNA	0.601	CBFB	CNA	0.526
PTEN	NGS	1.000	RAC1	CNA	0.592	CDK6	CNA	0.524
FOXL2	NGS	0.896	CDH1	CNA	0.590	ARID1A	NGS	0.524
Age	META	0.804	IKZF1	CNA	0.578	BCL9	CNA	0.523
JAZF1	CNA	0.797	SDHC	CNA	0.573	NUP214	CNA	0.517
Gender	META	0.766	CDKN2A	CNA	0.570	FANCF	CNA	0.510
C15orf65	CNA	0.725	ELK4	CNA	0.564	NTRK2	CNA	0.508
PIK3CA	NGS	0.724	PIK3R1	NGS	0.560	EP300	CNA	0.504
LHFPL6	CNA	0.710	MAP2K1	CNA	0.559	VHL	CNA	0.500
FGFR2	CNA	0.665	PPARG	CNA	0.557	GID4	CNA	0.499
TET1	CNA	0.654	FLT3	CNA	0.553	ETV1	CNA	0.499
TP53	NGS	0.651	PAX8	CNA	0.552	GNAS	CNA	0.499
MLLT11	CNA	0.650	BMPR1A	CNA	0.545	EWSR1	CNA	0.498
FNBP1	CNA	0.647	FLI1	CNA	0.542	NR4A3	CNA	0.497
GNAQ	CNA	0.635	CCNE1	CNA	0.534	CTNNA1	CNA	0.495
EGFR	CNA	0.633	HMGN2P46	CNA	0.534	TAF15	CNA	0.494
FANCC	CNA	0.604	PMS2	CNA	0.532	MECOM	CNA	0.491

Table 27: Endometrium Carcinoma Undifferentiated – FGTP

GENE	TECH	IMP	PRRX1	CNA	0.718	GATA2	CNA	0.547
PIK3CA	NGS	1.000	IKZF1	CNA	0.717	PCM1	NGS	0.533
MAF	CNA	0.994	SLC45A3	CNA	0.713	WISP3	CNA	0.523
Gender	META	0.991	RMI2	CNA	0.705	CCNB1IP1	CNA	0.520
FOXL2	NGS	0.976	TP53	NGS	0.688	CCDC6	CNA	0.518
ELK4	CNA	0.971	CDK6	CNA	0.670	PDE4DIP	CNA	0.504
GID4	CNA	0.952	GNA13	CNA	0.663	ARHGAP26	CNA	0.499
ARID1A	NGS	0.932	AURKB	CNA	0.619	PMS2	CNA	0.493
PTEN	NGS	0.881	KDM5C	NGS	0.605	FGFR1	CNA	0.486
H3F3A	CNA	0.873	NTRK1	CNA	0.603	GNAQ	CNA	0.484
PRCC	CNA	0.804	MLLT10	CNA	0.589	ETV6	CNA	0.477
HMGN2P46	CNA	0.775	RPL22	NGS	0.587	SOX2	CNA	0.472
HSP90AA1	CNA	0.765	TGFBR2	CNA	0.587	CDK8	CNA	0.470
HIST1H3B	CNA	0.753	SDC4	CNA	0.579	HEY1	CNA	0.468
SMARCA4	NGS	0.750	MYC	CNA	0.574	SPEN	CNA	0.468
PRKDC	CNA	0.737	HIST1H4I	CNA	0.571	EXT1	CNA	0.466
Age	META	0.727	TET1	CNA	0.560	EP300	CNA	0.465

5

Table 28: Endometrium Clear Cell Carcinoma – FGTP

GENE	TECH	IMP
PAX8	CNA	1.000
FOXL2	NGS	0.950
CDK12	CNA	0.941
Gender	META	0.871
Age	META	0.853
KLF4	CNA	0.823
FNBP1	CNA	0.780
NF2	CNA	0.754
WWTR1	CNA	0.735
MECOM	CNA	0.728
CHEK2	CNA	0.716
YWHAE	CNA	0.680
KAT6A	CNA	0.679
SUFU	CNA	0.675
AFF3	CNA	0.655
EWSR1	CNA	0.646
CLTCL1	CNA	0.637
CALR	CNA	0.628
CNTRL	CNA	0.626
STAT3	CNA	0.625
FANCC	CNA	0.617
CCNE1	CNA	0.600
NR4A3	CNA	0.600
TPM4	CNA	0.597
OMD	CNA	0.596
ERBB2	CNA	0.589
MKL1	CNA	0.577
EP300	CNA	0.557
TSC1	CNA	0.555
XPA	CNA	0.534
PCSK7	CNA	0.532
PAFAH1B2	CNA	0.521
BCL6	CNA	0.518
CRKL	CNA	0.511
GNAS	CNA	0.501
FGFR2	CNA	0.499
FUS	CNA	0.498
RAC1	CNA	0.496
ZNF217	CNA	0.495
NDRG1	CNA	0.490
KRAS	NGS	0.489
SETBP1	CNA	0.488
PMS2	CNA	0.488
FANCF	CNA	0.486
PIK3CA	NGS	0.476
CDKN2A	CNA	0.474
CREB3L2	CNA	0.472
TRIP11	CNA	0.461
GNA13	CNA	0.460
RNF213	NGS	0.459

Table 29: Esophagus Adenocarcinoma NOS – Esophagus

GENE	TECH	IMP
Gender	META	1.000
SETBP1	CNA	0.943
APC	NGS	0.932
ZNF217	CNA	0.931
ERG	CNA	0.922
TP53	NGS	0.908
Age	META	0.904
CDX2	CNA	0.856
SDC4	CNA	0.849
CDK12	CNA	0.827
IRF4	CNA	0.818
CREB3L2	CNA	0.803
U2AF1	CNA	0.802
KDSR	CNA	0.801
KRAS	CNA	0.796
MYC	CNA	0.758
ERBB2	CNA	0.757
BCL2	CNA	0.757
FHIT	CNA	0.743
KIAA1549	CNA	0.726
CDKN2A	CNA	0.694
CDKN2B	CNA	0.693
RUNX1	CNA	0.693
GNAS	CNA	0.672
TRRAP	CNA	0.671
AFF1	CNA	0.671
FLT3	CNA	0.670
ERBB3	CNA	0.655
CREBBP	CNA	0.652
JAZF1	CNA	0.651
CTNNA1	CNA	0.650
FOXO1	CNA	0.633
LHFPL6	CNA	0.633
SMAD4	CNA	0.631
SMAD2	CNA	0.630
CACNA1D	CNA	0.629
HSP90AB1	CNA	0.629
WWTR1	CNA	0.620
FGFR2	CNA	0.612
ASXL1	CNA	0.605
RAC1	CNA	0.602
MLLT11	CNA	0.601
EBF1	CNA	0.600
KRAS	NGS	0.600
TCF7L2	CNA	0.595
MALT1	CNA	0.593
CTCF	CNA	0.593
PRRX1	CNA	0.591
ARID1A	CNA	0.583
KMT2C	CNA	0.573

Table 30: Esophagus Carcinoma NOS – Esophagus

GENE	TECH	IMP
ERG	CNA	1.000
FOXL2	NGS	0.946
Gender	META	0.878
PDGFRA	CNA	0.873
Age	META	0.753
PRRX1	CNA	0.740
XPC	CNA	0.740
RUNX1	CNA	0.707
TP53	NGS	0.697
TCF7L2	CNA	0.674
YWHAE	CNA	0.665
FGFR1OP	CNA	0.658
FGF19	CNA	0.642
MLF1	CNA	0.629
APC	NGS	0.624
VHL	CNA	0.602
IDH1	NGS	0.585

VHL	NGS	0.572	CACNA1D	CNA	0.479	NOTCH2	CNA	0.449
FHIT	CNA	0.569	CBFB	CNA	0.475	CDKN2B	CNA	0.447
KIT	CNA	0.544	CREB3L2	CNA	0.473	CCND1	CNA	0.446
TFRC	CNA	0.532	NUTM2B	CNA	0.470	CDK4	CNA	0.446
KRAS	NGS	0.519	SETBP1	CNA	0.467	RHOH	CNA	0.442
WWTR1	CNA	0.507	FANCC	CNA	0.466	DAXX	CNA	0.440
RPN1	CNA	0.494	AURKB	CNA	0.462	FLT1	CNA	0.435
LHFPL6	CNA	0.486	USP6	CNA	0.460	FGFR2	CNA	0.434
FGF3	CNA	0.485	U2AF1	CNA	0.456	SRGAP3	CNA	0.431
JAK1	CNA	0.484	SOX2	CNA	0.455	TGFBR2	CNA	0.431
PHOX2B	CNA	0.482	FOXP1	CNA	0.453	MLLT11	CNA	0.428

Table 31: Esophagus Squamous Carcinoma – Esophagus

GENE	TECH	IMP						
KLHL6	CNA	1.000	FGF19	CNA	0.655	EP300	CNA	0.510
TFRC	CNA	0.969	CDKN2A	CNA	0.647	BCL6	CNA	0.499
SOX2	CNA	0.923	PPARG	CNA	0.637	CDKN2B	CNA	0.498
FOXL2	NGS	0.913	SRGAP3	CNA	0.637	XPC	CNA	0.495
EPHA3	CNA	0.898	YWHAE	CNA	0.610	EBF1	CNA	0.472
FHIT	CNA	0.879	CTNNA1	CNA	0.609	IDH1	NGS	0.471
FGF3	CNA	0.869	FGF4	CNA	0.609	KRAS	NGS	0.470
CCND1	CNA	0.811	EWSR1	CNA	0.591	WWTR1	CNA	0.464
TGFBR2	CNA	0.804	MAML2	CNA	0.588	NUP214	CNA	0.462
LPP	CNA	0.799	Age	META	0.571	EZR	CNA	0.440
MITF	CNA	0.783	ERG	CNA	0.560	FOXP1	CNA	0.436
Gender	META	0.750	RAC1	CNA	0.556	VHL	CNA	0.434
TP53	NGS	0.708	VHL	NGS	0.535	MYC	CNA	0.432
CACNA1D	CNA	0.706	RPN1	CNA	0.531	RABEP1	CNA	0.431
LHFPL6	CNA	0.700	APC	NGS	0.527	RAF1	CNA	0.430
ETV5	CNA	0.666	FANCC	CNA	0.524	GID4	CNA	0.428
			TP53	CNA	0.511	BCL2	NGS	0.423

Table 32: Extrahepatic Cholangio Common Bile Gallbladder Adenocarcinoma NOS – Liver, Gallbladder, Ducts

GENE	TECH	IMP						
Age	META	1.000	KDSR	CNA	0.760	JAZF1	CNA	0.686
Gender	META	0.953	CDKN2B	CNA	0.751	ZNF217	CNA	0.685
CDK12	CNA	0.868	CACNA1D	CNA	0.744	CD274	CNA	0.683
PDCD1LG2	CNA	0.847	LHFPL6	CNA	0.733	HEY1	CNA	0.651
APC	NGS	0.842	ERG	CNA	0.729	WWTR1	CNA	0.649
USP6	CNA	0.841	TP53	NGS	0.724	CALR	CNA	0.647
YWHAE	CNA	0.780	PTPN11	CNA	0.719	CCNE1	CNA	0.644
SETBP1	CNA	0.776	VHL	NGS	0.713	KRAS	NGS	0.640
STAT3	CNA	0.772	CDKN2A	CNA	0.710	TPM4	CNA	0.639
			FOXL2	NGS	0.686	TAF15	CNA	0.631

PRRX1	CNA	0.628	WISP3	CNA	0.617	PAX3	CNA	0.583
SPEN	CNA	0.627	CBFB	CNA	0.614	RABEP1	CNA	0.583
LPP	CNA	0.626	MDM2	CNA	0.614	EXT1	CNA	0.583
MAML2	CNA	0.626	HSP90AA1	CNA	0.606	H3F3B	CNA	0.582
FANCC	CNA	0.624	RAC1	CNA	0.593	ARID1A	CNA	0.580
NFIB	CNA	0.620	BCL6	CNA	0.592	SUZ12	CNA	0.580
KLHL6	CNA	0.619	BCL2	CNA	0.584	ETV5	CNA	0.578

Table 33: Fallopian tube Adenocarcinoma NOS – FGTP

GENE	TECH	IMP						
EWSR1	CNA	1.000	WDCP	CNA	0.568	CACNA1D	CNA	0.444
CDK12	CNA	0.973	TP53	NGS	0.551	KMT2D	CNA	0.444
FOXL2	NGS	0.942	PSIP1	CNA	0.545	HLF	CNA	0.437
STAT3	CNA	0.915	CDH1	NGS	0.522	NF2	CNA	0.428
ETV6	CNA	0.910	KLHL6	CNA	0.506	GNAS	CNA	0.428
KAT6B	CNA	0.851	MKL1	CNA	0.502	CDH1	CNA	0.423
ABL1	NGS	0.815	AFF3	CNA	0.496	c-KIT	NGS	0.421
SMARCE1	CNA	0.788	CDH11	CNA	0.496	STAT5B	CNA	0.411
Gender	META	0.778	NUTM1	CNA	0.495	SS18	CNA	0.411
RPN1	CNA	0.724	CBFB	CNA	0.493	ASXL1	CNA	0.410
TFRC	CNA	0.692	EP300	CNA	0.491	BMPR1A	CNA	0.409
CCNE1	CNA	0.670	SDHC	CNA	0.478	ZNF521	CNA	0.405
LPP	CNA	0.663	CDKN1B	CNA	0.478	USP6	CNA	0.401
WWTR1	CNA	0.655	PMS2	CNA	0.475	ETV5	CNA	0.398
Age	META	0.629	MYCN	CNA	0.466	MYD88	CNA	0.397
MAP2K1	CNA	0.616	MSH2	CNA	0.465	MAF	CNA	0.396
			EPHB1	CNA	0.463	DAXX	CNA	0.394

Table 34: Fallopian tube Carcinoma NOS – FGTP

GENE	TECH	IMP						
RPN1	CNA	1.000	PICALM	CNA	0.556	EMSY	CNA	0.466
MUC1	CNA	0.926	WWTR1	CNA	0.554	GMPS	CNA	0.463
FOXL2	NGS	0.926	LYL1	CNA	0.547	BCL2	CNA	0.456
ETV5	CNA	0.919	EP300	CNA	0.546	SPECC1	CNA	0.448
Gender	META	0.871	ELK4	CNA	0.545	SLC45A3	CNA	0.448
STAT3	CNA	0.772	CARS	CNA	0.540	TSC1	CNA	0.447
TP53	NGS	0.718	PDCD1LG2	CNA	0.539	TNFAIP3	CNA	0.446
SMARCE1	CNA	0.708	FOXL2	CNA	0.522	STAT5B	CNA	0.445
NF1	CNA	0.672	ABL1	NGS	0.518	CDK12	CNA	0.444
CDH1	NGS	0.668	NUMA1	CNA	0.515	NUP214	CNA	0.440
Age	META	0.658	MECOM	CNA	0.514	c-KIT	NGS	0.436
SOX2	CNA	0.625	NTRK3	CNA	0.499	NUP93	CNA	0.436
BCL6	CNA	0.608	KLHL6	CNA	0.494	C15orf65	CNA	0.429
NUP98	CNA	0.608	RAC1	CNA	0.491	LPP	CNA	0.426
MAP2K1	CNA	0.593	NDRG1	CNA	0.478	PSIP1	CNA	0.422
			RECQL4	CNA	0.467	VHL	CNA	0.418

MSI2	CNA	0.414	APC	NGS	0.412	FGF10	CNA	0.411

Table 35: Fallopian tube Carcinosarcoma NOS – FGTP

GENE	TECH	IMP	WIF1	CNA	0.481	CDK12	CNA	0.346
ASXL1	CNA	1.000	BRD4	CNA	0.466	STK11	CNA	0.345
ABL2	NGS	0.855	ERC1	CNA	0.458	CNBP	CNA	0.340
WDCP	CNA	0.795	ATIC	CNA	0.443	WISP3	CNA	0.338
MECOM	CNA	0.768	HMGN2P46	CNA	0.432	FSTL3	CNA	0.333
BCL11A	CNA	0.724	CDH1	NGS	0.428	GATA3	CNA	0.317
FOXL2	NGS	0.703	BRCA1	CNA	0.397	MLLT11	CNA	0.315
KLF4	CNA	0.661	ARNT	CNA	0.396	GNA13	CNA	0.312
AFF3	CNA	0.643	KRAS	NGS	0.375	PMS2	CNA	0.308
DDR2	CNA	0.598	MAP2K1	CNA	0.374	MLLT3	CNA	0.302
BCL9	CNA	0.592	CTLA4	CNA	0.367	KDSR	CNA	0.301
NUTM1	CNA	0.544	VHL	NGS	0.367	FGF23	CNA	0.299
Gender	META	0.531	HMGA2	CNA	0.365	KAT6A	CNA	0.293
GNAS	CNA	0.516	PAX3	CNA	0.364	BCL2	CNA	0.286
CDKN2A	CNA	0.493	CASP8	CNA	0.354	ASPSCR1	NGS	0.277
TP53	NGS	0.493	RET	CNA	0.352	NOTCH2	CNA	0.276
APC	NGS	0.488	CCND2	CNA	0.349	CALR	CNA	0.274

Table 36: Fallopian tube Serous Carcinoma – FGTP

GENE	TECH	IMP	CDH11	CNA	0.660	GNAS	CNA	0.552
MECOM	CNA	1.000	WWTR1	CNA	0.643	SMARCE1	CNA	0.550
TP53	NGS	0.955	RAC1	CNA	0.630	MLLT11	CNA	0.549
FOXL2	NGS	0.912	RPN1	CNA	0.629	STAT5B	CNA	0.545
TPM4	CNA	0.847	ASXL1	CNA	0.625	WT1	CNA	0.543
Gender	META	0.815	CDK12	CNA	0.613	FGFR2	CNA	0.538
CCNE1	CNA	0.812	NUP214	CNA	0.604	HEY1	CNA	0.531
CBFB	CNA	0.795	TSC1	CNA	0.600	KRAS	NGS	0.531
EP300	CNA	0.753	SUZ12	CNA	0.596	CDX2	CNA	0.528
Age	META	0.753	ETV5	CNA	0.590	CACNA1D	CNA	0.528
MAF	CNA	0.750	ZNF217	CNA	0.580	NF1	CNA	0.526
CTCF	CNA	0.738	BCL9	CNA	0.578	GID4	CNA	0.519
STAT3	CNA	0.735	FSTL3	CNA	0.576	BRD4	CNA	0.516
BCL6	CNA	0.700	TET2	CNA	0.573	CRKL	CNA	0.516
KLHL6	CNA	0.696	GNA11	CNA	0.572	KLF4	CNA	0.507
TAF15	CNA	0.675	PMS2	CNA	0.562	SRSF2	CNA	0.505
CDH1	CNA	0.671	EWSR1	CNA	0.560	AFF3	CNA	0.502

5

Table 37: Gastric Adenocarcinoma – Stomach

GENE	TECH	IMP	ERG	CNA	0.989	U2AF1	CNA	0.956
Age	META	1.000	FOXL2	NGS	0.962	CDX2	CNA	0.881

CDKN2B	CNA	0.866
ZNF217	CNA	0.850
EXT1	CNA	0.840
CACNA1D	CNA	0.825
LHFPL6	CNA	0.820
Gender	META	0.815
CDH1	NGS	0.807
SPECC1	CNA	0.799
FOXO1	CNA	0.795
CDKN2A	CNA	0.779
KRAS	NGS	0.751
FHIT	CNA	0.749
SETBP1	CNA	0.745
PRRX1	CNA	0.742
SDC4	CNA	0.739

TP53	NGS	0.738
IKZF1	CNA	0.737
TCF7L2	CNA	0.736
EWSR1	CNA	0.725
CBFB	CNA	0.725
WWTR1	CNA	0.723
MYC	CNA	0.721
KLHL6	CNA	0.719
FLT3	CNA	0.717
HMGN2P46	CNA	0.716
RUNX1	CNA	0.715
PMS2	CNA	0.713
MLLT11	CNA	0.709
JAZF1	CNA	0.704
EBF1	CNA	0.703

KDSR	CNA	0.703
CDK6	CNA	0.701
USP6	CNA	0.697
RAC1	CNA	0.690
FGFR2	CNA	0.685
FANCC	CNA	0.679
CDH11	CNA	0.678
XPC	CNA	0.677
CREB3L2	CNA	0.676
BCL2	CNA	0.673
FANCF	CNA	0.672
SBDS	CNA	0.670
CDK12	CNA	0.670
PPARG	CNA	0.669
TGFBR2	CNA	0.665

Table 38: Gastroesophageal junction Adenocarcinoma NOS – Esophagus

GENE	TECH	IMP
ERG	CNA	1.000
FOXL2	NGS	0.979
U2AF1	CNA	0.966
Gender	META	0.902
CDK12	CNA	0.896
Age	META	0.858
ZNF217	CNA	0.830
CREB3L2	CNA	0.828
ERBB2	CNA	0.793
SDC4	CNA	0.778
CDX2	CNA	0.776
RUNX1	CNA	0.764
ASXL1	CNA	0.742
EBF1	CNA	0.735
CACNA1D	CNA	0.734
KIAA1549	CNA	0.730

KDSR	CNA	0.720
EWSR1	CNA	0.712
RAC1	CNA	0.709
SETBP1	CNA	0.702
TP53	NGS	0.692
ARID1A	CNA	0.682
JAZF1	CNA	0.679
FHIT	CNA	0.676
CTNNA1	CNA	0.675
CDKN2A	CNA	0.670
GNAS	CNA	0.662
KRAS	NGS	0.661
IRF4	CNA	0.660
MYC	CNA	0.654
ACSL6	CNA	0.638
FNBP1	CNA	0.636
CBFB	CNA	0.636

LHFPL6	CNA	0.634
CHEK2	CNA	0.621
PCM1	CNA	0.619
RPN1	CNA	0.618
HOXA11	CNA	0.614
TCF7L2	CNA	0.612
SRGAP3	CNA	0.595
KLHL6	CNA	0.593
FGFR2	CNA	0.592
HOXD13	CNA	0.584
HOXA13	CNA	0.583
CRTC3	CNA	0.580
TOP1	CNA	0.576
WRN	CNA	0.575
CCNE1	CNA	0.574
CDKN2B	CNA	0.571
CDH11	CNA	0.566

Table 39: Glioblastoma – Brain

GENE	TECH	IMP
FGFR2	CNA	1.000
EGFR	CNA	0.993
FOXL2	NGS	0.953
TCF7L2	CNA	0.912
OLIG2	CNA	0.910
VTI1A	CNA	0.896
SBDS	CNA	0.889

Age	META	0.870
CDKN2A	CNA	0.820
PDGFRA	CNA	0.809
TET1	CNA	0.801
MYC	CNA	0.791
CREB3L2	CNA	0.787
CCDC6	CNA	0.779
SOX2	CNA	0.773

EXT1	CNA	0.756
TRRAP	CNA	0.755
CDKN2B	CNA	0.749
KAT6B	CNA	0.741
CDK6	CNA	0.738
SPECC1	CNA	0.734
JAZF1	CNA	0.719
NFKB2	CNA	0.713

NDRG1	CNA	0.711
GATA3	CNA	0.684
TPM3	CNA	0.683
NT5C2	CNA	0.668
HMGA2	CNA	0.660
KIT	CNA	0.658
ZNF217	CNA	0.658
FOXO1	CNA	0.657
KIAA1549	CNA	0.633

Gender	META	0.618
SPEN	CNA	0.614
ETV1	CNA	0.605
MCL1	CNA	0.598
NCOA2	CNA	0.594
FGF14	CNA	0.588
SUFU	CNA	0.585
KMT2C	CNA	0.582
PIK3CG	CNA	0.576

NUP214	CNA	0.570
IDH1	NGS	0.568
MET	CNA	0.568
TP53	NGS	0.564
HIP1	CNA	0.558
PTEN	CNA	0.550
PTEN	NGS	0.542
LCP1	CNA	0.528
LHFPL6	CNA	0.522

Table 40: Glioma NOS – Brain

GENE	TECH	IMP
Age	META	1.000
IDH1	NGS	0.871
FOXL2	NGS	0.738
Gender	META	0.709
CREB3L2	CNA	0.685
SETBP1	CNA	0.657
SOX2	CNA	0.656
PDGFRA	CNA	0.645
c-KIT	NGS	0.640
PDGFRA	NGS	0.612
TPM3	CNA	0.605
VHL	NGS	0.594
SPECC1	CNA	0.588
CDH1	NGS	0.571
STK11	CNA	0.567
MYC	CNA	0.556

OLIG2	CNA	0.549
KIAA1549	CNA	0.537
CDX2	CNA	0.536
VTI1A	CNA	0.533
KRAS	NGS	0.532
CDKN2B	CNA	0.531
CDKN2A	CNA	0.521
PIK3R1	CNA	0.515
EGFR	CNA	0.513
APC	NGS	0.493
TCF7L2	CNA	0.482
TP53	NGS	0.480
NDRG1	CNA	0.471
TERT	CNA	0.464
MSI2	CNA	0.459
SBDS	CNA	0.458
PMS2	CNA	0.449

KDR	CNA	0.448
MCL1	CNA	0.432
FAM46C	CNA	0.425
NR4A3	CNA	0.421
RPL22	CNA	0.420
CDK6	CNA	0.406
MYCL	CNA	0.406
PDE4DIP	CNA	0.405
KAT6B	CNA	0.402
IRF4	CNA	0.397
NFKB2	CNA	0.391
H3F3A	CNA	0.387
HMGA2	CNA	0.387
KIT	CNA	0.374
EIF4A2	CNA	0.374
EZH2	CNA	0.372
NT5C2	CNA	0.361

Table 41: Gliosarcoma – Brain

GENE	TECH	IMP
IKZF1	CNA	1.000
PTEN	NGS	0.916
FOXL2	NGS	0.899
CDH1	NGS	0.817
CREB3L2	CNA	0.774
TRRAP	CNA	0.732
NF1	NGS	0.713
CCDC6	CNA	0.703
JAZF1	CNA	0.619
TET1	CNA	0.604
Age	META	0.582
CDK6	CNA	0.575
MLLT10	CNA	0.550

ETV1	CNA	0.549
KAT6B	CNA	0.540
FGFR2	CNA	0.531
CDK12	CNA	0.510
SS18	CNA	0.504
EGFR	CNA	0.503
GATA3	CNA	0.492
EBF1	CNA	0.489
MYC	CNA	0.482
PDGFRA	CNA	0.480
VHL	NGS	0.477
RAC1	CNA	0.474
KRAS	NGS	0.466
KIF5B	CNA	0.461

NTRK2	CNA	0.448
ELK4	CNA	0.425
FHIT	CNA	0.423
ABI1	CNA	0.421
SOX10	CNA	0.416
Gender	META	0.416
ERG	CNA	0.415
c-KIT	NGS	0.409
TCF7L2	CNA	0.405
MSH2	NGS	0.404
VTI1A	CNA	0.402
KIAA1549	CNA	0.401
NR4A3	CNA	0.397
COX6C	CNA	0.396

CBFB	CNA	0.390
FOXP1	CNA	0.380
CDX2	CNA	0.378
STAT3	CNA	0.376
APC	NGS	0.371
ATP1A1	CNA	0.371
RBM15	CNA	0.368
IRF4	CNA	0.368
SOX2	CNA	0.360

Table 42: Head, face or neck NOS Squamous carcinoma – Head, face or neck, NOS

GENE	TECH	IMP
Gender	META	1.000
ETV5	CNA	0.977
KLHL6	CNA	0.947
NOTCH1	NGS	0.930
FOXL2	NGS	0.922
MN1	CNA	0.898
EWSR1	CNA	0.891
LPP	CNA	0.846
NF2	CNA	0.824
BCL6	CNA	0.786
WWTR1	CNA	0.728
Age	META	0.712
SOX2	CNA	0.704
MAML2	CNA	0.697
ATIC	CNA	0.689
MECOM	CNA	0.684
TFRC	CNA	0.666
MLF1	CNA	0.655
FNBP1	CNA	0.648
ARID1A	CNA	0.609
CDH1	CNA	0.609
NOTCH2	NGS	0.589
PAFAH1B2	CNA	0.584
SET	CNA	0.563
NDRG1	CNA	0.563
CDKN2A	CNA	0.560
GMPS	CNA	0.557
FGF3	CNA	0.552
CDKN2A	NGS	0.535
TBL1XR1	CNA	0.534
SPEN	CNA	0.523
KRAS	NGS	0.516
BCL9	CNA	0.503
TP53	NGS	0.501
CRKL	CNA	0.498
SETBP1	CNA	0.494
MAF	CNA	0.493
FAS	CNA	0.491
NTRK2	CNA	0.485
CREB3L2	CNA	0.484
FOXP1	CNA	0.483
JUN	CNA	0.482
PAX3	CNA	0.473
FLT1	CNA	0.466
GID4	CNA	0.464
DDX6	CNA	0.458
FLI1	CNA	0.451
FGF19	CNA	0.451
TSC1	CNA	0.447
ZBTB16	CNA	0.442

Table 43: Intrahepatic bile duct Cholangiocarcinoma – Liver, Gallbladder, Ducts

GENE	TECH	IMP
MDS2	CNA	1.000
Age	META	0.992
ARID1A	CNA	0.983
CACNA1D	CNA	0.975
FHIT	CNA	0.957
APC	NGS	0.952
MAF	CNA	0.948
CAMTA1	CNA	0.921
TP53	NGS	0.898
MTOR	CNA	0.857
VHL	NGS	0.851
ESR1	CNA	0.851
STAT3	CNA	0.834
CDKN2B	CNA	0.834
EZR	CNA	0.832
TSHR	CNA	0.829
Gender	META	0.821
CDKN2A	CNA	0.808
SPEN	CNA	0.799
U2AF1	CNA	0.799
PBRM1	CNA	0.794
NOTCH2	CNA	0.760
ELK4	CNA	0.755
ERG	CNA	0.747
MSI2	CNA	0.742
SDHB	CNA	0.740
TAF15	CNA	0.733
CDK12	CNA	0.733
FANCC	CNA	0.730
RPL22	CNA	0.725
LHFPL6	CNA	0.725
PTCH1	CNA	0.722
SETBP1	CNA	0.714
BCL3	CNA	0.713
KRAS	NGS	0.712
FANCF	CNA	0.705
WISP3	CNA	0.698
TGFBR2	CNA	0.696
FOXP1	CNA	0.696
NR4A3	CNA	0.694
EXT1	CNA	0.692
CBFB	CNA	0.691
ECT2L	CNA	0.686
MYB	CNA	0.686
FOXL2	NGS	0.686
ZNF331	CNA	0.683
ETV5	CNA	0.683
NTRK2	CNA	0.683
SRGAP3	CNA	0.681
ZNF217	CNA	0.676
MYC	CNA	0.673
LPP	CNA	0.673
IL2	CNA	0.673

5

Table 44: Kidney Carcinoma NOS – Kidney

GENE	TECH	IMP
EBF1	CNA	1.000
BTG1	CNA	0.971
FOXL2	NGS	0.931
FHIT	CNA	0.817
VHL	NGS	0.810
TP53	NGS	0.797
XPC	CNA	0.772
MAF	CNA	0.765
GID4	CNA	0.712
MYCN	CNA	0.671
SDHAF2	CNA	0.639
Gender	META	0.633
FANCC	CNA	0.626
CTNNA1	CNA	0.624
FANCA	CNA	0.622
SDHB	CNA	0.608
CDH11	CNA	0.593
CDKN1B	CNA	0.580
MAML2	CNA	0.564
CBFB	CNA	0.560
FGF23	CNA	0.558
Age	META	0.558
CNBP	CNA	0.555
FGF14	CNA	0.553
FGFR1OP	CNA	0.544
FAM46C	CNA	0.540
WWTR1	CNA	0.533
MTOR	CNA	0.528
USP6	CNA	0.520
TFRC	CNA	0.520
SPECC1	CNA	0.518
PAX3	CNA	0.516
HMGA2	CNA	0.513
ITK	CNA	0.505
HOXD13	CNA	0.502
SPEN	CNA	0.501
RMI2	CNA	0.497
CD74	CNA	0.494
HOXA13	CNA	0.494
MYC	CNA	0.489
CREBBP	CNA	0.477
c-KIT	NGS	0.475
ARID1A	CNA	0.467
EXT1	CNA	0.457
KRAS	NGS	0.452
ACSL6	CNA	0.452
CRKL	CNA	0.451
RAF1	CNA	0.446
BCL9	CNA	0.439
GNA13	CNA	0.437

Table 45: Kidney Clear Cell Carcinoma – Kidney

GENE	TECH	IMP
VHL	NGS	1.000
FOXL2	NGS	0.743
TP53	NGS	0.618
EBF1	CNA	0.577
VHL	CNA	0.569
XPC	CNA	0.535
MYD88	CNA	0.517
Gender	META	0.495
c-KIT	NGS	0.490
ITK	CNA	0.481
SRGAP3	CNA	0.446
MDM4	CNA	0.431
RAF1	CNA	0.430
ARNT	CNA	0.428
CTNNA1	CNA	0.411
TGFBR2	CNA	0.405
MLLT11	CNA	0.403
PRCC	CNA	0.382
Age	META	0.366
MAF	CNA	0.357
KRAS	NGS	0.349
APC	NGS	0.338
USP6	CNA	0.325
CDKN2A	CNA	0.319
PTPN11	CNA	0.312
MCL1	CNA	0.298
IL21R	CNA	0.296
RPN1	CNA	0.291
KDSR	CNA	0.289
PAX3	CNA	0.275
MUC1	CNA	0.273
STAT5B	NGS	0.265
MAX	CNA	0.265
CDH11	CNA	0.264
ABL2	CNA	0.264
HMGN2P46	CNA	0.261
CBLB	CNA	0.260
TSHR	CNA	0.259
YWHAE	CNA	0.254
SETD2	NGS	0.254
PPARG	CNA	0.252
ZNF217	CNA	0.247
TRIM33	NGS	0.247
SETBP1	CNA	0.245
CACNA1D	CNA	0.244
BTG1	CNA	0.242
CYP2D6	CNA	0.240
NUTM2B	CNA	0.239
FANCD2	CNA	0.238
BCL2	CNA	0.238

Table 46: Kidney Papillary Renal Cell Carcinoma – Kidney

GENE	TECH	IMP
MSI2	CNA	1.000
Gender	META	0.945
FOXL2	NGS	0.914
c-KIT	NGS	0.899
TP53	NGS	0.890
CREB3L2	CNA	0.873
HLF	CNA	0.825
SRSF2	CNA	0.763
IDH1	NGS	0.739
GNA13	CNA	0.717
AURKB	CNA	0.661

VHL	NGS	0.652
CDX2	CNA	0.619
APC	NGS	0.592
MAF	CNA	0.591
SNX29	CNA	0.584
KRAS	NGS	0.568
H3F3B	CNA	0.561
TPM3	CNA	0.559
PER1	CNA	0.525
KIAA1549	CNA	0.513
YWHAE	CNA	0.505
NKX2-1	CNA	0.491
CLTC	CNA	0.488
IRF4	CNA	0.478
STAT3	CNA	0.477
BRAF	CNA	0.476
EXT1	CNA	0.452
NUP93	CNA	0.451
SOX10	CNA	0.440
TAF15	CNA	0.428
RECQL4	CNA	0.425
Age	META	0.419
PRCC	CNA	0.419
RNF213	CNA	0.411
SPEN	CNA	0.411
RMI2	CNA	0.402
CBFB	CNA	0.397
CRKL	CNA	0.392
COX6C	CNA	0.391
DDX5	CNA	0.387
BCL7A	CNA	0.387
SRSF3	CNA	0.385
ERCC4	CNA	0.380
MAP2K4	CNA	0.367
SMARCE1	CNA	0.366
MLLT11	CNA	0.366
PRKAR1A	CNA	0.366
BRIP1	CNA	0.365
ASXL1	CNA	0.365

Table 47: Kidney Renal Cell Carcinoma NOS – Kidney

GENE	TECH	IMP
VHL	NGS	1.000
RAF1	CNA	0.977
EBF1	CNA	0.971
MAF	CNA	0.968
CTNNA1	CNA	0.939
FOXL2	NGS	0.916
TP53	NGS	0.898
c-KIT	NGS	0.870
SRGAP3	CNA	0.852
MUC1	CNA	0.831
XPC	CNA	0.826
Gender	META	0.807
NUP93	CNA	0.760
VHL	CNA	0.740
MTOR	CNA	0.710
Age	META	0.709
ITK	CNA	0.683
FLI1	CNA	0.666
CDH11	CNA	0.660
CACNA1D	CNA	0.654
FANCC	CNA	0.648
ACSL6	CNA	0.647
TRIM27	CNA	0.637
FANCF	CNA	0.630
FNBP1	CNA	0.623
CBFB	CNA	0.605
PDGFRA	NGS	0.598
CDX2	CNA	0.598
MLLT11	CNA	0.594
KRAS	NGS	0.577
CREB3L2	CNA	0.574
FANCD2	CNA	0.573
FHIT	CNA	0.573
TSC1	CNA	0.566
NUP214	CNA	0.563
KIAA1549	CNA	0.560
HSP90AA1	CNA	0.559
TPM3	CNA	0.556
ABL2	CNA	0.554
APC	NGS	0.548
SPEN	CNA	0.544
ETV5	CNA	0.540
BTG1	CNA	0.535
ZNF217	CNA	0.532
CD74	CNA	0.518
SNX29	CNA	0.513
PPARG	CNA	0.510
RANBP17	CNA	0.508
ARHGAP26	CNA	0.507
ARFRP1	NGS	0.505

Table 48: Larynx NOS Squamous carcinoma – Head, Face or Neck, NOS

GENE	TECH	IMP
TGFBR2	CNA	1.000
Gender	META	0.979
FOXL2	NGS	0.949
ETV5	CNA	0.896
KLHL6	CNA	0.803
BCL6	CNA	0.787
HMGN2P46	CNA	0.755
YWHAE	CNA	0.749
TFRC	CNA	0.745
EGFR	CNA	0.727
USP6	CNA	0.723
WWTR1	CNA	0.698
VHL	NGS	0.697
RAF1	CNA	0.683
SOX2	CNA	0.682
FOXP1	CNA	0.673
SETD2	CNA	0.660
NF2	CNA	0.644
MYD88	CNA	0.601
PIK3CA	CNA	0.592
LPP	CNA	0.589
VHL	CNA	0.561
CREB3L2	CNA	0.557
Age	META	0.557
CACNA1D	CNA	0.551
TP53	NGS	0.534
GNAS	CNA	0.533
FHIT	CNA	0.528
KRAS	NGS	0.525

MECOM	CNA	0.511
GID4	CNA	0.511
TBL1XR1	CNA	0.474
FLT3	CNA	0.473
SPECC1	CNA	0.470
CDKN2A	CNA	0.466
RABEP1	CNA	0.445

TOP1	CNA	0.438
EWSR1	CNA	0.433
ZNF217	CNA	0.419
EXT1	CNA	0.415
XPC	CNA	0.412
CTNNB1	CNA	0.402
PPARG	CNA	0.396

CAMTA1	CNA	0.394
FANCC	CNA	0.390
CHEK2	CNA	0.389
CDKN2A	NGS	0.385
CDH1	CNA	0.384
RUNX1	CNA	0.375
SETBP1	CNA	0.369

Table 49: Left Colon Adenocarcinoma NOS – Colon

GENE	TECH	IMP
CDX2	CNA	1.000
APC	NGS	0.989
FLT1	CNA	0.824
FOXL2	NGS	0.821
FLT3	CNA	0.793
SETBP1	CNA	0.773
BCL2	CNA	0.738
KRAS	NGS	0.733
Age	META	0.708
LHFPL6	CNA	0.696
ZNF521	CNA	0.664
ASXL1	CNA	0.649
SDC4	CNA	0.649
KDSR	CNA	0.644
CDK8	CNA	0.644
TOP1	CNA	0.621

CDH1	CNA	0.595
ZNF217	CNA	0.585
ZMYM2	CNA	0.585
CDKN2B	CNA	0.575
RB1	CNA	0.566
GNAS	CNA	0.557
HOXA9	CNA	0.548
SMAD4	CNA	0.547
SOX2	CNA	0.543
WWTR1	CNA	0.536
JAZF1	CNA	0.530
Gender	META	0.518
ERCC5	CNA	0.505
HOXA11	CNA	0.498
MSI2	CNA	0.497
FOXO1	CNA	0.492
WRN	CNA	0.487

TP53	NGS	0.485
COX6C	CNA	0.482
CDKN2A	CNA	0.479
LCP1	CNA	0.478
ETV5	CNA	0.475
PDE4DIP	CNA	0.467
PMS2	CNA	0.465
U2AF1	CNA	0.463
AURKA	CNA	0.460
RAC1	CNA	0.453
EBF1	CNA	0.452
BCL6	CNA	0.447
SPECC1	CNA	0.444
EP300	CNA	0.443
SS18	CNA	0.439
PTCH1	CNA	0.434
HOXA13	CNA	0.433

Table 50: Left Colon Mucinous Adenocarcinoma – Colon

GENE	TECH	IMP
APC	NGS	1.000
FOXL2	NGS	0.909
CDX2	CNA	0.902
KRAS	NGS	0.845
LHFPL6	CNA	0.814
CDK8	CNA	0.688
Age	META	0.661
Gender	META	0.658
FLT1	CNA	0.657
FLT3	CNA	0.638
ETV5	CNA	0.609
FANCC	CNA	0.605
SMAD4	NGS	0.594
SET	CNA	0.592
NTRK2	CNA	0.586

TOP1	CNA	0.586
WWTR1	CNA	0.582
SDHAF2	CNA	0.563
CDKN2A	CNA	0.527
HOXA9	CNA	0.525
SETBP1	CNA	0.522
SOX2	CNA	0.519
ABL1	CNA	0.510
CAMTA1	CNA	0.497
CDKN2B	CNA	0.494
SYK	CNA	0.484
PTCH1	CNA	0.472
VHL	NGS	0.455
MLLT3	CNA	0.446
BCL2	CNA	0.439
MAX	CNA	0.430

MYD88	CNA	0.421
MUC1	CNA	0.414
CACNA1D	CNA	0.412
WISP3	CNA	0.403
AFF3	CNA	0.396
MLLT11	CNA	0.395
RNF213	CNA	0.391
SDHB	CNA	0.384
ASXL1	CNA	0.384
TP53	NGS	0.382
ZNF217	CNA	0.379
FGF14	CNA	0.378
NF2	CNA	0.377
CDK12	CNA	0.376
CCNE1	CNA	0.370
IRS2	CNA	0.368

RPN1	CNA	0.366	ERG	CNA	0.365	GATA3	CNA	0.359

Table 51: Liver Hepatocellular Carcinoma NOS – Liver, Gallbladder, Ducts

GENE	TECH	IMP						
PRCC	CNA	1.000	COX6C	CNA	0.742	ETV6	CNA	0.651
HLF	CNA	0.992	NSD1	CNA	0.741	FLT1	CNA	0.637
FOXL2	NGS	0.981	HMGN2P46	CNA	0.732	KRAS	NGS	0.636
SDHC	CNA	0.955	YWHAE	CNA	0.727	ABL2	CNA	0.636
Gender	META	0.901	TRIM26	CNA	0.713	HIST1H4I	CNA	0.636
BCL9	CNA	0.894	SPEN	CNA	0.707	HEY1	CNA	0.636
ELK4	CNA	0.863	CACNA1D	CNA	0.706	BTG1	CNA	0.633
ERG	CNA	0.852	TPM3	CNA	0.704	AFF1	CNA	0.633
MLLT11	CNA	0.834	H3F3A	CNA	0.698	ZNF703	CNA	0.631
FGFR1	CNA	0.814	ACSL6	CNA	0.691	TP53	NGS	0.630
WRN	CNA	0.813	NCOA2	CNA	0.678	APC	NGS	0.627
Age	META	0.802	TRIM27	CNA	0.675	CDH11	CNA	0.617
CAMTA1	CNA	0.771	USP6	CNA	0.674	CDKN2A	CNA	0.613
FANCF	CNA	0.763	LHFPL6	CNA	0.669	MCL1	CNA	0.612
PCM1	CNA	0.762	MTOR	CNA	0.669	KLHL6	CNA	0.610
NSD3	CNA	0.746	EXT1	CNA	0.667	IRF4	CNA	0.601
			MECOM	CNA	0.651	ADGRA2	CNA	0.600

Table 52: Lung Adenocarcinoma NOS – Lung

GENE	TECH	IMP						
NKX2-1	CNA	1.000	HMGN2P46	CNA	0.578	CDK12	CNA	0.543
Age	META	0.890	FANCC	CNA	0.577	FLI1	CNA	0.542
TPM4	CNA	0.707	PPARG	CNA	0.575	YWHAE	CNA	0.540
TERT	CNA	0.685	CDKN2B	CNA	0.574	RAC1	CNA	0.540
KRAS	NGS	0.671	SDHC	CNA	0.572	XPC	CNA	0.535
CALR	CNA	0.667	IL7R	CNA	0.571	APC	NGS	0.529
MUC1	CNA	0.660	FGF10	CNA	0.571	TP53	NGS	0.525
Gender	META	0.656	CACNA1D	CNA	0.571	WWTR1	CNA	0.522
VHL	NGS	0.655	KDSR	CNA	0.562	FHIT	CNA	0.522
NFKBIA	CNA	0.625	TPM3	CNA	0.559	JAZF1	CNA	0.520
USP6	CNA	0.624	ASXL1	CNA	0.557	IKZF1	CNA	0.519
FOXA1	CNA	0.608	BCL2	CNA	0.555	NUTM2B	CNA	0.516
CDKN2A	CNA	0.607	SLC34A2	CNA	0.554	CCNE1	CNA	0.515
LHFPL6	CNA	0.606	EWSR1	CNA	0.550	CDKN1B	CNA	0.515
ESR1	CNA	0.588	WISP3	CNA	0.547	ELK4	CNA	0.514
FGFR2	CNA	0.585	PTCH1	CNA	0.547	LIFR	CNA	0.514
PMS2	CNA	0.579	MLLT11	CNA	0.547	SYK	CNA	0.513
BCL9	CNA	0.579	MCL1	CNA	0.546	LRP1B	NGS	0.512
SETBP1	CNA	0.578	SRGAP3	CNA	0.543			
			CDX2	CNA	0.543			

Table 53: Lung Adenosquamous Carcinoma – Lung

GENE	TECH	IMP
Age	META	1.000
FOXL2	NGS	0.928
TERT	CNA	0.848
CDKN2A	CNA	0.795
LRP1B	NGS	0.788
RUNX1	CNA	0.756
FLI1	CNA	0.756
CALR	CNA	0.746
ELK4	CNA	0.709
CACNA1D	CNA	0.707
CDKN2B	CNA	0.699
IL7R	CNA	0.695
MAML2	CNA	0.666
FANCC	CNA	0.645
HIST1H3B	CNA	0.634
Gender	META	0.631
FNBP1	CNA	0.614
FHIT	CNA	0.599
NKX2-1	CNA	0.583
MYD88	CNA	0.573
ERBB3	CNA	0.557
RHOH	CNA	0.556
PTPN11	CNA	0.549
TP53	NGS	0.549
LHFPL6	CNA	0.546
CDK4	CNA	0.541
NTRK2	CNA	0.541
FOXA1	CNA	0.537
SDHD	CNA	0.536
MAX	CNA	0.533
CBFB	CNA	0.528
USP6	CNA	0.520
KRAS	NGS	0.512
GNAS	CNA	0.511
KIT	CNA	0.509
PPARG	CNA	0.509
SOX2	CNA	0.503
CDX2	CNA	0.498
C15orf65	CNA	0.496
GNA13	CNA	0.496
EPHA3	CNA	0.483
APC	NGS	0.472
MLH1	CNA	0.470
RAF1	CNA	0.470
RPN1	CNA	0.468
MLLT11	CNA	0.465
VHL	NGS	0.462
HMGA2	CNA	0.457
MECOM	CNA	0.457
FLT1	CNA	0.456

Table 54: Lung Carcinoma NOS – Lung

GENE	TECH	IMP
Age	META	1.000
CDX2	CNA	0.870
FOXA1	CNA	0.798
VHL	NGS	0.777
KRAS	NGS	0.756
NKX2-1	CNA	0.742
APC	NGS	0.741
TP53	NGS	0.731
CALR	CNA	0.728
TPM4	CNA	0.726
CTNNA1	CNA	0.720
CACNA1D	CNA	0.719
Gender	META	0.687
FGFR2	CNA	0.672
ATP1A1	CNA	0.672
CDKN2A	CNA	0.660
XPC	CNA	0.647
SRGAP3	CNA	0.642
FHIT	CNA	0.641
FOXL2	NGS	0.640
TERT	CNA	0.628
ARID1A	CNA	0.627
LRP1B	NGS	0.625
BRD4	CNA	0.620
MSI2	CNA	0.620
FGF10	CNA	0.616
CDKN2B	CNA	0.614
LHFPL6	CNA	0.613
RPN1	CNA	0.613
PBX1	CNA	0.608
PCM1	CNA	0.607
WWTR1	CNA	0.606
FLT3	CNA	0.605
IL7R	CNA	0.603
HMGN2P46	CNA	0.597
CDK4	CNA	0.594
SETBP1	CNA	0.594
FLT1	CNA	0.592
RBM15	CNA	0.591
USP6	CNA	0.590
TRIM27	CNA	0.583
CDK12	CNA	0.581
TGFBR2	CNA	0.580
RAC1	CNA	0.577
PPARG	CNA	0.574
FANCC	CNA	0.573
CDKN1B	CNA	0.569
MYC	CNA	0.566
STAT3	CNA	0.566
MLLT11	CNA	0.564

Table 55: Lung Mucinous Adenocarcinoma – Lung

GENE	TECH	IMP
KRAS	NGS	1.000
Age	META	0.880
FOXL2	NGS	0.818
CDKN2B	CNA	0.687
TP53	NGS	0.636
CDKN2A	CNA	0.634
TPM4	CNA	0.626
ASXL1	CNA	0.624
Gender	META	0.614
IGF1R	CNA	0.596
C15orf65	CNA	0.593

BCL6	CNA	0.587	CTNNA1	CNA	0.483	CCNE1	CNA	0.448
CRKL	CNA	0.586	FLI1	CNA	0.481	ACSL6	CNA	0.446
HMGN2P46	CNA	0.550	ARHGAP26	CNA	0.477	BTG1	CNA	0.443
EBF1	CNA	0.534	CRTC3	CNA	0.474	CDH1	CNA	0.437
ETV5	CNA	0.526	EIF4A2	CNA	0.472	EPHB1	CNA	0.436
RPN1	CNA	0.519	CBFB	CNA	0.469	STK11	NGS	0.428
LPP	CNA	0.518	NUTM2B	CNA	0.468	TPM3	CNA	0.427
EXT1	CNA	0.512	ZNF521	CNA	0.467	GID4	CNA	0.419
SETBP1	CNA	0.512	CDK6	CNA	0.457	NUTM1	CNA	0.417
LHFPL6	CNA	0.511	FANCC	CNA	0.456	TRIM33	NGS	0.416
MAP2K1	CNA	0.509	FOXA1	CNA	0.456	EP300	CNA	0.416
ELK4	CNA	0.501	MLF1	CNA	0.450	FLT3	CNA	0.413
SDHC	CNA	0.484	APC	NGS	0.450	MUC1	CNA	0.408

Table 56: Lung Neuroendocrine Carcinoma NOS – Lung

GENE	TECH	IMP						
NKX2-1	CNA	1.000	RPL22	CNA	0.681	MSI2	CNA	0.580
FOXL2	NGS	0.955	FANCC	CNA	0.680	FOXO1	CNA	0.578
CAMTA1	CNA	0.870	MYD88	CNA	0.677	FLT1	CNA	0.574
VHL	CNA	0.813	PRF1	CNA	0.653	CDKN2C	CNA	0.562
PBRM1	CNA	0.801	FANCD2	CNA	0.650	ZNF217	CNA	0.553
TGFBR2	CNA	0.798	RB1	NGS	0.645	MYC	CNA	0.528
KDSR	CNA	0.752	BTG1	CNA	0.640	BCL2	CNA	0.515
SFPQ	CNA	0.751	HMGN2P46	CNA	0.634	CACNA1D	CNA	0.487
FANCG	CNA	0.746	TCF7L2	CNA	0.631	FLI1	CNA	0.481
FOXA1	CNA	0.739	LHFPL6	CNA	0.626	RAF1	CNA	0.481
SUFU	CNA	0.731	WWTR1	CNA	0.623	CDKN1B	CNA	0.477
SETBP1	CNA	0.730	FHIT	CNA	0.622	CDKN2A	CNA	0.463
PRRX1	CNA	0.702	Age	META	0.616	CDK4	CNA	0.462
XPC	CNA	0.701	MYCL	CNA	0.612	DDX5	CNA	0.461
BAP1	CNA	0.691	HIST1H3B	CNA	0.603	BCL9	CNA	0.460
FGFR2	CNA	0.682	PPARG	CNA	0.599	FLT3	CNA	0.451
			Gender	META	0.598	CDX2	CNA	0.451

Table 57: Lung Non-small Cell Carcinoma – Lung

GENE	TECH	IMP						
Age	META	1.000	FLT1	CNA	0.722	CDKN2A	CNA	0.650
NKX2-1	CNA	0.831	Gender	META	0.706	FGFR2	CNA	0.647
TP53	NGS	0.827	LHFPL6	CNA	0.697	BCL9	CNA	0.643
CDX2	CNA	0.800	HMGN2P46	CNA	0.692	KRAS	NGS	0.625
TERT	CNA	0.786	FLT3	CNA	0.682	CALR	CNA	0.624
TPM4	CNA	0.783	EWSR1	CNA	0.677	PTCH1	CNA	0.621
VHL	NGS	0.764	FANCC	CNA	0.667	CDKN2B	CNA	0.620
CTNNA1	CNA	0.741	FOXA1	CNA	0.662	GNA13	CNA	0.611
APC	NGS	0.735	FGF10	CNA	0.661	LRP1B	NGS	0.603
			CACNA1D	CNA	0.660	IKZF1	CNA	0.603

ARID1A	CNA	0.602	LIFR	CNA	0.576	PPARG	CNA	0.564
MSI2	CNA	0.601	EBF1	CNA	0.575	WWTR1	CNA	0.556
SRSF2	CNA	0.599	IL7R	CNA	0.573	KLHL6	CNA	0.556
SETBP1	CNA	0.593	MCL1	CNA	0.572	HEY1	CNA	0.550
RAC1	CNA	0.591	SPECC1	CNA	0.569	MUC1	CNA	0.547
MITF	CNA	0.590	VTI1A	CNA	0.567	SRGAP3	CNA	0.546
TGFBR2	CNA	0.590	BRD4	CNA	0.566	HMGA2	CNA	0.546
ZNF217	CNA	0.579	CCNE1	CNA	0.565	BTG1	CNA	0.545
FHIT	CNA	0.577	PAX8	CNA	0.565			
XPC	CNA	0.576	IRF4	CNA	0.565			

Table 58: Lung Sarcomatoid Carcinoma – Lung

GENE	TECH	IMP						
Age	META	1.000	BTG1	CNA	0.618	FCRL4	CNA	0.509
YWHAE	CNA	0.964	FANCC	CNA	0.617	JAK2	CNA	0.502
FOXL2	NGS	0.930	PRCC	CNA	0.614	MAML2	CNA	0.494
RAC1	CNA	0.915	LRP1B	NGS	0.602	WRN	NGS	0.486
KRAS	NGS	0.857	PBX1	CNA	0.600	FANCF	CNA	0.481
RHOH	CNA	0.855	c-KIT	NGS	0.588	KDM5C	NGS	0.472
CNBP	CNA	0.788	SPECC1	CNA	0.587	SRSF2	CNA	0.466
CD274	CNA	0.775	FOXP1	CNA	0.586	CCNE1	CNA	0.461
RPN1	CNA	0.769	ELK4	CNA	0.584	GNAS	NGS	0.455
CTNNA1	CNA	0.737	KRAS	CNA	0.573	H3F3A	CNA	0.455
POT1	NGS	0.731	MECOM	CNA	0.570	LHFPL6	CNA	0.451
PDCD1LG2	CNA	0.707	CREB3L2	CNA	0.563	IRF4	CNA	0.449
TP53	NGS	0.689	CBL	CNA	0.556	FH	CNA	0.446
GSK3B	CNA	0.662	FHIT	CNA	0.544	GMPS	CNA	0.443
CRKL	CNA	0.655	VTI1A	CNA	0.541	FLI1	CNA	0.441
Gender	META	0.624	WWTR1	CNA	0.533	TRRAP	CNA	0.440
			CTCF	CNA	0.518	APC	NGS	0.440

Table 59: Lung Small Cell Carcinoma NOS – Lung

GENE	TECH	IMP						
RB1	NGS	1.000	SRGAP3	CNA	0.701	BAP1	CNA	0.618
NKX2-1	CNA	0.924	ARID1A	CNA	0.699	KDSR	CNA	0.616
FOXL2	NGS	0.918	SS18	CNA	0.699	BCL9	CNA	0.612
SETBP1	CNA	0.892	RB1	CNA	0.693	MYCL	CNA	0.605
VHL	CNA	0.832	CBFB	CNA	0.691	SOX2	CNA	0.595
MSI2	CNA	0.829	PBRM1	CNA	0.688	HMGN2P46	CNA	0.588
TGFBR2	CNA	0.807	CDKN2C	CNA	0.685	HIST1H3B	CNA	0.576
MITF	CNA	0.797	FOXA1	CNA	0.672	LHFPL6	CNA	0.567
XPC	CNA	0.793	CDKN2B	CNA	0.665	KLHL6	CNA	0.560
FOXP1	CNA	0.778	BCL2	CNA	0.656	PPARG	CNA	0.550
CACNA1D	CNA	0.743	Age	META	0.652	FHIT	CNA	0.548
SMAD4	CNA	0.729	FLT3	CNA	0.640	FOXO1	CNA	0.535
			PBX1	CNA	0.625	DEK	CNA	0.532

TTL	CNA	0.527
Gender	META	0.518
FLT1	CNA	0.515
HIST1H4I	CNA	0.514

JAK1	CNA	0.509
FGFR2	CNA	0.509
MYD88	CNA	0.507
JUN	CNA	0.505

SFPQ	CNA	0.498
CDH11	CNA	0.498
DAXX	CNA	0.497
FANCD2	CNA	0.496

Table 60: Lung Squamous Carcinoma – Lung

GENE	TECH	IMP
Age	META	1.000
SOX2	CNA	0.971
FOXL2	NGS	0.917
CACNA1D	CNA	0.899
KLHL6	CNA	0.895
CTNNA1	CNA	0.865
XPC	CNA	0.826
CDKN2A	CNA	0.791
LPP	CNA	0.789
TP53	NGS	0.786
TFRC	CNA	0.783
CRKL	CNA	0.750
FHIT	CNA	0.748
CDKN2B	CNA	0.740
RPN1	CNA	0.739
FLT3	CNA	0.728

FGF10	CNA	0.717
BTG1	CNA	0.716
TERT	CNA	0.708
WWTR1	CNA	0.700
EWSR1	CNA	0.700
ETV5	CNA	0.698
MECOM	CNA	0.692
TGFBR2	CNA	0.691
Gender	META	0.685
PPARG	CNA	0.678
FLT1	CNA	0.677
CDX2	CNA	0.674
FOXP1	CNA	0.669
SPECC1	CNA	0.669
RAC1	CNA	0.664
LHFPL6	CNA	0.657
RAF1	CNA	0.655

SRGAP3	CNA	0.652
GNAS	CNA	0.649
MAF	CNA	0.645
CALR	CNA	0.645
BCL6	CNA	0.644
EBF1	CNA	0.644
IL7R	CNA	0.637
FGFR2	CNA	0.632
U2AF1	CNA	0.629
BCL11A	CNA	0.629
HMGN2P46	CNA	0.627
ERG	CNA	0.625
HMGA2	CNA	0.624
EP300	CNA	0.622
NF2	CNA	0.621
ACSL6	CNA	0.617
ELK4	CNA	0.617

Table 61: Meninges Meningioma NOS – Brain

GENE	TECH	IMP
CHEK2	CNA	1.000
MYCL	CNA	0.986
THRAP3	CNA	0.959
FOXL2	NGS	0.948
EWSR1	CNA	0.905
EBF1	CNA	0.863
TP53	NGS	0.857
MPL	CNA	0.823
PMS2	CNA	0.734
NF2	CNA	0.678
SPEN	CNA	0.661
Age	META	0.640
STIL	CNA	0.639
HLF	CNA	0.636
CDH11	CNA	0.628
FLI1	CNA	0.610

NTRK2	CNA	0.609
HOXA9	CNA	0.601
CDKN2C	CNA	0.601
RPL22	CNA	0.599
USP6	CNA	0.584
ZNF217	CNA	0.566
LHFPL6	CNA	0.553
EP300	CNA	0.550
Gender	META	0.538
NTRK3	CNA	0.538
HOXA13	CNA	0.537
RAC1	CNA	0.518
ERG	CNA	0.517
LCK	CNA	0.505
ECT2L	CNA	0.493
MTOR	CNA	0.484
SETBP1	CNA	0.483

MAP2K4	CNA	0.478
MYC	CNA	0.477
ELK4	CNA	0.473
CTNNA1	CNA	0.471
FANCF	CNA	0.466
SDHB	CNA	0.465
c-KIT	NGS	0.458
SPECC1	CNA	0.457
PDGFRB	CNA	0.455
GAS7	CNA	0.435
ZBTB16	CNA	0.435
U2AF1	CNA	0.433
RABEP1	CNA	0.427
FHIT	CNA	0.425
CSF3R	CNA	0.413
YWHAE	CNA	0.408
IGF1R	CNA	0.406

5

Table 62: Nasopharynx NOS Squamous Carcinoma – Head, Face or Neck, NOS

GENE	TECH	IMP
CTCF	CNA	1.000
FOXL2	NGS	0.955
TP53	NGS	0.870
SOX2	CNA	0.842
GNAS	CNA	0.838
CDH1	CNA	0.834
RPN1	CNA	0.833
Gender	META	0.828
KMT2A	CNA	0.770
ASXL1	CNA	0.739
MAP3K1	NGS	0.713
TGFBR2	CNA	0.703
SDHD	CNA	0.690
Age	META	0.690
CDKN2B	CNA	0.685
CBFB	CNA	0.680
PTPN11	CNA	0.673
ETV6	CNA	0.641
C15orf65	CNA	0.632
JAZF1	CNA	0.621
BCL6	CNA	0.612
TFRC	CNA	0.612
KDSR	CNA	0.598
MAML2	CNA	0.586
MLLT11	CNA	0.584
CBL	CNA	0.580
BUB1B	CNA	0.563
ABL2	NGS	0.553
EPHB1	CNA	0.550
APC	NGS	0.547
VHL	NGS	0.541
BTG1	CNA	0.540
PCM1	CNA	0.538
WIF1	CNA	0.537
TSC1	CNA	0.534
USP6	CNA	0.523
REL	CNA	0.509
CDK4	CNA	0.506
NUTM1	CNA	0.500
CYP2D6	CNA	0.496
CDX2	CNA	0.481
LHFPL6	CNA	0.478
SDHB	CNA	0.477
KRAS	NGS	0.460
RB1	NGS	0.453
PMS2	CNA	0.447
WRN	CNA	0.441
EGFR	CNA	0.441
CCDC6	CNA	0.432
MECOM	CNA	0.428

Table 63: Oligodendroglioma NOS – Brain

GENE	TECH	IMP
IDH1	NGS	1.000
Age	META	0.871
FOXL2	NGS	0.846
MPL	CNA	0.689
BCL3	CNA	0.651
FAM46C	CNA	0.640
ACSL6	CNA	0.624
RHOH	CNA	0.591
MLLT11	CNA	0.574
JAK1	CNA	0.564
ZNF331	CNA	0.560
OLIG2	CNA	0.560
ATP1A1	NGS	0.529
MCL1	CNA	0.498
Gender	META	0.486
KLK2	CNA	0.486
JUN	CNA	0.485
CD79A	CNA	0.463
MYCL	CNA	0.452
NUP93	CNA	0.450
PDE4DIP	CNA	0.432
RAD51	CNA	0.432
CTCF	CNA	0.399
TP53	NGS	0.396
PALB2	CNA	0.372
ERCC1	CNA	0.359
PPP2R1A	CNA	0.358
CSF3R	CNA	0.358
ZNF217	CNA	0.356
CBL	CNA	0.354
MYC	CNA	0.352
FLT1	CNA	0.352
SETBP1	CNA	0.351
SPECC1	CNA	0.351
ATP1A1	CNA	0.343
c-KIT	NGS	0.339
VHL	NGS	0.339
HIST1H4I	CNA	0.321
PAFAH1B2	CNA	0.320
MSI	NGS	0.320
EXT1	CNA	0.316
AXL	CNA	0.312
APC	NGS	0.309
NFKBIA	CNA	0.309
CACNA1D	CNA	0.306
RPL22	CNA	0.305
ELK4	CNA	0.304
MSI2	CNA	0.301
CCNE1	CNA	0.299
ARID1A	CNA	0.298

Table 64: Oligodendroglioma Anaplastic – Brain

GENE	TECH	IMP
IDH1	NGS	1.000
CCNE1	CNA	0.933
Age	META	0.917
FOXL2	NGS	0.916
ZNF703	CNA	0.844
JUN	CNA	0.763
SFPQ	CNA	0.752
RPL22	CNA	0.694
THRAP3	CNA	0.647
BCL3	CNA	0.619
ZNF331	CNA	0.610

SDHB	CNA	0.610
MPL	CNA	0.582
MCL1	CNA	0.564
ERCC1	CNA	0.555
CDH1	NGS	0.482
ERG	CNA	0.464
TNFRSF14	CNA	0.436
NF2	CNA	0.414
c-KIT	NGS	0.410
GRIN2A	CNA	0.409
RPL5	CNA	0.406
USP6	CNA	0.391
ZNF217	CNA	0.378

MUTYH	CNA	0.373
CDKN2C	CNA	0.373
AFF3	CNA	0.369
MYCL	CNA	0.366
NR4A3	CNA	0.359
ELK4	CNA	0.358
ACSL6	CNA	0.358
MUC1	CNA	0.354
APC	NGS	0.349
CSF3R	CNA	0.348
MLLT11	CNA	0.347
TET1	NGS	0.345
KRAS	NGS	0.341

SYK	CNA	0.334
CHEK2	CNA	0.332
EWSR1	CNA	0.325
PTEN	NGS	0.323
U2AF1	CNA	0.321
SETBP1	CNA	0.319
MDM4	NGS	0.318
SPECC1	CNA	0.316
ATP1A1	CNA	0.316
CBLC	CNA	0.312
ARID1A	CNA	0.307
SOX10	CNA	0.304
TP53	NGS	0.302

Table 65: Ovary Adenocarcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000
Gender	META	0.986
MECOM	CNA	0.875
KLHL6	CNA	0.834
APC	NGS	0.827
MYC	CNA	0.784
BCL6	CNA	0.761
TP53	NGS	0.760
KRAS	NGS	0.752
SPECC1	CNA	0.748
VHL	NGS	0.740
WWTR1	CNA	0.728
ZNF217	CNA	0.720
CBFB	CNA	0.703
MUC1	CNA	0.700
CDH1	CNA	0.691
c-KIT	NGS	0.680
CCNE1	CNA	0.678
KAT6B	CNA	0.671
GID4	CNA	0.665

CDH11	CNA	0.660
MLLT11	CNA	0.659
SUZ12	CNA	0.657
CDKN2B	CNA	0.652
CDKN2A	CNA	0.649
HMGN2P46	CNA	0.649
TPM4	CNA	0.644
RPN1	CNA	0.644
CDKN2C	CNA	0.644
WT1	CNA	0.642
SETBP1	CNA	0.640
BCL9	CNA	0.640
FANCC	CNA	0.637
EP300	CNA	0.633
NTRK2	CNA	0.633
LHFPL6	CNA	0.630
CACNA1D	CNA	0.625
ARID1A	CNA	0.625
CDX2	CNA	0.624
CTCF	CNA	0.624
RAC1	CNA	0.611

CNBP	CNA	0.607
NUP214	CNA	0.605
SOX2	CNA	0.604
GATA3	CNA	0.604
BCL2	CNA	0.603
ETV5	CNA	0.601
GNAS	CNA	0.600
PAX8	CNA	0.596
CDH1	NGS	0.595
C15orf65	CNA	0.595
ZNF331	CNA	0.594
CDKN1B	CNA	0.594
EWSR1	CNA	0.593
NDRG1	CNA	0.591
KDSR	CNA	0.584
EBF1	CNA	0.583
PMS2	CNA	0.582
MSI2	CNA	0.581
ASXL1	CNA	0.579

Table 66: Ovary Carcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000
Gender	META	0.996
MECOM	CNA	0.973
FOXL2	NGS	0.875
HMGN2P46	CNA	0.826

KLHL6	CNA	0.824
TP53	NGS	0.815
CDH11	CNA	0.797
RAC1	CNA	0.794
CDH1	CNA	0.788
RPN1	CNA	0.769

SUZ12	CNA	0.768
JAZF1	CNA	0.766
NF1	CNA	0.756
ETV5	CNA	0.754
CBFB	CNA	0.753
KRAS	NGS	0.753

ZNF217	CNA	0.748
ETV1	CNA	0.747
LHFPL6	CNA	0.732
MYC	CNA	0.731
MAF	CNA	0.731
ARID1A	CNA	0.716
TAF15	CNA	0.715
WWTR1	CNA	0.715
EP300	CNA	0.700
CARS	CNA	0.694
FGFR2	CNA	0.693
SPECC1	CNA	0.690
PMS2	CNA	0.689
TET2	CNA	0.681
C15orf65	CNA	0.673
FANCC	CNA	0.669
CDKN2A	CNA	0.668
CCNE1	CNA	0.664
NUP98	CNA	0.656
HOXD13	CNA	0.651
CACNA1D	CNA	0.650
NUP214	CNA	0.650
FANCF	CNA	0.648
CTCF	CNA	0.647
MUC1	CNA	0.646
EWSR1	CNA	0.645
CDKN2B	CNA	0.645
FOXA1	CNA	0.644
PDE4DIP	CNA	0.640
APC	NGS	0.639
MCL1	CNA	0.638
CDK12	CNA	0.630
CDX2	CNA	0.628
PRCC	CNA	0.627

Table 67: Ovary Carcinosarcoma – FGTP

GENE	TECH	IMP
ASXL1	CNA	1.000
STK11	CNA	0.951
FOXL2	NGS	0.945
MECOM	CNA	0.925
ZNF384	CNA	0.917
Gender	META	0.895
TP53	NGS	0.822
ETV5	CNA	0.815
GNAS	CNA	0.795
Age	META	0.783
WDCP	CNA	0.778
EP300	CNA	0.762
FGF6	CNA	0.715
FSTL3	CNA	0.708
EWSR1	CNA	0.691
PBX1	CNA	0.672
MYCN	CNA	0.666
AFF1	CNA	0.662
TRIM27	CNA	0.649
ALK	CNA	0.644
RAC1	CNA	0.642
BCL11A	CNA	0.640
CBFB	CNA	0.640
PRRX1	CNA	0.633
LHFPL6	CNA	0.630
CCND2	CNA	0.630
HMGA2	CNA	0.622
MAF	CNA	0.619
CDH1	CNA	0.606
TCF3	CNA	0.602
ETV6	CNA	0.600
NUTM1	CNA	0.592
DDR2	CNA	0.584
BCL2	NGS	0.571
PIK3CA	NGS	0.570
STAT3	CNA	0.568
CRKL	CNA	0.566
HMGN2P46	CNA	0.561
FGFR1	CNA	0.553
ERBB2	CNA	0.552
FGF23	CNA	0.550
ELK4	CNA	0.538
MAX	CNA	0.533
CCNE1	CNA	0.533
FANCF	CNA	0.532
PMS2	CNA	0.529
VEGFA	CNA	0.527
KLHL6	CNA	0.524
AURKA	CNA	0.522
NCOA1	CNA	0.516

Table 68: Ovary Clear Cell Carcinoma – FGTP

GENE	TECH	IMP
ZNF217	CNA	1.000
Age	META	0.965
FOXL2	NGS	0.935
ARID1A	NGS	0.920
TP53	NGS	0.887
PIK3CA	NGS	0.853
STAT3	CNA	0.826
Gender	META	0.810
HLF	CNA	0.755
EP300	CNA	0.743
MECOM	CNA	0.639
NF2	CNA	0.635
KAT6A	CNA	0.625
TRIM27	CNA	0.623
ERBB3	CNA	0.611
EXT1	CNA	0.610
ERCC5	CNA	0.608
NCOA2	CNA	0.597
FHIT	CNA	0.594
STAT5B	CNA	0.593
CDK12	CNA	0.592
CDKN2B	CNA	0.589
PAX8	CNA	0.588
FANCC	CNA	0.587
PLAG1	CNA	0.586
MED12	NGS	0.582
TSC1	CNA	0.581
CDKN2A	CNA	0.574
CCNE1	CNA	0.570
ACKR3	CNA	0.567
NR4A3	CNA	0.563
BCL2	CNA	0.560

WWTR1	CNA	0.558	FLI1	CNA	0.514	EWSR1	CNA	0.503
IRS2	CNA	0.553	NUTM1	CNA	0.510	SUFU	CNA	0.502
RAC1	CNA	0.537	BRCA1	CNA	0.509	PBX1	CNA	0.500
PDCD1LG2	CNA	0.531	BTG1	CNA	0.508	HMGN2P46	CNA	0.494
HSP90AB1	CNA	0.531	MSI2	CNA	0.508	CDH11	CNA	0.490
CBL	CNA	0.523	NUP214	CNA	0.503	APC	NGS	0.489

Table 69: Ovary Endometrioid Adenocarcinoma – FGTP

GENE	TECH	IMP						
Age	META	1.000	CDKN2A	CNA	0.604	CRKL	CNA	0.526
FOXL2	NGS	0.951	MDM4	CNA	0.596	FLI1	CNA	0.526
CTNNB1	NGS	0.936	ALK	CNA	0.594	NUP98	CNA	0.526
ARID1A	NGS	0.879	VTI1A	CNA	0.582	CBL	CNA	0.524
CHIC2	CNA	0.848	ZNF331	CNA	0.581	BCL6	CNA	0.524
FGFR2	CNA	0.834	CCDC6	CNA	0.578	PTEN	NGS	0.522
Gender	META	0.809	LHFPL6	CNA	0.575	MYCL	CNA	0.517
FANCF	CNA	0.791	BCL9	CNA	0.562	RAC1	CNA	0.517
MUC1	CNA	0.774	HMGN2P46	CNA	0.560	ARID1A	CNA	0.516
ELK4	CNA	0.675	CTNNA1	CNA	0.555	BCL11A	CNA	0.515
TP53	NGS	0.667	CDK12	CNA	0.547	TET1	CNA	0.509
PBX1	CNA	0.662	CACNA1D	CNA	0.541	FHIT	CNA	0.506
CBFB	CNA	0.656	ZNF384	CNA	0.540	CDKN1B	CNA	0.501
AFF3	CNA	0.655	HOXA13	CNA	0.535	STAT3	CNA	0.499
MAF	CNA	0.655	PPARG	CNA	0.534	CDKN2B	CNA	0.494
H3F3B	CNA	0.605	WWTR1	CNA	0.532	SETBP1	CNA	0.489
			PIK3CA	NGS	0.528	U2AF1	CNA	0.488

Table 70: Ovary Granulosa Cell Tumor – FGTP

GENE	TECH	IMP						
FOXL2	NGS	1.000	CYP2D6	CNA	0.319	NR4A3	CNA	0.248
EWSR1	CNA	0.475	CHEK2	CNA	0.317	CACNA1D	CNA	0.244
Gender	META	0.455	RMI2	CNA	0.317	MN1	CNA	0.242
NF2	CNA	0.454	GID4	CNA	0.312	BCR	CNA	0.241
MYH9	CNA	0.450	SOX2	CNA	0.306	ALDH2	CNA	0.237
TP53	NGS	0.425	CRKL	CNA	0.301	CEBPA	CNA	0.231
Age	META	0.422	HMGA2	CNA	0.290	IDH1	NGS	0.229
CBFB	CNA	0.408	PATZ1	CNA	0.281	TSC1	CNA	0.225
MKL1	CNA	0.388	SOX10	CNA	0.276	PTCH1	CNA	0.225
BCL3	CNA	0.377	ZNF217	CNA	0.276	APC	NGS	0.222
TSHR	CNA	0.368	EP300	CNA	0.274	KRAS	NGS	0.220
SPECC1	CNA	0.355	PTPN11	CNA	0.270	BLM	NGS	0.215
FHIT	CNA	0.346	ATF1	CNA	0.267	ERG	NGS	0.215
SMARCB1	CNA	0.346	PCM1	CNA	0.266	HLF	NGS	0.215
FANCC	CNA	0.331	IGF1R	CNA	0.266	NUP214	CNA	0.212
SOCS1	CNA	0.324	CCND2	CNA	0.261	PTEN	NGS	0.211
			FLT1	CNA	0.254	HOXA13	CNA	0.205

5

Table 71: Ovary High-grade Serous Carcinoma – FGTP

GENE	TECH	IMP	ETV1	CNA	0.615	ABL1	NGS	0.472
MECOM	CNA	1.000	ALDH2	NGS	0.607	AKT3	NGS	0.463
MLLT11	NGS	0.987	AURKB	NGS	0.606	Gender	META	0.459
KLHL6	CNA	0.984	ACSL3	NGS	0.589	HOXA9	CNA	0.448
ETV5	CNA	0.942	CBFB	NGS	0.589	RPN1	CNA	0.445
HIST1H4I	NGS	0.927	H3F3B	NGS	0.584	CBFB	CNA	0.434
BTG1	NGS	0.881	WWTR1	CNA	0.577	ATP1A1	NGS	0.433
EZR	CNA	0.791	ALK	NGS	0.554	RAP1GDS1	CNA	0.430
C15orf65	NGS	0.779	BRCA1	NGS	0.554	MAF	CNA	0.429
BCL2L11	NGS	0.776	AKT1	NGS	0.547	ASXL1	CNA	0.407
HMGN2P46	NGS	0.769	BCL6	CNA	0.536	GSK3B	CNA	0.402
AKT2	NGS	0.728	ACSL6	NGS	0.522	HEY1	CNA	0.390
ARFRP1	NGS	0.671	DDIT3	NGS	0.520	WRN	CNA	0.384
BAP1	NGS	0.658	ARHGAP26	NGS	0.502	FOXO1	CNA	0.376
BCL2	NGS	0.637	ABL2	NGS	0.500	SUZ12	CNA	0.372
ZNF384	CNA	0.635	NF1	CNA	0.486	GNA11	NGS	0.366
TAF15	CNA	0.615	TFRC	CNA	0.472	PIK3CA	CNA	0.366

Table 72: Ovary Low-grade Serous Carcinoma – FGTP

GENE	TECH	IMP	GNA11	NGS	0.544	SDHC	CNA	0.358
RPL22	CNA	1.000	H3F3A	CNA	0.484	HRAS	NGS	0.358
HMGN2P46	NGS	0.898	GID4	CNA	0.477	HMGN2P46	CNA	0.352
CDKN2A	CNA	0.780	ARFRP1	NGS	0.466	AURKB	NGS	0.350
CDKN2B	CNA	0.752	TNFRSF14	CNA	0.464	COX6C	CNA	0.343
WRN	CNA	0.712	DDIT3	NGS	0.456	ABL1	NGS	0.330
HOOK3	CNA	0.667	BCL2	NGS	0.451	ACKR3	NGS	0.329
PCM1	CNA	0.631	PSIP1	CNA	0.431	SBDS	CNA	0.325
BCL2L11	NGS	0.613	ALDH2	NGS	0.424	TCL1A	CNA	0.321
H3F3B	NGS	0.604	MCL1	CNA	0.423	CACNA1D	CNA	0.321
BTG1	NGS	0.598	AKT2	NGS	0.404	MLLT3	CNA	0.318
HIST1H4I	NGS	0.584	C15orf65	NGS	0.403	USP6	CNA	0.318
PLAG1	CNA	0.578	MLLT11	CNA	0.400	SDHB	CNA	0.312
NUTM2B	CNA	0.562	PRKDC	CNA	0.395	ABL2	NGS	0.312
SOX2	CNA	0.558	MAP2K1	CNA	0.389	ACSL6	NGS	0.310
WISP3	CNA	0.547	CDK4	NGS	0.387	AKT1	NGS	0.303
RUNX1T1	CNA	0.545	NRAS	NGS	0.362	RBM15	CNA	0.299

Table 73: Ovary Mucinous Adenocarcinoma – FGTP

GENE	TECH	IMP	Gender	META	0.784	CDKN2B	CNA	0.579
KRAS	NGS	1.000	CDKN2A	CNA	0.628	YWHAE	CNA	0.569
Age	META	0.941	HMGN2P46	CNA	0.620	TPM4	CNA	0.566
FOXL2	NGS	0.896	FUS	CNA	0.618	BCL6	CNA	0.565

LHFPL6	CNA	0.558	C15orf65	CNA	0.464	STAT3	CNA	0.424
SRGAP3	CNA	0.538	ASXL1	CNA	0.456	INHBA	CNA	0.418
ZNF217	CNA	0.534	APC	NGS	0.447	EBF1	CNA	0.418
c-KIT	NGS	0.524	NUTM1	CNA	0.447	RAC1	CNA	0.416
HEY1	CNA	0.523	BCL2	CNA	0.443	U2AF1	CNA	0.415
FNBP1	CNA	0.511	KLHL6	CNA	0.440	WT1	CNA	0.411
CDKN2C	CNA	0.506	MSI	NGS	0.438	CDX2	CNA	0.410
CTNNA1	CNA	0.502	NTRK2	CNA	0.436	CRKL	CNA	0.409
CACNA1D	CNA	0.495	RMI2	CNA	0.434	ERBB4	CNA	0.406
SETBP1	CNA	0.481	BRCA2	CNA	0.434	SDC4	CNA	0.404
SOX2	CNA	0.474	PDCD1LG2	CNA	0.432	SPECC1	CNA	0.401
KDM5C	NGS	0.471	FHIT	CNA	0.432	CDH1	CNA	0.394
MYC	CNA	0.470	PPARG	CNA	0.425	TP53	NGS	0.389

Table 74: Ovary Serous Carcinoma – FGTP

GENE	TECH	IMP						
WT1	CNA	1.000	FANCF	CNA	0.689	MLLT11	CNA	0.639
Gender	META	0.988	PAX8	CNA	0.686	HMGN2P46	CNA	0.634
Age	META	0.933	CDH1	CNA	0.685	NDRG1	CNA	0.634
EP300	CNA	0.821	PIK3CA	NGS	0.672	MYC	CNA	0.633
MECOM	CNA	0.819	CDKN1B	CNA	0.671	CTCF	CNA	0.632
APC	NGS	0.791	ARID1A	CNA	0.669	c-KIT	NGS	0.629
RPN1	CNA	0.778	RAC1	CNA	0.660	HOOK3	CNA	0.626
CBFB	CNA	0.773	TAF15	CNA	0.657	CDKN2A	CNA	0.625
TPM4	CNA	0.754	CDH11	CNA	0.653	SUZ12	CNA	0.616
TP53	NGS	0.748	JAZF1	CNA	0.650	ZNF384	CNA	0.616
KRAS	NGS	0.735	ETV1	CNA	0.649	CDKN2B	CNA	0.614
MUC1	CNA	0.729	FOXL2	NGS	0.646	SMARCE1	CNA	0.608
KLHL6	CNA	0.718	CRKL	CNA	0.645	BCL9	CNA	0.606
PMS2	CNA	0.712	ETV6	CNA	0.644	STAT3	CNA	0.602
MAF	CNA	0.709	CDX2	CNA	0.643	ZNF331	CNA	0.601
BCL6	CNA	0.698	CDK12	CNA	0.640	ETV5	CNA	0.596
			CCNE1	CNA	0.639	EWSR1	CNA	0.593

Table 75: Pancreas Adenocarcinoma NOS – Pancreas

GENE	TECH	IMP						
KRAS	NGS	1.000	USP6	CNA	0.588	FOXO1	CNA	0.546
APC	NGS	0.731	IRF4	CNA	0.584	BCL2	CNA	0.541
Age	META	0.706	TP53	NGS	0.584	SPEN	CNA	0.537
SETBP1	CNA	0.676	SPECC1	CNA	0.582	LHFPL6	CNA	0.536
CDKN2A	CNA	0.649	CACNA1D	CNA	0.577	HMGN2P46	CNA	0.536
FANCF	CNA	0.633	CBFB	CNA	0.567	YWHAE	CNA	0.524
CDKN2B	CNA	0.621	MDS2	CNA	0.561	ARID1A	CNA	0.513
ERG	CNA	0.610	Gender	META	0.561	CDX2	CNA	0.511
KDSR	CNA	0.594	SMAD4	CNA	0.559	RABEP1	CNA	0.509
			SMAD2	CNA	0.556	PDCD1LG2	CNA	0.508

CRTC3	CNA	0.507
MAF	CNA	0.504
WWTR1	CNA	0.502
VHL	NGS	0.502
CDH1	CNA	0.500
TGFBR2	CNA	0.497
EP300	CNA	0.493

SDHB	CNA	0.493
RAC1	CNA	0.493
FLI1	CNA	0.490
CDH11	CNA	0.482
EWSR1	CNA	0.481
MSI2	CNA	0.479
FHIT	CNA	0.478

HOXA9	CNA	0.477
EXT1	CNA	0.476
ELK4	CNA	0.475
CRKL	CNA	0.469
RPN1	CNA	0.468
ASXL1	CNA	0.468
PMS2	CNA	0.468

Table 76: Pancreas Carcinoma NOS – Pancreas

GENE	TECH	IMP
KRAS	NGS	1.000
FOXL2	NGS	0.850
CDKN2A	CNA	0.748
FHIT	CNA	0.724
CDKN2B	CNA	0.617
SETBP1	CNA	0.595
Gender	META	0.591
TP53	NGS	0.585
YWHAE	CNA	0.576
Age	META	0.576
PDE4DIP	CNA	0.553
RPL22	CNA	0.547
RMI2	CNA	0.530
CAMTA1	CNA	0.528
FSTL3	CNA	0.507
CREB3L2	CNA	0.499

FCRL4	CNA	0.483
RPN1	CNA	0.482
ACSL6	CNA	0.481
IRF4	CNA	0.475
TNFRSF17	CNA	0.472
ASXL1	CNA	0.471
CBFB	CNA	0.466
KLHL6	CNA	0.465
CTNNA1	CNA	0.461
FAM46C	CNA	0.456
EP300	CNA	0.454
BCL11A	CNA	0.454
ZNF521	CNA	0.452
USP6	CNA	0.452
IL6ST	CNA	0.450
FANCF	CNA	0.447
MAML2	CNA	0.444

PBX1	CNA	0.443
BTG1	CNA	0.440
ERG	CNA	0.440
EBF1	CNA	0.436
TFRC	CNA	0.435
CDH11	CNA	0.432
JAZF1	CNA	0.431
ZNF217	CNA	0.425
CTCF	CNA	0.424
MYC	CNA	0.424
GNAS	CNA	0.423
ESR1	CNA	0.421
NF2	CNA	0.418
CDH1	CNA	0.416
HEY1	CNA	0.409
CACNA1D	CNA	0.407
SOX2	CNA	0.404

Table 77: Pancreas Mucinous Adenocarcinoma – Pancreas

GENE	TECH	IMP
KRAS	NGS	1.000
APC	NGS	0.568
FOXL2	NGS	0.516
ASXL1	CNA	0.489
JUN	CNA	0.487
Gender	META	0.455
GNAS	NGS	0.442
FOXO1	CNA	0.436
NUTM1	CNA	0.429
STK11	NGS	0.425
ACKR3	NGS	0.406
CACNA1D	CNA	0.386
MUC1	CNA	0.382
SETBP1	CNA	0.379
ARID1A	CNA	0.373

STAT3	NGS	0.372
ZNF331	CNA	0.369
CDKN2A	CNA	0.369
TP53	NGS	0.367
RMI2	CNA	0.356
ERCC3	NGS	0.340
VHL	NGS	0.332
CDH1	NGS	0.332
NTRK2	CNA	0.327
CDKN2B	CNA	0.327
RAC1	CNA	0.314
HMGN2P46	CNA	0.311
ELK4	CNA	0.306
Age	META	0.305
FANCF	CNA	0.302
JAK1	CNA	0.281

FAM46C	CNA	0.277
C15orf65	CNA	0.273
AFF4	NGS	0.268
SDHB	CNA	0.264
MSI2	CNA	0.264
TAL2	CNA	0.257
RUNX1	CNA	0.247
SOCS1	CNA	0.242
COX6C	CNA	0.235
SMAD4	CNA	0.235
CREB3L2	CNA	0.234
RPN1	CNA	0.232
KDSR	CNA	0.229
EBF1	CNA	0.228
FANCC	CNA	0.226
FCRL4	CNA	0.224

USP6	CNA	0.224

EZR	CNA	0.222

CCDC6	CNA	0.222

Table 78: Pancreas Neuroendocrine Carcinoma – Pancreas

GENE	TECH	IMP
JAZF1	CNA	1.000
GATA3	CNA	0.992
FOXL2	NGS	0.973
WWTR1	CNA	0.962
Age	META	0.904
MECOM	CNA	0.874
FOXA1	CNA	0.856
EPHA3	CNA	0.825
MLLT3	CNA	0.774
BCL6	CNA	0.770
LHFPL6	CNA	0.769
PTPRC	CNA	0.764
CDK4	CNA	0.761
PTPN11	CNA	0.754
LPP	CNA	0.749
TFRC	CNA	0.730

ZNF217	CNA	0.722
BTG1	CNA	0.718
FCRL4	CNA	0.695
EBF1	CNA	0.678
NOTCH2	CNA	0.677
STAT5B	CNA	0.672
INHBA	CNA	0.665
TCL1A	CNA	0.657
KLHL6	CNA	0.646
SMAD4	CNA	0.635
MLF1	CNA	0.632
TP53	NGS	0.631
SETBP1	CNA	0.630
SOX2	CNA	0.610
TCEA1	CNA	0.609
GMPS	CNA	0.600
Gender	META	0.596

MYC	CNA	0.592
DICER1	CNA	0.589
NIN	CNA	0.576
CD79A	NGS	0.567
SPECC1	CNA	0.565
ITK	CNA	0.541
ETV1	CNA	0.530
KDSR	CNA	0.525
PMS2	CNA	0.522
CTCF	CNA	0.509
FGFR2	CNA	0.508
FLT1	CNA	0.508
DDIT3	CNA	0.507
NR4A3	CNA	0.507
IL7R	CNA	0.507
RUNX1	CNA	0.505
H3F3A	CNA	0.505

Table 79: Parotid Gland Carcinoma NOS – Head, Face or Neck, NOS

GENE	TECH	IMP
ERBB2	CNA	1.000
FOXL2	NGS	0.974
CACNA1D	CNA	0.864
CRTC3	CNA	0.829
RMI2	CNA	0.801
TRRAP	CNA	0.793
RUNX1	CNA	0.782
LRP1B	NGS	0.764
RPL22	CNA	0.754
Gender	META	0.749
SBDS	CNA	0.719
NDRG1	NGS	0.715
CBFB	CNA	0.701
GATA3	CNA	0.696
NSD3	CNA	0.695
APC	NGS	0.693

Age	META	0.690
PTEN	NGS	0.686
CDKN2A	CNA	0.676
VEGFA	CNA	0.673
LHFPL6	CNA	0.671
IGF1R	CNA	0.658
TFRC	CNA	0.638
SMAD2	CNA	0.632
HOXD13	CNA	0.621
CDH11	CNA	0.614
CDH1	NGS	0.609
HEY1	CNA	0.591
ACKR3	CNA	0.580
SOX2	CNA	0.565
c-KIT	NGS	0.560
HMGA2	CNA	0.535
IL7R	NGS	0.535

CREBBP	CNA	0.530
FUS	CNA	0.526
MDM2	CNA	0.509
GNA13	CNA	0.507
GNAS	CNA	0.505
NTRK3	CNA	0.504
TP53	NGS	0.504
CYLD	CNA	0.496
ASXL1	CNA	0.494
GRIN2A	CNA	0.494
CDK6	CNA	0.480
ELK4	CNA	0.479
VTI1A	CNA	0.474
PRDM1	CNA	0.473
ZRSR2	NGS	0.460
BCL11A	CNA	0.456
JAZF1	CNA	0.456

5

Table 80: Peritoneum Adenocarcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000

Gender	META	0.948
FOXL2	NGS	0.921

EWSR1	CNA	0.869
ETV5	CNA	0.830

EPHA3	CNA	0.828
GMPS	CNA	0.826
SYK	CNA	0.821
CCNE1	CNA	0.799
TP53	NGS	0.768
FANCC	CNA	0.767
CDH1	CNA	0.742
MECOM	CNA	0.741
LPP	CNA	0.734
FGFR2	CNA	0.734
FNBP1	CNA	0.679
TFRC	CNA	0.677
MAF	CNA	0.676
NTRK2	CNA	0.675
RPN1	CNA	0.653

SETBP1	CNA	0.648
ZNF384	CNA	0.635
SOX2	CNA	0.632
LHFPL6	CNA	0.628
JAZF1	CNA	0.626
RAC1	CNA	0.618
NUP214	CNA	0.615
PRCC	CNA	0.615
CALR	CNA	0.612
CHEK2	CNA	0.602
KLHL6	CNA	0.586
PTCH1	CNA	0.582
WT1	CNA	0.582
ERCC4	CNA	0.577
CDKN2A	CNA	0.571

TRIM27	CNA	0.564
MAML2	CNA	0.556
MLLT11	CNA	0.555
TPM4	CNA	0.551
TAF15	CNA	0.550
CCND1	CNA	0.548
NSD1	CNA	0.548
RNF213	NGS	0.545
BCL9	CNA	0.540
MYC	CNA	0.537
WWTR1	CNA	0.535
MED12	NGS	0.535
CAMTA1	CNA	0.531
BCL6	CNA	0.531
FHIT	CNA	0.526

Table 81: Peritoneum Carcinoma NOS – FGTP

GENE	TECH	IMP
Age	META	1.000
FOXL2	NGS	0.940
Gender	META	0.875
TP53	NGS	0.777
KAT6B	CNA	0.772
WWTR1	CNA	0.757
CDK12	CNA	0.732
RPN1	CNA	0.687
MLF1	CNA	0.681
TFRC	CNA	0.679
RAC1	CNA	0.679
XPC	CNA	0.675
NTRK2	CNA	0.669
NF1	CNA	0.662
EWSR1	CNA	0.660
EXT1	CNA	0.647

WRN	CNA	0.631
CDK6	CNA	0.628
CDH11	CNA	0.624
VHL	CNA	0.604
LPP	CNA	0.597
SRGAP3	CNA	0.592
GMPS	CNA	0.589
MLLT3	CNA	0.579
CDH1	CNA	0.571
NUTM2B	CNA	0.570
EP300	CNA	0.558
INHBA	CNA	0.557
MECOM	CNA	0.550
CTCF	CNA	0.549
SUZ12	CNA	0.548
HOXA9	CNA	0.545
ETV5	CNA	0.545

APC	NGS	0.537
STAT5B	CNA	0.534
ETV1	CNA	0.530
KRAS	NGS	0.522
TPM4	CNA	0.522
CHEK2	CNA	0.521
BCL6	CNA	0.521
HMGN2P46	CNA	0.519
PAFAH1B2	CNA	0.505
CRTC3	CNA	0.505
LHFPL6	CNA	0.500
SOX2	CNA	0.497
FGFR2	CNA	0.496
MAML2	CNA	0.494
PAX5	CNA	0.493
KDSR	CNA	0.483
NDRG1	CNA	0.479

Table 82: Peritoneum Serous Carcinoma – FGTP

GENE	TECH	IMP
TPM4	CNA	1.000
BCL6	CNA	0.984
FOXL2	NGS	0.978
SUZ12	CNA	0.978
Gender	META	0.973
Age	META	0.955
CTCF	CNA	0.940

TP53	NGS	0.933
TAF15	CNA	0.902
RAC1	CNA	0.877
CDK12	CNA	0.875
EP300	CNA	0.866
CDKN2B	CNA	0.865
MECOM	CNA	0.865
RPN1	CNA	0.863

PMS2	CNA	0.853
WWTR1	CNA	0.845
ETV1	CNA	0.838
CDH1	CNA	0.822
LPP	CNA	0.807
ASXL1	CNA	0.794
CDH11	CNA	0.793
KLHL6	CNA	0.793

FANCA	CNA	0.786	EWSR1	CNA	0.726	EBF1	CNA	0.681
CBFB	CNA	0.786	TAL2	CNA	0.716	TFRC	CNA	0.677
FANCF	CNA	0.784	CDKN2A	CNA	0.713	SMARCE1	CNA	0.676
ETV5	CNA	0.778	GMPS	CNA	0.711	CCNE1	CNA	0.671
NUP93	CNA	0.766	NF1	CNA	0.710	WT1	CNA	0.668
FGFR2	CNA	0.760	NUP214	CNA	0.706	ZNF217	CNA	0.666
JAZF1	CNA	0.753	CRKL	CNA	0.702	MLF1	CNA	0.665
FHIT	CNA	0.740	SPECC1	CNA	0.700	ETV6	CNA	0.664
CYP2D6	CNA	0.738	KLF4	CNA	0.700	BCL9	CNA	0.664

Table 83: Pleural Mesothelioma NOS – Lung

GENE	TECH	IMP						
Age	META	1.000	ASXL1	CNA	0.684	PBRM1	CNA	0.488
FOXL2	NGS	0.954	FOXP1	CNA	0.658	CDX2	CNA	0.487
EWSR1	CNA	0.938	RAC1	CNA	0.630	CALR	CNA	0.484
CDKN2B	CNA	0.909	FSTL3	CNA	0.619	BAP1	CNA	0.484
TP53	NGS	0.849	ARID1A	CNA	0.602	ITK	CNA	0.484
EPHA3	CNA	0.848	NUTM2B	CNA	0.550	CDH1	CNA	0.483
CDKN2A	CNA	0.834	LYL1	CNA	0.543	CDH11	CNA	0.482
Gender	META	0.834	EGFR	CNA	0.528	KRAS	NGS	0.479
WT1	CNA	0.825	CDKN2C	CNA	0.526	c-KIT	NGS	0.477
MAF	CNA	0.822	HMGN2P46	CNA	0.520	NFIB	CNA	0.473
EBF1	CNA	0.778	WISP3	CNA	0.516	MAP2K1	CNA	0.471
NF2	CNA	0.754	KDR	CNA	0.513	C15orf65	CNA	0.468
PRDM1	CNA	0.714	NTRK3	CNA	0.504	VHL	NGS	0.465
MSI2	CNA	0.712	RUNX1T1	CNA	0.502	FGF10	CNA	0.461
ACSL6	CNA	0.707	FGFR2	CNA	0.500	HLF	CNA	0.460
EP300	CNA	0.698	TPM4	CNA	0.497	ERG	CNA	0.454
			FAM46C	CNA	0.491	CREB3L2	CNA	0.452

Table 84: Prostate Adenocarcinoma NOS – Prostate

GENE	TECH	IMP						
Gender	META	1.000	GNA11	NGS	0.562	ELK4	CNA	0.430
FOXA1	CNA	0.875	NCOA2	CNA	0.537	SDC4	CNA	0.430
PTEN	CNA	0.825	LCP1	CNA	0.531	MAF	CNA	0.411
KRAS	NGS	0.783	PTCH1	CNA	0.530	FGF14	CNA	0.404
Age	META	0.697	c-KIT	NGS	0.510	RB1	CNA	0.403
KLK2	CNA	0.693	TP53	NGS	0.500	CACNA1D	CNA	0.401
FOXO1	CNA	0.675	CDKN1B	CNA	0.491	CDKN2B	CNA	0.394
FANCA	CNA	0.664	HOXA11	CNA	0.466	HEY1	CNA	0.388
GATA2	CNA	0.663	FGFR2	CNA	0.457	TP53	CNA	0.384
APC	NGS	0.623	IDH1	NGS	0.456	COX6C	CNA	0.381
LHFPL6	CNA	0.608	IRF4	CNA	0.454	CDX2	CNA	0.377
ETV6	CNA	0.580	PCM1	CNA	0.452	SOX10	CNA	0.376
ERCC3	CNA	0.579	CDKN2A	CNA	0.442	BRAF	NGS	0.374
			VHL	NGS	0.431	SRGAP3	CNA	0.373

FGFR1	CNA	0.371
CDH11	CNA	0.370
SPECC1	CNA	0.368
CREBBP	CNA	0.366
TGFBR2	CNA	0.366
CBFB	CNA	0.365
MLH1	CNA	0.364
PRDM1	CNA	0.363
HOXA13	CNA	0.355

Table 85: Rectosigmoid Adenocarcinoma NOS – Colon

GENE	TECH	IMP
APC	NGS	1.000
CDX2	CNA	0.877
FOXL2	NGS	0.771
FLT3	CNA	0.769
BCL2	CNA	0.750
FLT1	CNA	0.705
SETBP1	CNA	0.704
ZNF521	CNA	0.657
CDK8	CNA	0.645
KDSR	CNA	0.638
LHFPL6	CNA	0.628
ASXL1	CNA	0.603
SMAD4	CNA	0.584
RB1	CNA	0.578
MALT1	CNA	0.568
HOXA9	CNA	0.563
Age	META	0.561
RAC1	CNA	0.550
TOP1	CNA	0.540
CDKN2A	CNA	0.532
FOXO1	CNA	0.523
KRAS	NGS	0.521
ZMYM2	CNA	0.518
SDC4	CNA	0.515
ZNF217	CNA	0.510
CDKN2B	CNA	0.500
BRCA2	CNA	0.492
HOXA11	CNA	0.491
Gender	META	0.488
PMS2	CNA	0.477
FCRL4	CNA	0.475
WWTR1	CNA	0.471
BCL2	NGS	0.454
SS18	CNA	0.449
CAMTA1	CNA	0.440
BRAF	NGS	0.437
NSD3	CNA	0.437
MTOR	CNA	0.432
CTCF	CNA	0.420
SOX2	CNA	0.419
VHL	NGS	0.418
PRRX1	CNA	0.412
GNAS	CNA	0.405
PIK3CA	NGS	0.404
FANCF	CNA	0.398
MECOM	CNA	0.397
LCP1	CNA	0.397
HOXA13	CNA	0.396
CARS	CNA	0.396
ERCC5	CNA	0.393

Table 86: Rectum Adenocarcinoma NOS – Colon

GENE	TECH	IMP
APC	NGS	1.000
CDX2	CNA	0.904
SETBP1	CNA	0.745
KRAS	NGS	0.738
ASXL1	CNA	0.701
FLT3	CNA	0.698
Age	META	0.669
SDC4	CNA	0.663
KDSR	CNA	0.649
FLT1	CNA	0.649
ZNF217	CNA	0.631
CDK8	CNA	0.614
BCL2	CNA	0.601
LHFPL6	CNA	0.583
Gender	META	0.545
ZNF521	CNA	0.536
TP53	NGS	0.521
SPECC1	CNA	0.519
SMAD4	CNA	0.514
AMER1	NGS	0.503
FOXL2	NGS	0.503
ERCC5	CNA	0.499
GNAS	CNA	0.498
CDKN2B	CNA	0.493
RB1	CNA	0.481
HOXA9	CNA	0.458
VHL	NGS	0.456
HOXA11	CNA	0.455
TOP1	CNA	0.449
MALT1	CNA	0.443
EBF1	CNA	0.442
RAC1	CNA	0.441
BCL9	CNA	0.441
PTCH1	CNA	0.438
FOXO1	CNA	0.435
SS18	CNA	0.427
WWTR1	CNA	0.424
CCNE1	CNA	0.424
USP6	CNA	0.423
JAZF1	CNA	0.422
CAMTA1	CNA	0.421
CDKN2A	CNA	0.417
EXT1	CNA	0.417
ERG	CNA	0.416
CDH1	CNA	0.415
FNBP1	CNA	0.413
BRCA2	CNA	0.413
NSD2	CNA	0.412
HMGN2P46	CNA	0.406
ABL1	CNA	0.403

5

Table 87: Rectum Mucinous Adenocarcinoma – Colon

GENE	TECH	IMP
KRAS	NGS	1.000
APC	NGS	0.917
FOXL2	NGS	0.887
CDKN2A	CNA	0.665
CDKN2B	CNA	0.643
NUP214	CNA	0.641
GPHN	CNA	0.625
TSC1	CNA	0.605
KLF4	CNA	0.554
CDH1	NGS	0.550
PRKDC	CNA	0.542
Gender	META	0.538
ASPSCR1	NGS	0.521
Age	META	0.519
CDX2	CNA	0.512
BCL2	CNA	0.503
SDC4	CNA	0.498
RPL22	CNA	0.471
SOX2	CNA	0.469
PPARG	CNA	0.466
CTCF	CNA	0.456
LHFPL6	CNA	0.456
ARFRP1	CNA	0.449
TAL2	CNA	0.441
SETBP1	CNA	0.441
SYK	CNA	0.440
CACNA1D	CNA	0.415
LIFR	CNA	0.413
NTRK2	CNA	0.411
TP53	NGS	0.403
IRS2	CNA	0.403
KDSR	CNA	0.400
FHIT	CNA	0.397
PDGFRA	CNA	0.395
EPHA3	CNA	0.394
VTI1A	CNA	0.394
RMI2	CNA	0.394
NDRG1	CNA	0.394
USP6	CNA	0.393
WWTR1	CNA	0.389
EXT1	CNA	0.384
PMS2	CNA	0.380
RAF1	CNA	0.369
TGFBR2	CNA	0.363
SMAD4	NGS	0.360
ARID1A	CNA	0.359
JAK2	CNA	0.355
CCND2	CNA	0.352
HOXD13	CNA	0.352
TRIM27	CNA	0.350

Table 88: Retroperitoneum Dedifferentiated Liposarcoma – FGTP

GENE	TECH	IMP
CDK4	CNA	1.000
MDM2	CNA	0.760
RET	CNA	0.379
SBDS	CNA	0.334
ASXL1	CNA	0.245
VTI1A	CNA	0.216
KMT2D	CNA	0.212
GRIN2A	CNA	0.178
HMGA2	CNA	0.173
PTCH1	CNA	0.156
CYP2D6	CNA	0.156
BMPR1A	CNA	0.145
CDX2	CNA	0.137
GID4	CNA	0.134
ETV1	CNA	0.134
GATA2	CNA	0.128
USP6	CNA	0.120
MUC1	CNA	0.116
STAT5B	NGS	0.114
BCL9	CNA	0.112
PAX3	CNA	0.112
TP53	NGS	0.107
FGF4	CNA	0.106
SOX2	CNA	0.091
RABEP1	CNA	0.090
PTEN	CNA	0.090
FUBP1	NGS	0.089
RAD51	CNA	0.089
MLLT11	CNA	0.089
ACKR3	NGS	0.089
ZNF217	CNA	0.089
NF2	CNA	0.087
Age	META	0.082
KAT6B	CNA	0.079
ZNF521	CNA	0.079
IL2	CNA	0.079
KDM5C	NGS	0.079
IRS2	CNA	0.078
BCL6	CNA	0.077
ELK4	CNA	0.076
MNX1	CNA	0.070
WRN	CNA	0.068
CDK6	CNA	0.068
AFDN	CNA	0.068
POU2AF1	CNA	0.068
ESR1	NGS	0.067
ELN	CNA	0.067
NTRK2	CNA	0.067
NUMA1	CNA	0.067
SRC	CNA	0.067

Table 89: Retroperitoneum Leiomyosarcoma NOS – FGTP

GENE	TECH	IMP
GID4	CNA	1.000
FOXL2	NGS	0.916
NFKB2	CNA	0.905
SUFU	CNA	0.874
TGFBR2	CNA	0.870
SPECC1	CNA	0.817
TET1	CNA	0.786
TCF7L2	CNA	0.763
PDGFRA	CNA	0.727
MSH2	CNA	0.696
FGFR2	CNA	0.670
BCL11A	CNA	0.662
JUN	CNA	0.659
RET	CNA	0.620
MAP2K4	CNA	0.614
CHIC2	CNA	0.586
ALK	CNA	0.585

NT5C2	CNA	0.578	CRKL	CNA	0.466	BAP1	CNA	0.365
ATIC	CNA	0.572	VHL	NGS	0.458	NCOA4	CNA	0.356
EBF1	CNA	0.535	LHFPL6	CNA	0.457	CDH1	NGS	0.354
PRF1	CNA	0.521	WDCP	CNA	0.438	TP53	NGS	0.351
KAT6B	CNA	0.506	LCP1	CNA	0.422	EML4	CNA	0.345
TP53	CNA	0.502	CCDC6	CNA	0.416	KIAA1549	CNA	0.337
FHIT	CNA	0.500	IL2	CNA	0.414	KRAS	NGS	0.336
EP300	CNA	0.491	FUBP1	CNA	0.406	RB1	CNA	0.335
Gender	META	0.480	NTRK3	CNA	0.384	GNA11	CNA	0.328
JAK1	CNA	0.478	CRTC3	CNA	0.382	FLCN	CNA	0.326
MLH1	CNA	0.471	CDX2	CNA	0.368	CACNA1D	CNA	0.323

Table 90: Right Colon Adenocarcinoma NOS – Colon

GENE	TECH	IMP						
CDX2	CNA	1.000	EBF1	CNA	0.626	ERCC5	CNA	0.513
APC	NGS	0.952	MYC	CNA	0.619	SDC4	CNA	0.512
FLT3	CNA	0.842	HOXA11	CNA	0.584	BRCA2	CNA	0.509
FOXL2	NGS	0.827	ASXL1	CNA	0.583	USP6	CNA	0.506
KRAS	NGS	0.823	U2AF1	CNA	0.577	RB1	CNA	0.503
FLT1	CNA	0.798	Gender	META	0.574	CTCF	CNA	0.503
BRAF	NGS	0.784	CDKN2A	CNA	0.570	PDGFRA	CNA	0.503
RNF43	NGS	0.770	CDK8	CNA	0.565	RAC1	CNA	0.502
LHFPL6	CNA	0.759	WWTR1	CNA	0.563	FOXO1	CNA	0.498
SETBP1	CNA	0.748	SPECC1	CNA	0.560	TRIM27	CNA	0.495
HOXA9	CNA	0.705	CDH1	CNA	0.551	ZNF217	CNA	0.495
Age	META	0.703	ZNF521	CNA	0.551	CACNA1D	CNA	0.490
GID4	CNA	0.659	ETV5	CNA	0.548	ERG	CNA	0.488
SOX2	CNA	0.634	LCP1	CNA	0.533	FGF14	CNA	0.482
CDKN2B	CNA	0.631	ZMYM2	CNA	0.526	PMS2	CNA	0.481
BCL2	CNA	0.629	KDSR	CNA	0.526	SLC34A2	CNA	0.479
			SMAD4	CNA	0.522	LIFR	CNA	0.477

Table 91: Right Colon Mucinous Adenocarcinoma – Colon

GENE	TECH	IMP						
KRAS	NGS	1.000	WWTR1	CNA	0.634	CBFB	CNA	0.520
CDX2	CNA	0.891	HMGN2P46	CNA	0.610	PDGFRA	CNA	0.513
FOXL2	NGS	0.876	Gender	META	0.606	GNA13	CNA	0.506
APC	NGS	0.864	PRRX1	CNA	0.591	TCF7L2	CNA	0.499
Age	META	0.864	RPL22	NGS	0.591	FOXL2	CNA	0.494
RNF43	NGS	0.793	MYC	CNA	0.575	FLT1	CNA	0.492
LHFPL6	CNA	0.730	BRAF	NGS	0.568	SETBP1	CNA	0.487
CDK6	CNA	0.685	HOXA9	CNA	0.564	KLF4	CNA	0.484
RPN1	CNA	0.678	ASXL1	CNA	0.553	ETV5	CNA	0.481
PTCH1	CNA	0.670	FLT3	CNA	0.543	SOX2	CNA	0.481
CDKN2A	CNA	0.668	CDKN2B	CNA	0.543	ELK4	CNA	0.479
			GPHN	CNA	0.537	EBF1	CNA	0.479

SPEN	CNA	0.478	BRAF	CNA	0.467	MAML2	CNA	0.448
HOXA13	CNA	0.477	MSI2	CNA	0.466	PDCD1LG2	CNA	0.447
RPL22	CNA	0.472	EZH2	CNA	0.457	RUNX1T1	CNA	0.446
KIAA1549	CNA	0.469	RMI2	CNA	0.453	TCEA1	CNA	0.445
KMT2C	CNA	0.468	CDH1	CNA	0.453	GATA2	CNA	0.443

Table 92: Salivary Gland Adenoid Cystic Carcinoma – Head, Face or Neck, NOS

GENE	TECH	IMP						
SOX10	CNA	1.000	MDS2	CNA	0.553	TRRAP	CNA	0.451
TP53	NGS	0.825	ERBB3	CNA	0.548	TGFBR2	CNA	0.446
BCL2	CNA	0.791	BTG1	CNA	0.532	PDGFRA	NGS	0.441
Age	META	0.771	RUNX1	CNA	0.531	WDCP	CNA	0.435
ATF1	CNA	0.742	PMS2	CNA	0.531	TLX1	CNA	0.427
FOXL2	NGS	0.736	CEBPA	CNA	0.527	CDH11	CNA	0.421
IDH1	NGS	0.684	HOXC11	CNA	0.519	ABL1	NGS	0.412
c-KIT	NGS	0.677	DDIT3	CNA	0.515	FNBP1	CNA	0.412
APC	NGS	0.669	PTEN	NGS	0.512	NCOA1	NGS	0.412
CDK4	CNA	0.653	ASXL1	CNA	0.510	MAF	CNA	0.409
FANCF	CNA	0.624	MYH9	CNA	0.502	BCL6	CNA	0.405
FANCC	CNA	0.605	RPN1	CNA	0.501	BCL11A	CNA	0.405
Gender	META	0.603	PDCD1LG2	CNA	0.498	SDC4	CNA	0.404
KRAS	NGS	0.591	IRF4	CNA	0.474	FGFR2	CNA	0.404
VHL	NGS	0.579	LHFPL6	CNA	0.471	SETBP1	CNA	0.403
KMT2D	CNA	0.554	PAX3	CNA	0.452	HEY1	CNA	0.403
			CDH1	NGS	0.452	IKZF1	CNA	0.400

Table 93: Skin Merkel Cell Carcinoma – Skin

GENE	TECH	IMP						
Age	META	1.000	SMAD2	CNA	0.495	SDC4	CNA	0.355
RB1	NGS	0.980	KRAS	NGS	0.493	HOOK3	CNA	0.353
AKT1	NGS	0.902	FOXO1	CNA	0.468	SDHB	CNA	0.352
SFPQ	CNA	0.881	MAX	CNA	0.462	VHL	CNA	0.346
FOXL2	NGS	0.874	MDS2	CNA	0.452	PBX1	CNA	0.344
WWTR1	CNA	0.843	ECT2L	CNA	0.452	GOPC	NGS	0.344
TGFBR2	CNA	0.799	PRKDC	CNA	0.439	MYCL	CNA	0.335
Gender	META	0.795	CBFB	CNA	0.438	LCP1	CNA	0.332
JAK1	CNA	0.719	STAT5B	CNA	0.423	RB1	CNA	0.327
WISP3	CNA	0.716	HMGA2	CNA	0.419	PTCH1	CNA	0.323
SETBP1	CNA	0.694	MYC	CNA	0.413	ELL	NGS	0.318
CHIC2	CNA	0.632	RAC1	CNA	0.401	SRSF3	CNA	0.317
AFDN	CNA	0.615	MSI2	CNA	0.399	TP53	NGS	0.315
VHL	NGS	0.592	ZNF217	CNA	0.388	LMO1	CNA	0.311
CDKN2C	CNA	0.518	HLF	CNA	0.379	ERBB3	CNA	0.308
HSP90AB1	CNA	0.507	CALR	CNA	0.362	ARID1A	CNA	0.307
			CAMTA1	CNA	0.361	SPEN	CNA	0.304

5

Table 94: Skin Nodular Melanoma – Skin

GENE	TECH	IMP
CDKN2A	CNA	1.000
EZR	CNA	0.956
FOXL2	NGS	0.946
DAXX	CNA	0.833
BRAF	NGS	0.792
ABL1	NGS	0.752
CREB3L2	CNA	0.729
TP53	NGS	0.725
KIAA1549	CNA	0.722
CD274	CNA	0.710
NRAS	NGS	0.697
CDH1	NGS	0.679
c-KIT	NGS	0.655
FOXO3	CNA	0.634
EBF1	CNA	0.624
TRIM27	CNA	0.624
PDCD1LG2	CNA	0.614
CDKN2B	CNA	0.609
NFIB	CNA	0.603
ZNF217	CNA	0.598
SDHAF2	CNA	0.574
SOX10	CNA	0.573
POT1	CNA	0.544
Gender	META	0.513
SOX2	CNA	0.497
MLLT10	CNA	0.489
BRAF	CNA	0.488
IRF4	CNA	0.482
FOXL2	CNA	0.478
FANCG	CNA	0.478
FNBP1	CNA	0.472
FGFR2	CNA	0.468
CCDC6	CNA	0.466
ESR1	CNA	0.459
HIST1H4I	CNA	0.457
ABL1	CNA	0.456
TNFAIP3	CNA	0.449
Age	META	0.447
NUP214	CNA	0.421
MTOR	CNA	0.421
GMPS	CNA	0.418
CACNA1D	CNA	0.403
BTG1	CNA	0.402
SMAD2	CNA	0.400
KRAS	NGS	0.397
MLLT11	CNA	0.395
CARS	CNA	0.391
TCF7L2	CNA	0.389
PRDM1	CNA	0.386
HSP90AA1	CNA	0.384

Table 95: Skin Squamous Carcinoma – Skin

GENE	TECH	IMP
Age	META	1.000
NOTCH1	NGS	0.943
LRP1B	NGS	0.884
FOXL2	NGS	0.873
Gender	META	0.765
CACNA1D	CNA	0.744
EWSR1	CNA	0.726
ARFRP1	NGS	0.698
DDIT3	CNA	0.687
TP53	NGS	0.672
FNBP1	CNA	0.668
CDK4	CNA	0.647
KMT2D	NGS	0.646
MLH1	CNA	0.636
NTRK2	CNA	0.627
KLHL6	CNA	0.626
ARID1A	CNA	0.576
CHEK2	CNA	0.574
TAL2	CNA	0.554
FHIT	CNA	0.547
CAMTA1	CNA	0.536
SPECC1	CNA	0.536
FOXP1	CNA	0.532
PPARG	CNA	0.530
ASXL1	NGS	0.528
ABL1	CNA	0.518
SDHD	CNA	0.514
VHL	NGS	0.511
CCNE1	CNA	0.511
HOXD13	CNA	0.508
RAF1	CNA	0.507
KRAS	NGS	0.505
NUP214	CNA	0.500
NR4A3	CNA	0.499
JAZF1	CNA	0.495
RABEP1	CNA	0.491
GNAS	CNA	0.490
NOTCH2	NGS	0.487
FANCC	CNA	0.486
CDH11	CNA	0.485
SPEN	CNA	0.484
GPHN	CNA	0.483
ATR	NGS	0.483
TGFBR2	CNA	0.481
SETD2	CNA	0.474
HMGN2P46	CNA	0.471
GRIN2A	NGS	0.467
ZNF217	CNA	0.459
XPC	CNA	0.457
SDHB	CNA	0.455

Table 96: Skin Melanoma – Skin

GENE	TECH	IMP
IRF4	CNA	1.000
SOX10	CNA	0.977
FGFR2	CNA	0.807
FOXL2	NGS	0.799
EP300	CNA	0.785
BRAF	NGS	0.772
TP53	NGS	0.744
LRP1B	NGS	0.738
CCDC6	CNA	0.731
MITF	CNA	0.675
CREB3L2	CNA	0.645

Age	META	0.636
TRIM27	CNA	0.632
Gender	META	0.624
PDCD1LG2	CNA	0.620
CDKN2A	CNA	0.615
NRAS	NGS	0.609
TCF7L2	CNA	0.597
MTOR	CNA	0.594
NF2	CNA	0.590
CDKN2B	CNA	0.575
ESR1	CNA	0.562
GATA3	CNA	0.560
FOXA1	CNA	0.547

GRIN2A	NGS	0.542
NF1	NGS	0.536
CCND2	CNA	0.534
PRDM1	CNA	0.531
KRAS	NGS	0.528
EZR	CNA	0.525
MECOM	CNA	0.502
PAX3	CNA	0.497
NFIB	CNA	0.497
CNBP	CNA	0.494
CAMTA1	CNA	0.486
TNFAIP3	CNA	0.485
KIF5B	CNA	0.483

SOX2	CNA	0.482
LHFPL6	CNA	0.478
CHEK2	CNA	0.478
MLLT3	CNA	0.477
VTI1A	CNA	0.472
CTNNA1	CNA	0.471
KIAA1549	CNA	0.471
ARID1A	CNA	0.466
CDX2	CNA	0.459
DEK	CNA	0.458
CD274	CNA	0.453
CRKL	CNA	0.453
BTG1	CNA	0.453

Table 97: Small Intestine Gastrointestinal Stromal Tumor NOS – Small Intestine

GENE	TECH	IMP
c-KIT	NGS	1.000
ABL1	NGS	0.908
JAK1	CNA	0.861
SPEN	CNA	0.836
FOXL2	NGS	0.766
EPS15	CNA	0.732
STIL	CNA	0.727
HMGN2P46	CNA	0.721
Age	META	0.713
TP53	NGS	0.641
BLM	CNA	0.615
THRAP3	CNA	0.602
CDH11	CNA	0.602
MSI2	CNA	0.578
CRTC3	CNA	0.550
MYCL	NGS	0.543

MYCL	CNA	0.538
ATP1A1	CNA	0.532
TNFAIP3	CNA	0.521
SFPQ	CNA	0.480
APC	NGS	0.471
ERG	CNA	0.450
NOTCH2	CNA	0.441
RB1	NGS	0.426
CAMTA1	CNA	0.421
RPL22	CNA	0.413
PIK3CG	CNA	0.410
PTCH1	CNA	0.403
KNL1	CNA	0.398
ABL2	CNA	0.390
BTG1	CNA	0.389
ACSL6	CNA	0.386
ELK4	CNA	0.386

SETBP1	CNA	0.382
C15orf65	CNA	0.372
ARID1A	CNA	0.370
CDKN2B	CNA	0.361
MPL	CNA	0.338
CACNA1D	CNA	0.320
EGFR	CNA	0.319
JUN	CNA	0.318
TSHR	CNA	0.305
SUFU	CNA	0.303
AMER1	NGS	0.297
MTOR	CNA	0.297
FGFR2	CNA	0.293
NUP93	CNA	0.290
BCL9	CNA	0.286
VHL	NGS	0.284
U2AF1	CNA	0.281

Table 98: Small Intestine Adenocarcinoma – Small Intestine

GENE	TECH	IMP
KRAS	NGS	1.000
CDX2	CNA	0.866
FOXL2	NGS	0.862
SETBP1	CNA	0.853
FLT3	CNA	0.837
AURKB	CNA	0.762
FLT1	CNA	0.733
LCP1	CNA	0.691
SPECC1	CNA	0.621

LHFPL6	CNA	0.620
LPP	CNA	0.619
POU2AF1	CNA	0.613
Age	META	0.602
CDK8	CNA	0.590
BCL2	CNA	0.573
RB1	CNA	0.559
TP53	NGS	0.552
MYC	CNA	0.552
APC	NGS	0.551

Gender	META	0.535
RPN1	CNA	0.510
EBF1	CNA	0.499
ERCC5	CNA	0.497
KDSR	CNA	0.493
SDHC	CNA	0.488
HOXA11	CNA	0.479
SDHD	CNA	0.477
AFF3	CNA	0.474
GID4	CNA	0.473

ASXL1	CNA	0.469
GMPS	CNA	0.468
CDH1	CNA	0.465
ZNF217	CNA	0.457
FOXO1	CNA	0.456
CCNE1	CNA	0.455
EXT1	CNA	0.448

MLF1	CNA	0.441
FGF14	CNA	0.437
ABL2	CNA	0.435
CTCF	CNA	0.433
ARNT	CNA	0.428
C15orf65	CNA	0.427
CDKN2B	CNA	0.427

FHIT	CNA	0.422
ATP1A1	CNA	0.422
JAZF1	CNA	0.418
CDKN2A	CNA	0.417
EWSR1	CNA	0.410
CHIC2	CNA	0.408
MLLT11	CNA	0.407

Table 99: Stomach Gastrointestinal Stromal Tumor NOS – Stomach

GENE	TECH	IMP
c-KIT	NGS	1.000
PDGFRA	NGS	0.838
MAX	CNA	0.815
FOXL2	NGS	0.802
TSHR	CNA	0.684
BCL2L2	CNA	0.628
TP53	NGS	0.610
FOXA1	CNA	0.601
MSI2	CNA	0.591
NIN	CNA	0.578
NKX2-1	CNA	0.568
PDGFRA	CNA	0.536
SETBP1	CNA	0.460
CDH11	CNA	0.451
Age	META	0.449
Gender	META	0.440

CCNB1IP1	CNA	0.440
ROS1	CNA	0.439
BCL11B	CNA	0.438
CDH1	NGS	0.438
HSP90AA1	CNA	0.419
BCL2	CNA	0.405
CHEK2	CNA	0.391
ECT2L	CNA	0.371
NFKBIA	CNA	0.348
RAD51B	CNA	0.329
KRAS	NGS	0.301
JUN	CNA	0.300
PER1	CNA	0.299
PTEN	NGS	0.298
MPL	CNA	0.297
PDGFB	CNA	0.295
FGFR1	CNA	0.293

VHL	NGS	0.292
KTN1	CNA	0.292
USP6	CNA	0.274
ADGRA2	CNA	0.272
GPHN	CNA	0.271
TPM3	CNA	0.266
LPP	CNA	0.262
APC	NGS	0.261
BCL6	CNA	0.258
PMS2	NGS	0.255
AKT1	CNA	0.255
CTCF	CNA	0.254
GOLGA5	CNA	0.247
FGFR4	CNA	0.246
MUC1	CNA	0.244
TCL1A	CNA	0.240
PDE4DIP	CNA	0.240

Table 100: Stomach Signet Ring Cell Adenocarcinoma – Stomach

GENE	TECH	IMP
Age	META	1.000
CDX2	CNA	0.936
FOXL2	NGS	0.911
CDH1	NGS	0.898
LHFPL6	CNA	0.858
AFF3	CNA	0.815
BCL3	CNA	0.790
ERG	CNA	0.783
HOXD13	CNA	0.755
Gender	META	0.709
FANCC	CNA	0.686
EXT1	CNA	0.674
PBX1	CNA	0.664
RUNX1	CNA	0.663
CDKN2B	CNA	0.622

TGFBR2	CNA	0.616
BCL2	CNA	0.598
PRCC	CNA	0.595
NSD2	CNA	0.583
FNBP1	CNA	0.579
RPN1	CNA	0.578
MLLT11	CNA	0.577
CDK4	CNA	0.562
CTNNA1	CNA	0.561
c-KIT	NGS	0.554
HMGN2P46	CNA	0.552
TCF7L2	CNA	0.550
HIST1H4I	CNA	0.549
H3F3B	CNA	0.549
U2AF1	CNA	0.546
KRAS	NGS	0.546

USP6	CNA	0.546
FGFR2	CNA	0.543
FANCF	CNA	0.531
SETBP1	CNA	0.531
HOXD11	CNA	0.516
CDKN2A	CNA	0.514
WWTR1	CNA	0.513
MYC	CNA	0.509
CCNE1	CNA	0.499
CALR	CNA	0.485
HMGA2	CNA	0.483
LPP	CNA	0.473
TP53	NGS	0.466
CHEK2	CNA	0.464
NUTM2B	CNA	0.462
CDH11	CNA	0.461

BTG1	CNA	0.459	GID4	CNA	0.457	WRN	CNA	0.457

Table 101: Thyroid Carcinoma NOS – Thyroid

GENE	TECH	IMP	HOXA13	CNA	0.612	FHIT	CNA	0.533
NKX2-1	CNA	1.000	DDX6	CNA	0.600	TMPRSS2	CNA	0.531
Age	META	0.988	NDRG1	CNA	0.577	FANCF	CNA	0.530
FOXL2	NGS	0.980	CRKL	CNA	0.574	MUC1	CNA	0.524
HOXA9	CNA	0.756	BCL2	CNA	0.570	HOXA11	CNA	0.520
SBDS	CNA	0.750	CDH11	CNA	0.566	CARS	CNA	0.518
TP53	NGS	0.740	EBF1	CNA	0.559	DAXX	CNA	0.514
SOX10	CNA	0.728	KNL1	CNA	0.558	MYC	CNA	0.510
NF2	CNA	0.726	RAD51	CNA	0.554	HIST1H3B	CNA	0.506
ERG	CNA	0.719	HMGN2P46	CNA	0.553	DDIT3	CNA	0.497
HMGA2	CNA	0.686	CD274	CNA	0.553	LCP1	CNA	0.493
EWSR1	CNA	0.683	STAT5B	CNA	0.541	ERC1	CNA	0.492
GNAS	CNA	0.671	TSHR	CNA	0.541	SETBP1	CNA	0.489
MLLT11	CNA	0.662	CRTC3	CNA	0.534	TRIM33	NGS	0.488
KDSR	CNA	0.646	FANCA	CNA	0.533	TTL	CNA	0.481
Gender	META	0.636	AKAP9	NGS	0.533	PAK3	NGS	0.479
LHFPL6	CNA	0.628	BRCA1	CNA	0.533	PAX8	CNA	0.478

Table 102: Thyroid Carcinoma Anaplastic NOS – Thyroid

GENE	TECH	IMP	ERBB3	CNA	0.603	FLT1	CNA	0.474
TRRAP	CNA	1.000	KIAA1549	CNA	0.594	BCL9	CNA	0.469
BRAF	NGS	0.847	FUS	CNA	0.578	CBFB	CNA	0.463
CDH1	NGS	0.842	SPEN	CNA	0.559	BCL11A	NGS	0.459
WISP3	CNA	0.832	PDGFRA	CNA	0.548	CDKN2A	CNA	0.453
Age	META	0.782	NRAS	NGS	0.547	MN1	CNA	0.451
Gender	META	0.744	KDSR	CNA	0.534	AFF3	CNA	0.448
MYC	CNA	0.706	LHFPL6	CNA	0.533	BAP1	CNA	0.434
VHL	NGS	0.705	FGF14	CNA	0.520	CDKN2B	CNA	0.433
CDX2	CNA	0.680	IGF1R	CNA	0.517	HOXA9	CNA	0.432
PDE4DIP	CNA	0.670	EBF1	CNA	0.515	RB1	NGS	0.431
SBDS	CNA	0.666	HOOK3	CNA	0.510	PTCH1	CNA	0.424
KRAS	NGS	0.637	NCKIPSD	CNA	0.494	TP53	NGS	0.421
IDH1	NGS	0.636	ARID1A	CNA	0.490	PBRM1	CNA	0.417
FHIT	CNA	0.636	PBX1	CNA	0.482	CHIC2	CNA	0.412
PTEN	NGS	0.629	SPECC1	CNA	0.479	ABL2	NGS	0.412
ELK4	CNA	0.619	CLP1	CNA	0.475	HOXA13	CNA	0.409

5

Table 103: Thyroid Papillary Carcinoma of Thyroid – Thyroid

GENE	TECH	IMP		FOXL2	NGS	0.922		MYC	CNA	0.752	
BRAF	NGS	1.000		NKX2-1	CNA	0.798		RALGDS	NGS	0.728	

TP53	NGS	0.727		CTNNA1	CNA	0.477		FNBP1	CNA	0.405	
SETBP1	CNA	0.642		H3F3B	CNA	0.465		TPR	CNA	0.404	
EXT1	CNA	0.608		AFF1	CNA	0.465		TCEA1	CNA	0.404	
KDSR	CNA	0.604		APC	CNA	0.460		MAF	CNA	0.399	
KLHL6	CNA	0.560		ITK	CNA	0.452		WWTR1	CNA	0.395	
EBF1	CNA	0.560		ABL1	CNA	0.441		USP6	CNA	0.395	
YWHAE	CNA	0.555		Gender	META	0.440		PRKDC	CNA	0.385	
FHIT	CNA	0.529		NR4A3	CNA	0.431		TAL2	CNA	0.383	
Age	META	0.515		NDRG1	CNA	0.431		SET	CNA	0.379	
U2AF1	CNA	0.512		IGF1R	CNA	0.429		MCL1	CNA	0.372	
SLC34A2	CNA	0.498		FBXW7	CNA	0.422		CRKL	CNA	0.371	
SRSF2	CNA	0.498		RUNX1T1	CNA	0.422		ZNF521	CNA	0.370	
AKT3	CNA	0.492		FANCF	CNA	0.421		ETV5	CNA	0.367	
COX6C	CNA	0.490		PDE4DIP	CNA	0.414		CDX2	CNA	0.365	
TFRC	CNA	0.485		IKZF1	CNA	0.411		ERG	CNA	0.361	

Table 104: Tonsil Oropharynx Tongue Squamous Carcinoma – Head, Face or Neck, NOS

GENE	TECH	IMP								
SOX2	CNA	1.000	FHIT	CNA	0.773	TPM3	CNA	0.675
LPP	CNA	0.999	PRCC	CNA	0.768	NF2	CNA	0.667
KLHL6	CNA	0.995	CHEK2	CNA	0.758	FGF10	CNA	0.661
FOXL2	NGS	0.977	FLI1	CNA	0.757	MITF	CNA	0.661
Gender	META	0.897	CRKL	CNA	0.757	VHL	CNA	0.660
CACNA1D	CNA	0.888	TP53	NGS	0.740	BCL9	CNA	0.660
SDHD	CNA	0.860	PPARG	CNA	0.736	CREB3L2	CNA	0.659
ZBTB16	CNA	0.859	CBL	CNA	0.729	EWSR1	CNA	0.658
BCL6	CNA	0.851	FANCG	CNA	0.727	HSP90AA1	CNA	0.658
RPN1	CNA	0.846	NTRK2	CNA	0.716	FANCC	CNA	0.658
TGFBR2	CNA	0.845	PBRM1	CNA	0.715	NDRG1	CNA	0.644
Age	META	0.810	POU2AF1	CNA	0.705	CDKN2A	CNA	0.641
SYK	CNA	0.807	PRKDC	CNA	0.705	ETV5	CNA	0.639
TFRC	CNA	0.793	KIAA1549	CNA	0.699	RAF1	CNA	0.633
PCSK7	CNA	0.789	EGFR	CNA	0.692	EPHB1	CNA	0.628
KMT2A	CNA	0.780	WWTR1	CNA	0.691	PAFAH1B2	CNA	0.628
			TRIM27	CNA	0.680	ASXL1	CNA	0.618

Table 105: Transverse Colon Adenocarcinoma NOS – Colon

GENE	TECH	IMP								
APC	NGS	1.000	BCL2	CNA	0.763	SOX2	CNA	0.574
CDX2	CNA	0.969	Age	META	0.732	ERCC5	CNA	0.568
FLT3	CNA	0.902	KRAS	NGS	0.701	ZNF217	CNA	0.563
FOXL2	NGS	0.880	BRAF	NGS	0.637	TRRAP	NGS	0.554
SETBP1	CNA	0.842	KDSR	CNA	0.637	EPHA5	CNA	0.552
LHFPL6	CNA	0.778	ASXL1	CNA	0.620	MCL1	CNA	0.550
FLT1	CNA	0.769	HOXA9	CNA	0.595	SFPQ	CNA	0.548
			AURKA	CNA	0.584	LCP1	CNA	0.547

KLHL6	CNA	0.538	FOXO1	CNA	0.487	CBFB	CNA	0.450
EBF1	CNA	0.528	CDKN2B	CNA	0.479	GNA13	CNA	0.447
WWTR1	CNA	0.521	SMAD4	CNA	0.477	SDC4	CNA	0.443
ZNF521	NGS	0.516	COX6C	CNA	0.469	CACNA1D	CNA	0.442
CCNE1	CNA	0.511	SPEN	CNA	0.465	RB1	CNA	0.442
GNAS	CNA	0.505	PRRX1	CNA	0.464	TOP1	CNA	0.437
Gender	META	0.501	U2AF1	CNA	0.464	JAZF1	CNA	0.436
CDH1	CNA	0.493	CDKN2A	CNA	0.455	RUNX1	CNA	0.436
ZMYM2	CNA	0.492	TP53	NGS	0.453	HMGN2P46	CNA	0.422

Table 106: Urothelial Bladder Adenocarcinoma NOS – Bladder

GENE	TECH	IMP						
CTNNA1	CNA	1.000	IKZF1	CNA	0.546	RAC1	CNA	0.453
FOXL2	NGS	0.945	Gender	META	0.544	CEBPA	CNA	0.451
ZNF217	CNA	0.770	FGF10	CNA	0.533	PCSK7	CNA	0.448
FNBP1	CNA	0.693	SDC4	CNA	0.533	CBFB	CNA	0.447
EWSR1	CNA	0.687	HOXA13	CNA	0.518	SET	CNA	0.445
IL7R	CNA	0.686	WWTR1	CNA	0.517	STAT3	CNA	0.441
TP53	NGS	0.643	ARID2	NGS	0.513	RICTOR	CNA	0.439
ACSL6	CNA	0.642	APC	NGS	0.508	STAT5B	CNA	0.433
CTCF	CNA	0.639	MTOR	CNA	0.497	MYC	CNA	0.432
BCL3	CNA	0.637	ACSL3	CNA	0.497	SDHB	CNA	0.425
LIFR	CNA	0.636	CREB3L2	CNA	0.496	HOXA11	CNA	0.425
CHEK2	CNA	0.628	EPHA3	CNA	0.475	SETBP1	CNA	0.422
Age	META	0.606	EP300	CNA	0.468	HLF	CNA	0.418
CDH1	NGS	0.577	DDX6	CNA	0.461	PAFAH1B2	CNA	0.410
VHL	NGS	0.577	CDK4	CNA	0.457	FANCD2	NGS	0.410
CD79A	NGS	0.562	BCL2L11	CNA	0.455	CDK6	CNA	0.404
			CDX2	CNA	0.455	GNAS	CNA	0.391

Table 107: Urothelial Bladder Carcinoma NOS – Bladder

GENE	TECH	IMP						
Age	META	1.000	MYCL	CNA	0.709	USP6	CNA	0.574
VHL	CNA	0.971	FGFR2	CNA	0.694	CBFB	CNA	0.559
CREBBP	CNA	0.939	KDM6A	NGS	0.658	MDS2	CNA	0.558
FOXL2	NGS	0.912	TP53	NGS	0.656	HEY1	CNA	0.556
Gender	META	0.836	CTNNA1	CNA	0.648	EWSR1	CNA	0.554
CDKN2B	CNA	0.835	KRAS	NGS	0.623	ZNF331	CNA	0.551
FANCC	CNA	0.806	XPC	CNA	0.612	CARS	CNA	0.550
GATA3	CNA	0.797	LHFPL6	CNA	0.612	FBXW7	CNA	0.545
GNA13	CNA	0.755	CCNE1	CNA	0.608	TMPRSS2	CNA	0.544
IL7R	CNA	0.748	U2AF1	CNA	0.602	ARID1A	CNA	0.539
RAF1	CNA	0.736	PPARG	CNA	0.602	PAX3	CNA	0.533
WISP3	CNA	0.728	ERG	CNA	0.596	MECOM	CNA	0.526
ASXL1	CNA	0.722	ACKR3	CNA	0.580	CACNA1D	CNA	0.524
			CDKN2A	CNA	0.579	WWTR1	CNA	0.523

CTCF	CNA	0.520
CDH11	CNA	0.518
RPN1	CNA	0.518

CDH1	CNA	0.515
ABL2	NGS	0.510
ETV5	CNA	0.505

HMGN2P46	CNA	0.501
FANCD2	CNA	0.501
VHL	NGS	0.500

Table 108: Urothelial Bladder Squamous Carcinoma– Bladder

GENE	TECH	IMP
Age	META	1.000
FOXL2	NGS	0.934
IL7R	CNA	0.857
CDH1	NGS	0.808
ABL2	NGS	0.808
TFRC	CNA	0.785
KLHL6	CNA	0.733
LPP	CNA	0.696
WWTR1	CNA	0.696
EBF1	CNA	0.689
CDKN2C	CNA	0.665
c-KIT	NGS	0.656
AFF1	CNA	0.591
ETV5	CNA	0.574
Gender	META	0.566
CNBP	CNA	0.559

FHIT	CNA	0.522
KRAS	NGS	0.519
TP53	NGS	0.512
SOX2	CNA	0.510
MLLT11	CNA	0.506
FANCF	CNA	0.503
CDKN2A	CNA	0.501
EPS15	CNA	0.497
RPN1	CNA	0.484
CDH1	CNA	0.478
CDK4	CNA	0.474
INHBA	CNA	0.474
MLF1	CNA	0.467
JAK2	CNA	0.467
PRKDC	CNA	0.463
JAZF1	CNA	0.458
KMT2A	CNA	0.452

EPHB1	CNA	0.448
COX6C	CNA	0.445
ARID1A	CNA	0.445
CTLA4	CNA	0.443
CACNA1D	CNA	0.439
BAP1	CNA	0.433
EXT1	CNA	0.432
NUP98	CNA	0.431
NPM1	CNA	0.429
GID4	CNA	0.429
LIFR	CNA	0.425
FANCC	CNA	0.425
NOTCH1	NGS	0.422
GRIN2A	CNA	0.420
MAML2	CNA	0.416
STAT3	CNA	0.412
TERT	CNA	0.410

Table 109: Urothelial Carcinoma NOS – Bladder

GENE	TECH	IMP
GATA3	CNA	1.000
Age	META	0.820
ASXL1	CNA	0.698
CDKN2A	CNA	0.637
Gender	META	0.637
CDKN2B	CNA	0.634
ATIC	CNA	0.577
EBF1	CNA	0.575
NSD1	CNA	0.567
PPARG	CNA	0.550
ZNF331	CNA	0.545
ACSL6	CNA	0.535
TP53	NGS	0.532
RAF1	CNA	0.517
KRAS	NGS	0.517
CARS	CNA	0.511

KMT2D	NGS	0.510
FGFR2	CNA	0.501
EWSR1	CNA	0.492
VHL	CNA	0.491
NR4A3	CNA	0.482
FGFR3	NGS	0.481
c-KIT	NGS	0.479
PAX3	CNA	0.479
CTNNA1	CNA	0.477
ZNF217	CNA	0.475
XPC	CNA	0.473
FGF10	CNA	0.473
MYC	CNA	0.465
MYCL	CNA	0.463
KDM6A	NGS	0.461
EXT2	CNA	0.459
CTLA4	CNA	0.457

ELK4	CNA	0.455
BARD1	CNA	0.454
LHFPL6	CNA	0.453
KLHL6	CNA	0.452
APC	NGS	0.449
CCNE1	CNA	0.445
IL7R	CNA	0.441
DDB2	CNA	0.440
PTCH1	CNA	0.440
ARID1A	CNA	0.438
PBX1	CNA	0.432
FLT1	CNA	0.432
MLLT11	CNA	0.431
BCL6	CNA	0.431
CASP8	CNA	0.426
ITK	CNA	0.424
FANCF	CNA	0.422

5

Table 110: Uterine Endometrial Stromal Sarcoma NOS – FGTP

GENE	TECH	IMP
ETV1	CNA	1.000
FOXL2	NGS	0.967
HNRNPA2B1	CNA	0.957
PMS2	CNA	0.809
TGFBR2	CNA	0.734
Gender	META	0.726
TP53	NGS	0.690
Age	META	0.688
SPECC1	CNA	0.684
FANCC	CNA	0.683
INHBA	CNA	0.601
CDH1	CNA	0.570
RAC1	CNA	0.570
PTCH1	CNA	0.569
PDE4DIP	CNA	0.565
MAP2K4	CNA	0.541
CDH1	NGS	0.539
AFF1	CNA	0.520
ERG	CNA	0.512
DDR2	CNA	0.507
TERT	CNA	0.498
NR4A3	CNA	0.497
SDC4	CNA	0.483
VHL	NGS	0.447
RPN1	CNA	0.440
FANCE	CNA	0.430
PCM1	NGS	0.415
TOP1	CNA	0.414
ZNF217	CNA	0.409
PPARG	CNA	0.396
PDCD1LG2	CNA	0.396
RUNX1	CNA	0.368
RAP1GDS1	CNA	0.367
KRAS	NGS	0.360
FAM46C	CNA	0.359
FCRL4	CNA	0.357
HOXD13	CNA	0.341
FH	CNA	0.337
CDX2	CNA	0.328
CACNA1D	CNA	0.327
CNBP	CNA	0.326
BCL6	CNA	0.325
NDRG1	CNA	0.321
XPC	CNA	0.310
PTEN	NGS	0.310
CDK12	CNA	0.308
WRN	CNA	0.306
SRGAP3	CNA	0.302
JAK1	CNA	0.289
ESR1	CNA	0.289

Table 111: Uterine Leiomyosarcoma NOS – FGTP

GENE	TECH	IMP
RB1	CNA	1.000
FOXL2	NGS	0.966
SPECC1	CNA	0.943
Age	META	0.868
JAK1	CNA	0.830
PDCD1	CNA	0.825
PRRX1	CNA	0.795
Gender	META	0.790
ACKR3	CNA	0.771
ATIC	CNA	0.767
LCP1	CNA	0.762
HERPUD1	CNA	0.740
FANCC	CNA	0.739
GID4	CNA	0.728
NUP93	CNA	0.716
CDH1	CNA	0.692
PTCH1	CNA	0.686
PAX3	CNA	0.676
EBF1	CNA	0.665
SYK	CNA	0.659
WDCP	CNA	0.619
CBFB	CNA	0.612
ESR1	CNA	0.605
KLHL6	CNA	0.604
NTRK2	CNA	0.587
MYCN	CNA	0.578
JUN	CNA	0.574
CTCF	CNA	0.573
CRTC3	CNA	0.566
SOX2	CNA	0.560
RPN1	CNA	0.559
FOXO1	CNA	0.556
LHFPL6	CNA	0.548
LRIG3	CNA	0.547
PDGFRA	CNA	0.540
PBX1	CNA	0.538
NTRK3	CNA	0.531
IGF1R	CNA	0.530
MAP2K4	CNA	0.522
KDR	CNA	0.518
DNMT3A	CNA	0.494
CDKN2B	CNA	0.491
IDH1	CNA	0.482
BMPR1A	CNA	0.478
NUTM2B	CNA	0.477
KDSR	CNA	0.475
KIT	CNA	0.474
AFF3	CNA	0.470
TP53	NGS	0.467
TPM4	CNA	0.462

Table 112: Uterine Sarcoma NOS – FGTP

GENE	TECH	IMP
HOXD13	CNA	1.000
FOXL2	NGS	0.972
CACNA1D	CNA	0.887
Gender	META	0.870
MAX	CNA	0.799
TTL	CNA	0.778
Age	META	0.773
HMGA2	CNA	0.751
MITF	CNA	0.739
PRRX1	CNA	0.736
NF2	CNA	0.728
PRDM1	CNA	0.718
PML	CNA	0.697
RB1	CNA	0.678
CDKN2B	CNA	0.677
DDR2	CNA	0.676
HOXA11	CNA	0.665

HOXA9	CNA	0.645	USP6	CNA	0.545	ADGRA2	CNA	0.473
KIT	CNA	0.643	PDE4DIP	CNA	0.538	SEPT5	CNA	0.463
CDKN2A	CNA	0.630	IDH2	CNA	0.537	FGFR2	CNA	0.454
PDGFRA	CNA	0.614	TP53	NGS	0.534	PSIP1	CNA	0.441
ALK	NGS	0.610	MYC	CNA	0.531	FGFR1	CNA	0.439
FNBP1	CNA	0.600	PLAG1	CNA	0.519	FHIT	CNA	0.438
CDH1	CNA	0.597	ERCC3	CNA	0.497	ZNF217	CNA	0.433
WRN	CNA	0.593	HOXD11	CNA	0.495	RALGDS	CNA	0.431
SNX29	CNA	0.574	FANCA	CNA	0.487	AFF3	CNA	0.428
GID4	CNA	0.572	FCRL4	CNA	0.485	SFPQ	CNA	0.421
BCL11A	CNA	0.559	JAZF1	CNA	0.484	MAP2K4	CNA	0.417

Table 113: Uveal Melanoma – Eye

GENE	TECH	IMP						
IRF4	CNA	1.000	LPP	CNA	0.565	ETV5	CNA	0.419
HEY1	CNA	0.873	MLF1	CNA	0.525	UBR5	CNA	0.415
FOXL2	NGS	0.858	KLHL6	CNA	0.523	FOXL2	CNA	0.406
EXT1	CNA	0.826	NCOA2	CNA	0.522	HSP90AB1	CNA	0.401
PAX3	CNA	0.785	c-KIT	NGS	0.519	HIST1H4I	CNA	0.401
TRIM27	CNA	0.780	TFRC	CNA	0.511	SETBP1	CNA	0.389
TP53	NGS	0.730	WWTR1	CNA	0.509	KRAS	NGS	0.383
GNA11	NGS	0.710	COX6C	CNA	0.507	NR4A3	CNA	0.378
GNAQ	NGS	0.707	HIST1H3B	CNA	0.503	DEK	CNA	0.372
RUNX1T1	CNA	0.679	BAP1	NGS	0.491	TCEA1	CNA	0.362
SOX10	CNA	0.668	SF3B1	NGS	0.466	MUC1	CNA	0.354
MYC	CNA	0.658	GATA2	CNA	0.465	USP6	CNA	0.351
BCL6	CNA	0.650	EWSR1	CNA	0.457	YWHAE	CNA	0.348
RPN1	CNA	0.616	GMPS	CNA	0.456	SOX2	CNA	0.345
ABL2	NGS	0.598	BCL2	CNA	0.453	IDH1	NGS	0.341
SRGAP3	CNA	0.570	CNBP	CNA	0.452	VHL	NGS	0.340
			DAXX	CNA	0.427	CDX2	CNA	0.333

Table 114: Vaginal Squamous Carcinoma – FGTP

GENE	TECH	IMP						
CNBP	CNA	1.000	FNBP1	CNA	0.792	ARID2	NGS	0.623
RPN1	CNA	0.985	CD274	CNA	0.778	WT1	CNA	0.605
FOXL2	NGS	0.980	CBFB	CNA	0.774	ABI1	CNA	0.602
KMT2D	NGS	0.961	PPARG	CNA	0.755	KMT2C	NGS	0.586
VHL	NGS	0.927	MLLT3	CNA	0.750	TFRC	CNA	0.578
SPEN	CNA	0.917	WWTR1	CNA	0.749	RAF1	CNA	0.560
Gender	META	0.909	FANCC	CNA	0.682	SOX2	CNA	0.552
FHIT	CNA	0.894	PDCD1LG2	CNA	0.661	ETV5	CNA	0.548
CDH1	NGS	0.874	PAX3	CNA	0.651	CDKN2C	CNA	0.546
TP53	NGS	0.872	KLHL6	CNA	0.640	BARD1	CNA	0.545
JUN	CNA	0.807	SDHC	CNA	0.629	Age	META	0.531
			HOXD13	CNA	0.626	MAF	CNA	0.523

MECOM	CNA	0.514	LPP	CNA	0.474	MLLT10	CNA	0.454
SDHB	CNA	0.511	ESR1	CNA	0.472	KDSR	CNA	0.450
MDS2	CNA	0.498	CDH11	CNA	0.467	CDKN2B	CNA	0.447
ASXL1	CNA	0.492	GSK3B	CNA	0.466	TRRAP	CNA	0.447
EP300	CNA	0.481	CLP1	CNA	0.464	HOXD11	CNA	0.446

Table 115: Vulvar Squamous Carcinoma – FGTP

GENE	TECH	IMP	KLHL6	CNA	0.674	U2AF1	CNA	0.596
CNBP	CNA	1.000	SPECC1	CNA	0.666	PRDM1	CNA	0.592
CACNA1D	CNA	0.975	EXT1	CNA	0.665	SET	CNA	0.591
FOXL2	NGS	0.973	CDKN2B	CNA	0.653	NTRK2	CNA	0.590
Gender	META	0.967	CAMTA1	CNA	0.651	GNAS	CNA	0.583
SDHB	CNA	0.928	CHEK2	CNA	0.642	FNBP1	CNA	0.579
SYK	CNA	0.924	RPL22	CNA	0.641	PDCD1LG2	CNA	0.579
Age	META	0.832	RPN1	CNA	0.641	PBX1	CNA	0.579
TAL2	CNA	0.817	NR4A3	CNA	0.634	TRIM27	CNA	0.578
TGFBR2	CNA	0.807	CREB3L2	CNA	0.629	CD274	CNA	0.576
MTOR	CNA	0.807	TP53	NGS	0.629	TFRC	CNA	0.567
HOOK3	CNA	0.802	NUP93	CNA	0.624	STIL	CNA	0.566
SETD2	CNA	0.773	ARID1A	CNA	0.623	PAX3	CNA	0.559
PRKDC	CNA	0.729	CBFB	CNA	0.623	ETV5	CNA	0.556
PBRM1	CNA	0.709	FANCC	CNA	0.614	EWSR1	CNA	0.555
MDS2	CNA	0.704	BCL9	CNA	0.614	BCL11A	CNA	0.555
KAT6A	CNA	0.699	FGF4	CNA	0.604	XPC	CNA	0.554

Table 116: Skin Trunk Melanoma – Skin

GENE	TECH	IMP	Gender	META	0.558	FHIT	CNA	0.453
IRF4	CNA	1.000	SDHAF2	CNA	0.547	BCL6	CNA	0.444
FOXL2	NGS	0.900	HIST1H4I	CNA	0.540	MKL1	CNA	0.442
BRAF	NGS	0.853	ELK4	CNA	0.519	DAXX	CNA	0.428
SOX10	CNA	0.842	NRAS	NGS	0.518	KRAS	NGS	0.419
TP53	NGS	0.777	CCDC6	CNA	0.518	Age	META	0.414
TCF7L2	CNA	0.757	FLI1	CNA	0.517	PTCH1	CNA	0.409
FGFR2	CNA	0.734	SOX2	CNA	0.516	c-KIT	NGS	0.401
CDKN2A	CNA	0.734	TET1	CNA	0.511	NF2	CNA	0.399
EP300	CNA	0.686	TRIM26	CNA	0.509	BRAF	CNA	0.394
CDKN2B	CNA	0.669	CREB3L2	CNA	0.506	POT1	CNA	0.392
DEK	CNA	0.660	NOTCH2	CNA	0.505	MYCN	CNA	0.388
SYK	CNA	0.644	KIAA1549	CNA	0.504	CACNA1D	CNA	0.383
TRIM27	CNA	0.607	USP6	CNA	0.500	APC	NGS	0.378
LHFPL6	CNA	0.580	FOXP1	CNA	0.482	LRP1B	NGS	0.376
CRTC3	CNA	0.575	ESR1	CNA	0.466	TET1	NGS	0.372
FANCC	CNA	0.572	SDHD	CNA	0.458	BCL2	CNA	0.363

5

In many cases, the features in the biosignatures in **Tables 2-116** comprise gene copy number (CNA or CNV). Cells are typically diploid with two copies of each gene. However, cancer may lead to various genomic alterations which can alter copy number. In some instances, copies of genes are amplified (gained), whereas in other instances copies of genes are lost. Genomic alterations can affect different regions of a chromosome. For example, gain or loss may occur within a gene, at the gene level, or within groups of neighboring genes. Gain or loss may also be observed at the level of cytogenetic bands or even larger portions of chromosomal arms. Thus, analysis of such proximate regions to a gene may provide similar or even identical information to the gene itself. Accordingly, the methods provided herein are not limited to determining copy number of the specified genes, but also expressly contemplate the analysis of proximate regions to the genes, wherein such proximate regions provide similar or the same level of information. Copy analysis of genes, SNPs or other features within the band may be used within the scope of the systems and methods described herein.

As described in the **Examples** herein, the methods for classifying the attributes of the cancer may calculate a probability that the biosignature corresponds to the at least one pre-determined biosignature. In some embodiments, the method comprises a pairwise comparison between two candidate attributes, and a probability is calculated that the sample biosignature corresponds to either one of the at least one pre-determined biosignatures. In some embodiments, the pairwise comparison between the two candidate attributes is determined using a machine learning classification algorithm, wherein optionally the machine learning classification algorithm comprises a voting module. In some embodiments, the voting module is as provided herein, e.g., as described above. In some embodiments, a plurality of probabilities are calculated for a plurality of pre-determined biosignatures. In some embodiments, the probabilities are ranked. In some embodiments, the probabilities are compared to a threshold, wherein optionally the comparison to the threshold is used to determine whether the classification of the desired attribute of the cancer is likely, unlikely, or indeterminate. Systems and methods for implementing the classifications are provided herein. For example, see **FIGs. 1A-I** and related text.

In some embodiments, the levels of specificity for the attributes of the patient sample are determined at the level of an organ group. In one non-limiting example, the organ group that is predicted may be selected from bladder; skin; lung; head, face or neck (NOS); esophagus; female genital tract (FGT); brain; colon; prostate; liver, gall bladder, ducts; breast; eye; stomach; kidney; and pancreas. As desired, the systems and methods provided herein may employ biosignatures determined at the level of a primary tumor location and a histology, see, e.g., **Tables 2-116**, and the organ group is then determined based on the most probable primary tumor location + histology. As a non-limiting example, **Tables 2-116** herein provide biosignatures for primary tumor location + histology, and the table headers report both the primary tumor location + histology and corresponding organ group.

The disclosure contemplates that selections may be made from the biosignatures provided herein, e.g., in **Tables 2-116** for primary tumor location + histology. Use of the features in the tables

may provide optimal origin prediction, although selection may be made so long as the selections retain the ability to meet desired performance criteria, such as but not limited to accuracy of at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99%. In some embodiments, the biosignature comprises the top 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the feature biomarkers with the highest Importance value in the corresponding table (i.e., **Tables 2-116**). In some embodiments, the biosignature comprises the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 feature biomarkers with the highest Importance value in the corresponding table (i.e., **Tables 2-116**). In some embodiments, the biosignature comprises at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 40%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 feature biomarkers with the highest Importance value in the corresponding table (i.e., **Tables 2-116**). In some embodiments, the biosignature comprises at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, or 100 feature biomarkers with the highest Importance value in the corresponding table. As a non-limiting example, the biosignature may comprise at least 1, 2, 3, 4, or 5 of the top 10, 20 or 50 features. Provided herein is any selection of biomarkers that can be used to obtain a desired performance for predicting the attribute of interest, be it a primary location, organ group, histology, or disease/cancer type.

Systems for implementing the methods are also provided herein. See, e.g., **FIGs. 1F-1G** and related disclosure.

In some embodiments, the systems and methods of the invention implement systems and methods for predicting sample attributes as detailed in International Patent Publication WO/2020/146554, entitled Genomic Profiling Similarity and based on International Patent Application PCT/US2020/012815 filed on January 8, 2020, the entire contents of which application is hereby incorporated by reference in its entirety.

Expression-based Predictor of Disease Type

The section above provides a machine learning based classifier to predict attributes of a cancer sample based on molecular analysis of the sample, such attributes comprising a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof. The methods and

systems provided accordingly can be applied with various biological analytes as desired, e.g., nucleic acids, e.g., DNA and RNA, and protein. The section above and WO/2020/146554 demonstrated such analysis using genomic DNA. There have been attempts to use mRNA expression profiling to build classifiers or predictors of such attributes. mRNA is an attractive analyte because it can be assessed using well established techniques, e.g., PCR or microarray. mRNA sequences and expression can also be assessed in a high throughput manner using next generation sequencing, including without limitation whole transcriptome sequencing. However, RNA also has drawbacks. Consider analysis of a tumor sample using IHC for protein expression. A stained IHC slide will show areas of normal versus tumor tissue, and also other features such as nuclear or membrane staining of the protein. Thus a pathologist can focus on areas of interest for analysis of the protein expression levels and patterns. However, RNA would comprise a mix of RNA from different cells and cell types within the sample, without cellular location, and wherein background amounts of various RNA transcripts may vary greatly between cells. In particular, RNA classifiers may struggle with low neoplastic percentage in metastatic sites which is where TOO identification is often most needed. Accordingly, an RNA expression based assay may be confounded by the particular sample and cells from which the RNA is extracted. See, e.g., Hayashi et al., Randomized Phase II Trial Comparing Site-Specific Treatment Based on Gene Expression Profiling with Carboplatin and Paclitaxel for Patients with Cancer of Unknown Primary Site, J Clin Oncol **37**:57-579 (finding no significant improvement in one-year survival based on site-specific treatment as determined by gene expression profiling). Thus, there is a need to improve analysis of RNA based characterization of cancer samples.

Herein, we provide systems and methods to predict sample origin of a tumor sample based on RNA expression analysis with much higher accuracy than previously achieved. The general scheme **400** for performing the prediction is shown in **FIG. 4A**. RNA expression data **401** is collected for the desired transcripts. Any useful method of acquiring such data can be employed. For example, we used whole transcriptome sequencing analysis (WTS; RNA-seq) using the Illumina NGS platform, which methodology queries over 22,000 transcripts in a single assay. The raw expression data is processed via any desired methodology for processing. See, e.g., Li et al., Comparing the Normalization Methods for the Differential Analysis of Illumina High-Throughput RNA-Seq Data, BMC Bioinformatics. 2015 Oct 28;16:347. doi: 10.1186/s12859-015-0778-7; Abbas-Aghababazadeh and Fridley, Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing, PLoS One. 2018; 13(10): e0206312. In some embodiments, the RNA expression data **402** is normalized using Trimmed Mean of M-values (TMM). See Robinson and Oshlack, A Scaling Normalization Method for Differential Expression Analysis of RNA-seq Data, Genome Biol. 2010;11(3):R25. doi: 10.1186/gb-2010-11-3-r25. Epub 2010 Mar 2.

Continuing with **FIG. 4A**, normalized expression data for the target transcripts can be used to train machine learning models for various attributes of interest, including without limitation a primary tumor origin, cancer/disease type **403**, organ group **404**, and/or histology **405**. In some embodiments,

the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, or all 38 of prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas, rectum, colon, breast, intrahepatic bile duct, cecum, gastroesophageal junction, frontal lobe, kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin. In some embodiments, the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or all 21 of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma. In some embodiments, the cancer/disease type **403** consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or all 28 of adrenal cortical carcinoma; bile duct, cholangiocarcinoma; breast carcinoma; central nervous system (CNS); cervix carcinoma; colon carcinoma; endometrium carcinoma; gastrointestinal stromal tumor (GIST); gastroesophageal carcinoma; kidney renal cell carcinoma; liver hepatocellular carcinoma; lung carcinoma; melanoma; meningioma; Merkel; neuroendocrine; ovary granulosa cell tumor; ovary, fallopian, peritoneum; pancreas carcinoma; pleural mesothelioma; prostate adenocarcinoma; retroperitoneum; salivary and parotid; small intestine adenocarcinoma; squamous cell carcinoma; thyroid carcinoma; urothelial carcinoma; uterus. In some embodiments, the organ group **404** consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or all 17 of adrenal gland; bladder; brain; breast; colon; eye; female genital tract and peritoneum (FGTP); gastroesophageal; head, face or neck, NOS; kidney; liver, gallbladder, ducts; lung; pancreas; prostate; skin; small intestine; thyroid. In some embodiments, the histology **405** consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or all 29 of adenocarcinoma, adenoid cystic carcinoma, adenosquamous carcinoma, adrenal cortical carcinoma, astrocytoma, carcinoma, carcinosarcoma, cholangiocarcinoma, clear cell carcinoma, ductal carcinoma in situ (DCIS), glioblastoma (GBM), GIST, glioma, granulosa cell tumor, infiltrating lobular carcinoma, leiomyosarcoma, liposarcoma, melanoma, meningioma, Merkel cell carcinoma, mesothelioma,

neuroendocrine, non-small cell carcinoma, oligodendroglioma, sarcoma, sarcomatoid carcinoma, serous, small cell carcinoma, squamous.

Various classification methodology can be applied to the chosen attributes as desired, including without limitation a neural network model, a linear regression model, a random forest model, a logistic regression model, a naive Bayes model, a quadratic discriminant analysis model, a K-nearest neighbor model, a support vector machine, or various forms of or combinations thereof. In some embodiments, the machine learning approach comprises an XGBoost multi-class classification. XGBoost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework. Combinations of classification methods can be employed. Calculations can be performed using various statistical analysis platforms, including without limitation R.

FIG. 4A illustrates a scenario wherein three different classifications **403-405** performed on the same transcript expression data. The classifications from each of these three models can be combined using another model, such as those described above. In some embodiments, the combination is also made using an XGBoost model. This mechanism of combining intermediate classifications of the chose attributes, such as the illustrated **403-405**, is an implementation of the voting scheme described herein (*see, e.g.,* **FIG. 1F** and related text) and provides for dynamic voting **406**. As a non-limiting example, consider that one of the intermediate models **403-405** is very accurate at making a given classification. In such case, that single model's classification may carry more weight than the two other intermediate models when making the final classification **407**. In such case, that model's classification may dominate the other intermediate models when making the final classification **407**. The various intermediate models can be assigned different weights when performing the dynamic voting **406**. Any such combination of one or more of the intermediate models can outweigh others. Thus the dynamic voting **406** can provide classification **407** based on trained and optimized contributions from each of the intermediate models.

In some embodiments, analysis of different types of analytes are combined in order to classify the input sample and estimate the desired one or more attributes. In this regard, **FIG. 4B** presents an exemplary variation **410** of scheme **400** that is shown in **FIG. 4A**. In this variation, both RNA transcript levels **411** and DNA **416** are used to classify the input sample. As noted herein, DNA and RNA have various strengths and weaknesses for predicting attributes of a biological sample. For example, DNA is relatively more stable and more uniform amongst different types of cells, whereas RNA is more dynamic and may be more indicative of differences within individual cells. Without being bound by theory, we hypothesized that a combination of genomic DNA analysis with RNA transcriptome analysis may provide optimal results. We term this combined classifier a "panomic" predictor. As desired, analysis from additional analytes such as other types of RNA and/or protein could also be input into the system in a similar manner. In the embodiment illustrated in **FIG. 4B**, the three intermediate RNA transcript models **412-414** are identical to **FIG. 4A 403-405** as described above, respectively. In addition, the figure shows DNA **416** input into the system. In some

embodiments, the DNA is processed using the 115 disease types as described above. *See, e.g.,* **Tables 2-116** and related discussion; *see also* **Examples 2-3**. In this case, the dynamic voting **415** is applied to the four intermediate models comprising RNA **412-414** and DNA **416**. Models assessing attributes based on alternate analytes may also be input into the dynamic voting module **415** in a similar manner. As described above, the dynamic voting mechanism is a variation of the voting scheme described herein (*see, e.g.,* **FIG. 1F** and related text) and provides for essentially dynamic voting between the inputs into the dynamic voting module **415** in order to provide the prediction/classification **417**. As a non-limiting example, consider that one of the intermediate models **412-414** or **416** are very accurate at making a given classification. In such case, that model's classification may outweigh the other intermediate models when making the final classification **417**. Similarly, two of the intermediate models may outperform the two other intermediate models for a given classification and may thus dominate in that setting, or three of the intermediate models may combine to provide a better classification with lesser input from the remaining model. Thus the dynamic voting **415** can provide classification **417** based on trained and optimized contributions from each of the intermediate models.

FIG. 4C illustrates a flowchart of an example of a process **400C** for training a dynamic voting engine. Process **400C** may be performed by a system such as the system **400** of **FIG. 4A** or **410** of **FIG. 4B**.

The dynamic voting engine such as the dynamic voting engine of **FIG. 4A**, **406**, **FIG. 4B**, **415** or **FIG. 1G**, **400** can be trained in a number of different ways. In one implementation, the dynamic voting engine can be trained to predict a target classification for a biological sample based on processing, by the dynamic voting engine, data corresponding to one or more initial classifications that were previously determined for a biological sample. In some implementations, the biological sample can include a cancer sample and the target classification can include an attribute for the cancer, including without limitation a TOO. In some implementations, the one or more previously determined classifications can be based on processing of DNA sequences of the biological sample, RNA sequences of the biological sample, or both.

The system can begin performance of the process **400C** by using one or more computers to obtain **410C**, from a database of labeled training data items, a labeled training data item. Each labeled training data item can include one or more initial classifications and a target classification. The one or more initial classifications can be based on or derived from actual data generated by one or more initial classification engines such as cancer type classification engine (e.g., **FIG. 4A**, **403** or **FIG. 4B**, **412**), an initial organ of origin engine (e.g., **FIG. 4A**, **404** or **FIG. 4B**, **413**), a histology engine (e.g., **FIG. 4A**, **405** or **FIG. 4B**, **414**), or a DNA analysis engine (e.g., **FIG. 4B**, **416**), based on processing, by one or more of the respective initial classification engines, data derived from the biological sample. The data derived from the biological sample can include DNA sequences of the sample, RNA sequences of the sample, or both. In other implementations, the one or more initial classifications can

be based on or derived from simulated data that is generated to represent initial classifications that ought to be generated by such initial classification models when such initial classification models process data such as DNA sequences, RNA sequences, or both, derived from the biological sample.

The system can continue performance of the process **400C** by using one or more computers to generate **420C** training input data for input to the dynamic voting engine. In some implementations, the training input data can include, for example, a numerical representation of the one or more initial classifications. For example, data that represents each of the initial classifications can be encoded into one or more fields of a data structure that is formatted for input to the dynamic voting engine.

The system can continue performance of the process **400C** by using one or more computers to process **430C** the generated training input data through the dynamic voting engine. In some implementations, the dynamic voting engine can include one or more machine learning models, e.g., one or more of a random forests, support vector machines, logistic regressions, K-nearest neighbors, artificial neural networks, naïve Bayes, quadratic discriminant analysis, Gaussian processes models, decision trees, or any combination thereof. In such implementations, processing the generated training input data through the dynamic voting engine can include processing the generated training input data through each layer of the one or more machine learning models. In some implementations, the dynamic voting engine includes an XGBoost decision-tree-based ensemble machine learning algorithm.

The system can continue performance of the process **400C** by using one or more computers to obtain **440C** the output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the training input data generated at stage **420C**. The system can then use one or more computers to determine a level of similarity between the output data generated by the dynamic voting engine that is obtained at stage **440C** and the label for the training data item obtained at stage **410C**. In some implementations, the level of similarity between the label of the training data item obtained at stage **410C** and the output data that is obtained at stage **440C** can include the difference between the label and the output data.

The system can continue performance of the process **400C** by using one or more computers to adjust **460C** one or more parameters of the dynamic voting engine based on the level of similarity between the output data and the label of the training data item obtained at stage 410C. The system can then continue to iteratively perform the process **400C** until the output data generated by the system and obtained at stage **440C** begins to match the label for the training data item obtained at stage **410C** within a threshold amount of error. In some implementations, the threshold of error can be zero error. In other implementations, the threshold can include less than 1% error, less than 2 % error, less than 5% error, less than 10% error, or the like. Once the system begins to detect that the dynamic voting engine is predicting output data that matches the label for the training input data processed by the

dynamic voting engine within a threshold amount of error, then the dynamic voting engine may be considered to be fully trained.

The systems **400**, **410** and variations thereof can be trained to desired panels of RNA transcripts in order to classify the at least one attribute of the cancer of interest. In some embodiments, the systems are trained using NGS based whole transcriptome sequencing data, e.g., mRNA from 22,000 genes. To avoid overfitting or similar error, analysis of such panels may require training data on tens of thousands of tumor samples. To further avoid issues faced relying on RNA transcript analysis, such as overfitting of data based on the high number of total mRNAs, we may train the systems using more limited sets of transcripts. Traditionally, proteins that have been used in IHC based tumor classification. *See, e.g.,* Lin and Liu, Immunohistochemistry in Undifferentiated Neoplasm/Tumor of Uncertain Origin, Arch Pathol Lab Med. 2014;138:1583–1610, which reference is incorporated herein by reference in its entirety. In some embodiments, the panel of mRNA transcripts used to implement the system comprise the mRNA encoding such proteins, and may further include various isoforms or related family members thereof. The correlation between RNA transcript expression and protein expression levels is noisy and tissue dependent, and thus one would not be able to predict *a priori* whether such an approach would yield acceptable results. *See, e.g.,* Edfors et al, Gene-specific correlation of RNA and protein levels in human cells and tissues, Mol Syst Biol. (2016) 12: 883; Franks A, et al (2017) Post-transcriptional regulation across human tissues. PLoS Comput Biol 13(5): e1005535. However, we hypothesized that the analysis of multiple genes would improve noise levels to achieve acceptable accuracy and unexpectedly found our approach to perform with high levels of accuracy.

Based on the above rational for identifying a subset of potentially useful RNA transcripts, we constructed a list of candidate biomarkers shown in **Table 117**. The table provides the official gene symbol and full name as reported by the National Center for Biotechnology Information (NCBI) Gene database with reference to the HUGO Gene Nomenclature Committee (HGNC) database. *See* www.ncbi.nlm.nih.gov/gene (NCBI Gene); www.genenames.org (HGNC). The NCBI's Gene ID is also provided. The "Aliases" column provides a non-exhaustive list of alternate descriptions for the genes such as alternate gene names, e.g., that may also be used herein. Comprehensive listings of alternate symbols are provided by the NCBI and HGNC databases, among others available and known to those of skill in the art (e.g., Ensembl, Genecards, etc).

Table 117 – RNA Transcripts used to Characterize Tumor Sample

Gene Symbol	Full Name	Aliases	NCBI Gene ID
ACVRL1	activin A receptor like type 1		94
AFP	alpha fetoprotein		174
ALPP	alkaline phosphatase, placental		250
AMACR	alpha-methylacyl-CoA racemase		23600
ANKRD30A	ankyrin repeat domain 30A	NY-BR-1	91074

ANO1	anoctamin 1	DOG1	55107
AR	androgen receptor		367
ARG1	arginase 1		383
BCL2	BCL2 apoptosis regulator		596
BCL6	BCL6 transcription repressor		604
CA9	carbonic anhydrase 9		768
CALB2	calbindin 2		794
CALCA	calcitonin related polypeptide alpha		796
CALD1	caldesmon 1		800
CCND1	cyclin D1	CYCLIND1	595
CD1A	CD1a molecule		909
CD2	CD2 molecule		914
CD34	CD34 molecule		947
CD3G	CD3g molecule		917
CD5	CD5 molecule		921
CD79A	CD79a molecule		973
CD99L2	CD99 molecule like 2		83692
CDH1	cadherin 1	E-cadherin	999
CDH17	cadherin 17		1015
CDK4	cyclin dependent kinase 4		1019
CDKN2A	cyclin dependent kinase inhibitor 2A	p16	1029
CDX2	caudal type homeobox 2		1806
CEACAM1	CEA cell adhesion molecule 1		634
CEACAM16	CEA cell adhesion molecule 16, tectorial membrane component		388551
CEACAM18	CEA cell adhesion molecule 18		729767
CEACAM19	CEA cell adhesion molecule 19		56971
CEACAM20	CEA cell adhesion molecule 20		125931
CEACAM21	CEA cell adhesion molecule 21		90273
CEACAM3	CEA cell adhesion molecule 3		1084
CEACAM4	CEA cell adhesion molecule 4		1089
CEACAM5	CEA cell adhesion molecule 5		1048
CEACAM6	CEA cell adhesion molecule 6		4680
CEACAM7	CEA cell adhesion molecule 7		1087
CEACAM8	CEA cell adhesion molecule 8		1088
CGA	glycoprotein hormones, alpha polypeptide		1081
CGB3	chorionic gonadotropin subunit beta 3		1082
CNN1	calponin 1		1264
COQ2	coenzyme Q2, polyprenyltransferase		27235
CPS1	carbamoyl-phosphate synthase 1	HepPar-1 antibody target	1373
CR1	complement C3b/C4b receptor 1 (Knops blood group)		1378
CR2	complement C3d receptor 2		1380
CTNNB1	catenin beta 1		1499

DES	desmin		1674
DSC3	desmocollin 3		1825
ENO2	enolase 2		2026
ERBB2	erb-b2 receptor tyrosine kinase 2	HER2, HER2/neu	2064
ERG	ETS transcription factor ERG		2078
ESR1	estrogen receptor 1	ER	2099
FLI1	Fli-1 proto-oncogene, ETS transcription factor		2313
FOXL2	forkhead box L2		668
FUT4	fucosyltransferase 4	CD15	2526
GATA3	GATA binding protein 3		2625
GPC3	glypican 3		2719
HAVCR1	hepatitis A virus cellular receptor 1		26762
HNF1B	HNF1 homeobox B		6928
IL12B	interleukin 12B		3593
IMP3	IMP U3 small nucleolar ribonucleoprotein 3		55272
INHA	inhibin subunit alpha	Inhibin-alpha	3623
ISL1	ISL LIM homeobox 1		3670
KIT	KIT proto-oncogene, receptor tyrosine kinase		3815
KL	klotho		9365
KLK3	kallikrein related peptidase 3	PSA	354
KRT1	keratin 1		3848
KRT10	keratin 10		3858
KRT14	keratin 14		3861
KRT15	keratin 15		3866
KRT16	keratin 16		3868
KRT17	keratin 17	CK17	3872
KRT18	keratin 18	CK18	3875
KRT19	keratin 19	CK19	3880
KRT2	keratin 2		3849
KRT20	keratin 20	CK20	54474
KRT3	keratin 3		3850
KRT4	keratin 4		3851
KRT5	keratin 5		3852
KRT6A	keratin 6A	CK6A	3853
KRT6B	keratin 6B	CK6B	3854
KRT6C	keratin 6C	CK6C	28688
KRT7	keratin 7	CK7	3855
KRT8	keratin 8	CK8	3856
LIN28A	lin-28 homolog A		79727
LIN28B	lin-28 homolog B		389421
MAGEA2	MAGE family member A2		4101
MDM2	MDM2 proto-oncogene		4193

MIB1	mindbomb E3 ubiquitin protein ligase 1		57534
MITF	melanocyte inducing transcription factor		4286
MLANA	melan-A		2315
MLH1	mutL homolog 1		4292
MME	membrane metalloendopeptidase		4311
MPO	myeloperoxidase		4353
MS4A1	membrane spanning 4-domains A1		931
MSH2	mutS homolog 2		4436
MSH6	mutS homolog 6		2956
MSLN	mesothelin		10232
MTHFR	methylenetetrahydrofolate reductase		4524
MUC1	mucin 1, cell surface associated		4582
MUC2	mucin 2, oligomeric mucus/gel-forming		4583
MUC4	mucin 4, cell surface associated		4585
MUC5AC	mucin 5AC, oligomeric mucus/gel-forming		4586
MYOD1	myogenic differentiation 1		4654
MYOG	myogenin		4656
NANOG	Nanog homeobox		79923
NAPSA	napsin A aspartic peptidase	Napsin A	9476
NCAM1	neural cell adhesion molecule 1	CD56	4684
NCAM2	neural cell adhesion molecule 2		4685
NKX2-2	NK2 homeobox 2		4821
NKX3-1	NK3 homeobox 1		4824
OSCAR	osteoclast associated Ig-like receptor		126014
PAX2	paired box 2		5076
PAX5	paired box 5		5079
PAX8	paired box 8		7849
PDPN	podoplanin		10630
PDX1	pancreatic and duodenal homeobox 1		3651
PECAM1	platelet and endothelial cell adhesion molecule 1		5175
PGR	progesterone receptor	PR	5241
PIP	prolactin induced protein		5304
PMEL	premelanosome protein (gp100)	GP100, PMEL17, SILV, HMB-45 target	6490
PMS2	PMS1 homolog 2, mismatch repair system component		5395
POU5F1	POU class 5 homeobox 1		5460
PSAP	prosaposin		5660
PTPRC	protein tyrosine phosphatase receptor type C		5788
S100A1	S100 calcium binding protein A1		6271
S100A10	S100 calcium binding protein A10		6281
S100A11	S100 calcium binding protein A11		6282

S100A12	S100 calcium binding protein A12		6283
S100A13	S100 calcium binding protein A13		6284
S100A14	S100 calcium binding protein A14		57402
S100A16	S100 calcium binding protein A16		140576
S100A2	S100 calcium binding protein A2		6273
S100A4	S100 calcium binding protein A4		6275
S100A5	S100 calcium binding protein A5		6276
S100A6	S100 calcium binding protein A6		6277
S100A7	S100 calcium binding protein A7		6278
S100A7A	S100 calcium binding protein A7A		338324
S100A7L2	S100 calcium binding protein A7 like 2		645922
S100A8	S100 calcium binding protein A8		6279
S100A9	S100 calcium binding protein A9		6280
S100B	S100 calcium binding protein B		6285
S100P	S100 calcium binding protein P		6286
S100PBP	S100P binding protein		64766
S100Z	S100 calcium binding protein Z		170591
SALL4	spalt like transcription factor 4		57167
SATB2	SATB homeobox 2		23314
SDC1	syndecan 1	CD138	6382
SERPINA1	serpin family A member 1	α1-antitrypsin, antitrypsin	5265
SERPINB5	serpin family B member 5	PI5, maspin	5268
SF1	splicing factor 1		7536
SFTPA1	surfactant protein A1		653509
SMAD4	SMAD family member 4		4089
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1		6598
SMN1	survival of motor neuron 1, telomeric		6606
SOX2	SRY-box transcription factor 2		6657
SPN	sialophorin		6693
SYP	synaptophysin		6855
TFE3	transcription factor binding to IGHM enhancer 3		7030
TFF1	trefoil factor 1		7031
TFF3	trefoil factor 3		7033
TG	thyroglobulin		7038
TLE1	TLE family member 1, transcriptional corepressor		7088
TMPRSS2	transmembrane serine protease 2		7113
TNFRSF8	TNF receptor superfamily member 8		943
TP63	tumor protein p63	P63	8626
TPM1	tropomyosin 1		7168
TPM2	tropomyosin 2		7169
TPM3	tropomyosin 3		7170

TPM4	tropomyosin 4		7171
TPSAB1	tryptase alpha/beta 1		7177
TTF1	transcription termination factor 1		7270
UPK2	uroplakin 2	UPII	7379
UPK3A	uroplakin 3A		7380
UPK3B	uroplakin 3B		105375355
VHL	von Hippel-Lindau tumor suppressor		7428
VIL1	villin 1	Villin	7429
VIM	vimentin		7431
WT1	WT1 transcription factor		7490

In some embodiments, data for the chosen features, here transcript expression levels, is used to train the prediction models for the attributes of interest, e.g., as in **FIG. 4B 412-414** or **FIG. 4A 403-405**. Although we rationalized selection of the group of transcripts in **Table 117** by tissue classification based on IHC protein expression, we did not replicate classification schemes based on the protein – tissue correlations. Rather, expression data for the RNA transcripts in **Table 117** were used to build machine learning models to predict tissue characteristics. The machine learning algorithms selected the appropriate transcript features during the training phase. The transcript INSM1 (Full name: INSM transcriptional repressor 1; NCBI Gene ID: 3642) was also used as a verification for neuroendocrine tumors but was not included when training the machine learning framework. *See, e.g.,* Mukhopadhyay, M *et al.*, Insulinoma-associated protein 1 (INSM1) is a sensitive and highly specific marker of neuroendocrine differentiation in primary lung neoplasms: an immunohistochemical study of 345 cases, including 292 whole-tissue sections, Modern Pathology (2019) 32:100–109.

The models were trained as described herein. See, e.g., **FIGs. 4A-B** and related discussion; **Examples 2-3**. The training was performed using all transcript features in **Table 117**. Features of most importance for each prediction of the attributes cancer type, organ group, and histology are listed in **Tables 118-120**, respectively. In some embodiments, the prediction models for individual attributes use features found to contribute most to the predictions. In **Tables 118-120**, the "importance" values represent the relative contribution of each corresponding transcript to the noted classification model. Higher values indicate greater importance. Abbreviations in **Table 118** include ACC (adrenal cortical carcinoma), BDC (bile duct, cholangiocarcinoma), BC (breast cancer), Cerv (cervix carcinoma), Colon (colon carcinoma), EC (endometrium carcinoma), GC (gastroesophageal carcinoma), KRCC (kidney renal cell carcinoma), LHC (liver hepatocellular carcinoma), Lung (lung carcinoma), Mel (melanoma), Men (meningioma), Merk (Merkel), Neu (neuroendocrine), OGCT (ovary granulosa cell tumor), OFP (ovary, fallopian, peritoneum), Panc (pancreas carcinoma), PM (pleural mesothelioma), PA (prostate adenocarcinoma), Ret (retroperitoneum), SP (salivary and parotid), SIA (small intestine adenocarcinoma), SCC (squamous cell carcinoma), TC (thyroid carcinoma), UC (urothelial

carcinoma), Ute (uterus). Abbreviations in **Table 119** include AG (adrenal gland), Bla (bladder), Br (breast), Gast (Gastroesophageal), HFN (head, face or neck, NOS), Kid (kidney), LGD (liver, gallbladder, ducts), Panc (pancreas), Pros (prostate), SI (small intestine), Thy (thyroid). **Table 119** omits leading zeros before the decimal for brevity. Abbreviations in **Table 120** include Adeno (adenocarcinoma), ACyC (Adenoid cystic carcinoma), AC (adenosquamous carcinoma), ACC (adrenal cortical carcinoma), Astro (astrocytoma), Carc (carcinoma), CS (carcinosarcoma), Chol (cholangiocarcinoma), CCC (clear cell carcinoma), DCIS (ductal carcinoma in situ), GBM (glioblastoma), GIST (gastrointestinal stromal tumor), Gli (glioma), GCT (granulosa cell tumor), ILC (infiltrating lobular carcinoma), Lei (leiomyosarcoma), Lipo (liposarcoma), Mel (melanoma), Men (meningioma), Merk (Merkel cell carcinoma), Meso (mesothelioma), Neuro (neuroendocrine), NSCC (non-small cell carcinoma), Oligo (oligodendroglioma), Sarc (sarcoma), SerC (sarcomatoid carcinoma), SCC (small cell carcinoma), Sq (squamous).

Table 118 – Importance of RNA Transcripts used to Classify Cancer/Disease Type

Transcript	ACC	BDC	BC	CNS	Cerv	Colon	EC	GIST	GC	KRCC	LHC	Lung	Mel	Men
ACVRL1	0.0004	0.1199	0.0248	0.0000	0.0040	0.0230	0.2195	0.0976	0.0108	0.0470	0.0000	0.0301	0.1601	0.0000
AFP	0.0000	0.0571	0.0321	0.0019	0.0517	0.1342	0.1118	0.0000	0.0883	0.0000	0.3803	0.0209	0.0000	0.0000
ALPP	0.0000	0.0609	0.1331	0.0000	0.0828	0.1160	0.1729	0.0000	0.0256	0.0107	0.0000	0.0050	0.0000	0.0000
AMACR	0.0000	0.0712	0.1790	0.0000	0.0459	0.0142	0.0219	0.0000	0.0882	0.2849	0.0154	0.0116	0.0005	0.0000
ANKRD30A	0.0000	0.0758	0.7886	0.0000	0.1003	0.0019	0.0370	0.0000	0.0189	0.0000	0.0019	0.0762	0.0000	0.0000
ANO1	0.0000	0.3746	0.0930	0.5582	0.0019	0.0349	0.2271	0.4210	0.3991	0.0424	0.0000	0.1994	0.0000	0.3991
ARG1	0.0282	0.0159	0.1184	0.0000	0.0283	0.1287	0.2650	0.0000	0.0299	0.0073	0.0668	0.1887	0.0371	0.0000
AR	0.0000	0.2429	0.1239	0.0020	0.0000	0.0612	0.1165	0.0000	0.4879	0.0346	0.0000	0.3547	0.0242	0.0099
BCL2	0.0000	0.0847	0.0213	0.0169	0.0092	0.2816	0.1625	0.0000	0.1195	0.0038	0.0000	0.0585	0.0000	0.0000
BCL6	0.0000	0.1002	0.0250	0.0000	0.0231	0.0347	0.2506	0.0000	0.1025	0.2594	0.2069	0.0962	0.0625	0.0211
CA9	0.0000	0.1177	0.1194	0.0102	0.1060	0.0113	0.0136	0.0000	0.0518	0.1982	0.0000	0.0247	0.0073	0.0000
CALB2	0.0706	0.1980	0.1016	0.0000	0.0087	0.0390	0.0345	0.0000	0.0509	0.0000	0.0000	0.0571	0.0071	0.0000
CALCA	0.0000	0.0940	0.0409	0.0000	0.0054	0.0173	0.0291	0.0000	0.0737	0.1475	0.0000	0.1323	0.0000	0.0000
CALD1	0.0000	0.1236	0.0360	0.0251	0.0086	0.0145	0.4457	0.0000	0.0079	0.0959	0.0005	0.0906	0.0008	0.0068
CCND1	0.0000	0.0379	0.1132	0.0089	0.3474	0.0401	0.1933	0.0000	0.0121	0.0296	0.0166	0.0612	0.0949	0.0549
CD1A	0.0000	0.0580	0.1178	0.0000	0.0814	0.0362	0.0680	0.0000	0.2925	0.0000	0.0054	0.0327	0.0000	0.0000
CD2	0.0000	0.0484	0.0221	0.0393	0.0715	0.0662	0.0299	0.0000	0.0187	0.0000	0.0000	0.0615	0.0434	0.0194
CD34	0.0306	0.0250	0.0079	0.0000	0.0026	0.1113	0.1006	0.0000	0.2945	0.1061	0.1227	0.0378	0.0000	0.0000
CD3G	0.0000	0.0054	0.0465	0.0391	0.2238	0.0182	0.0326	0.0000	0.0453	0.0021	0.0246	0.0313	0.0247	0.0000
CD5	0.0000	0.1825	0.1934	0.0000	0.0554	0.1106	0.0434	0.0000	0.0416	0.0000	0.0071	0.0879	0.0004	0.0777
CD79A	0.0000	0.0582	0.1118	0.0000	0.2401	0.0662	0.0711	0.0000	0.0238	0.0046	0.0000	0.0242	0.0113	0.0000
CD99L2	0.0000	0.0427	0.1201	0.0579	0.0221	0.0134	0.0553	0.0000	0.0594	0.0000	0.0022	0.2901	0.0064	0.0000
CDH17	0.0000	0.0835	0.0034	0.0000	0.0018	0.4591	0.0785	0.0000	0.0357	0.0070	0.0055	0.1139	0.0000	0.0000
CDH1	0.0771	0.0161	0.1336	0.0544	0.0152	0.0166	0.0474	0.0320	0.2661	0.6591	0.0000	0.0191	0.0000	0.0563
CDK4	0.0000	0.1843	0.0275	0.0000	0.1197	0.0310	0.0171	0.0000	0.0430	0.0037	0.0000	0.1193	0.0000	0.0000
CDKN2A	0.0000	0.0972	0.1531	0.0093	0.3759	0.1270	0.1142	0.0000	0.0196	0.5109	0.0000	0.1210	0.1606	0.0086

CDX2	0.0000	0.0206	0.1544	0.0000	0.0308	1.6534	0.0274	0.0000	0.7635	0.0000	0.0000	0.0740	0.0000	0.0000
CEACAM16	0.0000	0.0676	0.1928	0.0000	0.0755	0.0727	0.2698	0.0000	0.0194	0.0000	0.5075	0.1828	0.0000	0.0000
CEACAM18	0.0000	0.0365	0.1524	0.0000	0.0000	0.2429	0.0217	0.0000	0.0788	0.0000	0.0000	0.0262	0.0000	0.0000
CEACAM19	0.0000	0.0464	0.0252	0.0038	0.1472	0.0772	0.1867	0.0000	0.1050	0.0656	0.0109	0.0851	0.0677	0.0000
CEACAM1	0.0000	0.0654	0.0122	0.1894	0.0085	0.0939	0.1046	0.0000	0.0521	0.0363	0.0389	0.2672	0.1125	0.2127
CEACAM20	0.0000	0.0059	0.0003	0.0000	0.0142	0.3682	0.0789	0.0000	0.0508	0.0000	0.1473	0.0159	0.0020	0.0000
CEACAM21	0.0000	0.0538	0.0382	0.0000	0.1321	0.0130	0.0591	0.0000	0.0035	0.0000	0.0000	0.0286	0.0000	0.0000
CEACAM3	0.0000	0.0270	0.0197	0.0000	0.0000	0.0169	0.0405	0.0000	0.0582	0.0000	0.0018	0.0340	0.0066	0.0000
CEACAM4	0.0000	0.0434	0.2064	0.0000	0.2952	0.0293	0.0162	0.0000	0.0622	0.0033	0.0000	0.0449	0.0149	0.0000
CEACAM5	0.0000	0.0342	0.0884	0.0016	0.0573	0.4906	0.0259	0.0000	0.0291	0.0783	0.2582	0.0113	0.0000	0.0061
CEACAM6	0.0000	0.0119	0.0048	0.0000	0.0065	0.0995	0.1930	0.0000	0.3695	0.0202	0.0160	0.4092	0.0020	0.0000
CEACAM7	0.0000	0.1211	0.1673	0.0000	0.1162	0.0211	0.0715	0.0000	0.0231	0.0023	0.0000	0.5022	0.0000	0.0000
CEACAM8	0.0000	0.0331	0.0057	0.0000	0.0361	0.0392	0.0932	0.0000	0.0093	0.0311	0.0078	0.0264	0.0046	0.0000
CGA	0.0000	0.0561	0.0075	0.0000	0.0083	0.0392	0.1350	0.0000	0.0293	0.0000	0.0000	0.0149	0.0000	0.0039
CGB3	0.0000	0.1212	0.0666	0.0987	0.0144	0.0253	0.0389	0.0000	0.1087	0.0064	0.0000	0.0295	0.0063	0.0000
CNN1	0.0000	0.2455	0.1790	0.0000	0.0246	0.1649	0.1165	0.0000	0.0061	0.0043	0.0000	0.1622	0.0000	0.0000
COQ2	0.0000	0.1545	0.0434	0.0000	0.0460	0.0509	0.0186	0.0000	0.0911	0.0454	0.0000	0.0338	0.0000	0.0000
CPS1	0.0000	0.0376	0.0288	0.0000	0.0337	0.2157	0.0971	0.0000	0.0678	0.1034	0.0030	0.1469	0.0815	0.0000
CR1	0.0000	0.0067	0.0219	0.0000	0.0680	0.1208	0.0306	0.0000	0.0547	0.0000	0.0000	0.0552	0.0160	0.0017
CR2	0.0000	0.0702	0.0070	0.0000	0.0613	0.1518	0.1308	0.0000	0.0320	0.0000	0.0010	0.0254	0.0081	0.0000
CTNNB1	0.0000	0.0503	0.0477	0.0027	0.1224	0.0602	0.0430	0.0000	0.1372	0.0000	0.0000	0.1204	0.0081	0.0000
DES	0.0000	0.1269	0.2030	0.0019	0.0049	0.0554	0.3589	0.0000	0.2451	0.0278	0.0047	0.0532	0.0000	0.0000
DSC3	0.0000	0.0947	0.0479	0.0240	0.2025	0.1638	0.2982	0.0000	0.0491	0.0146	0.1840	0.0709	0.0055	0.0174
ENO2	0.0000	0.2213	0.1018	0.0484	0.0245	0.1621	0.0513	0.0025	0.3330	0.1448	0.0021	0.0740	0.0155	0.0000
ERBB2	0.0000	0.0523	0.0108	0.1156	0.0067	0.0140	0.1281	0.0145	0.0472	0.0674	0.1205	0.1194	0.0050	0.0021
ERG	0.0000	0.0378	0.0427	0.0071	0.1084	0.1028	0.0444	0.0000	0.0110	0.0037	0.0097	0.0424	0.0000	0.0000
ESR1	0.0000	0.4155	0.0774	0.0000	0.6968	0.1522	0.5633	0.0000	0.0694	0.0454	0.0191	0.1661	0.0141	0.0000
FLI1	0.0003	0.0191	0.0309	0.0037	0.0111	0.0253	0.3088	0.0000	0.0185	0.0108	0.0000	0.1259	0.0007	0.0000
FOXL2	0.0000	0.0337	0.0212	0.0000	0.1575	0.1196	0.0875	0.0000	0.1158	0.0000	0.0380	0.0138	0.0000	0.0000
FUT4	0.0000	0.0441	0.0859	0.0000	0.2820	0.3326	0.0713	0.0000	0.7653	0.1120	0.0447	0.0897	0.0148	0.0000
GATA3	0.0000	0.1473	1.9751	0.0409	0.0403	0.1323	0.1365	0.0000	0.0156	0.0369	0.0086	0.1119	0.1175	0.0234
GPC3	0.0000	0.0757	0.0184	0.1721	0.0000	0.1183	0.1398	0.0000	0.0291	0.0271	0.1407	0.1804	0.0000	0.0003
HAVCR1	0.0000	0.0760	0.0267	0.0000	0.0102	0.0567	0.0489	0.0000	0.0167	0.4287	0.0121	0.1936	0.0000	0.0000
HNF1B	0.0000	0.9014	0.4113	0.0000	0.0330	0.2249	0.0448	0.0000	0.0365	0.3831	0.0073	0.0741	0.0000	0.0000
IL12B	0.0000	0.0407	0.0351	0.0000	0.0778	0.0270	0.0236	0.0000	0.0367	0.0026	0.0000	0.1886	0.0000	0.0000
IMP3	0.0000	0.0395	0.0232	0.0000	0.0363	0.2060	0.0144	0.0000	0.0197	0.0000	0.0006	0.1069	0.0000	0.0000
INHA	0.1270	0.1763	0.0491	0.0337	0.0644	0.1489	0.1608	0.0000	0.1896	0.0112	0.0000	0.0843	0.0610	0.0769
ISL1	0.0000	0.0894	0.1559	0.0043	0.1671	0.0771	0.0211	0.0000	0.4124	0.0081	0.0187	0.1219	0.0000	0.0000
KIT	0.0000	0.0272	0.1239	0.0000	0.0029	0.0612	0.0580	0.0677	0.1704	0.0761	0.0026	0.1541	0.0000	0.0000
KLK3	0.0000	0.0507	0.0645	0.0000	0.0174	0.1677	0.0545	0.0000	0.0066	0.0558	0.0000	0.0553	0.0000	0.0000
KL	0.0000	0.1828	0.1707	0.0000	0.0316	0.0214	0.0754	0.0000	0.0900	0.3624	0.0000	0.0176	0.0024	0.0000
KRT10	0.0000	0.0200	0.0073	0.0000	0.0214	0.1886	0.0352	0.0000	0.0303	0.0000	0.0076	0.2021	0.0267	0.1797

KRT14	0.0000	0.1351	0.1228	0.0047	0.0079	0.0936	0.1089	0.0000	0.1042	0.0000	0.0000	0.0556	0.0000	0.0000
KRT15	0.0000	0.0453	0.6266	0.0156	0.0438	0.0457	0.0559	0.0000	0.1042	0.0032	0.1799	0.2116	0.0000	0.0000
KRT16	0.0000	0.0358	0.2420	0.0008	0.0467	0.0180	0.0128	0.0000	0.0260	0.0000	0.0792	0.0515	0.0000	0.0452
KRT17	0.0000	0.1331	0.0193	0.0061	0.1592	0.0570	0.0143	0.0008	0.0463	0.0581	0.0004	0.1115	0.0349	0.0000
KRT18	0.0000	0.0201	0.4157	1.0434	0.0172	0.2612	0.0282	0.0000	0.0531	0.0007	0.0831	0.0396	0.0586	0.0000
KRT19	0.0670	0.0128	0.0489	0.3758	0.0000	0.0356	0.0527	0.3005	0.0545	0.0108	0.4374	0.0656	0.5359	0.0000
KRT1	0.0000	0.0148	0.0119	0.0008	0.0177	0.0026	0.0414	0.0000	0.0274	0.0043	0.0037	0.0204	0.0000	0.0000
KRT20	0.0000	0.0344	0.0877	0.0000	0.0826	0.7625	0.0481	0.0000	0.0898	0.0000	0.0031	0.1707	0.0000	0.0000
KRT2	0.0000	0.0212	0.0551	0.0000	0.0544	0.0247	0.0444	0.0000	0.1291	0.0657	0.0000	0.0423	0.0000	0.0000
KRT3	0.0000	0.0490	0.0538	0.0000	0.0224	0.0041	0.0061	0.0000	0.0014	0.0000	0.0000	0.0127	0.0807	0.0000
KRT4	0.0000	0.1454	0.0520	0.0000	0.0932	0.1828	0.0783	0.0000	0.0421	0.0000	0.0024	0.0245	0.0000	0.0000
KRT5	0.0000	0.2816	0.1591	0.0042	0.0038	0.0270	0.3821	0.0000	0.0270	0.0033	0.0000	0.2748	0.0000	0.0000
KRT6A	0.0000	0.0124	0.0774	0.0010	0.0022	0.2649	0.0206	0.0000	0.0639	0.0000	0.0446	0.1030	0.0006	0.0000
KRT6B	0.0000	0.0895	0.2370	0.0000	0.0026	0.3555	0.0083	0.0000	0.0319	0.0084	0.0000	0.0573	0.0007	0.0000
KRT6C	0.0000	0.0171	0.0874	0.0000	0.0809	0.0272	0.0616	0.0000	0.0422	0.0000	0.0000	0.0705	0.0007	0.0000
KRT7	0.0000	0.2611	0.5100	0.1042	0.0374	1.4166	0.0785	0.0164	0.0742	0.3134	0.0000	0.4525	0.0000	0.0051
KRT8	0.0295	0.1635	0.0546	1.0032	0.0436	0.0185	0.0389	0.2585	0.0500	0.0092	0.0000	0.1172	0.8518	0.4163
LIN28A	0.0000	0.0122	0.0287	0.0000	0.3409	0.0741	0.0268	0.0000	0.0244	0.0000	0.0150	0.0186	0.0975	0.0000
LIN28B	0.0000	0.0373	0.0432	0.0021	0.0000	0.0228	0.4217	0.0000	0.0021	0.0000	0.0000	0.0462	0.0000	0.0000
MAGEA2	0.0000	0.1055	0.0066	0.0000	0.0013	0.0025	0.0102	0.0000	0.0554	0.0000	0.0000	0.0529	0.0123	0.0126
MDM2	0.0000	0.1220	0.2848	0.0019	0.2589	0.0265	0.1140	0.0000	0.0116	0.1901	0.0000	0.0210	0.0000	0.0471
MIB1	0.1185	0.0235	0.1144	0.0000	0.0718	0.0828	0.0719	0.0000	0.0092	0.0410	0.0000	0.0132	0.0000	0.0000
MITF	0.0000	0.0981	0.0159	0.0053	0.1067	0.0571	0.2480	0.0000	0.0311	0.0005	0.0040	0.1927	0.2270	0.0108
MLANA	0.0000	0.0948	0.0481	0.0132	0.1234	0.0678	0.0679	0.0000	0.0640	0.0174	0.0000	0.1531	0.4586	0.0000
MLH1	0.0000	0.0557	0.0199	0.0000	0.0783	0.2382	0.2500	0.0000	0.0131	0.0100	0.0000	0.0699	0.0000	0.0000
MME	0.0000	0.0823	0.0803	0.0000	0.1093	0.1141	0.0662	0.0000	0.0227	0.0685	0.0000	0.0496	0.0000	0.0000
MPO	0.0000	0.0714	0.0100	0.0000	0.0560	0.0020	0.0441	0.0000	0.0248	0.0075	0.0000	0.0580	0.0000	0.0165
MS4A1	0.0000	0.1279	0.0470	0.0000	0.0626	0.0565	0.0126	0.0000	0.0050	0.0113	0.0033	0.1088	0.1585	0.0000
MSH2	0.0000	0.0366	0.0268	0.2361	0.0199	0.0610	0.0421	0.0000	0.0532	0.0544	0.2183	0.0431	0.0000	0.2008
MSH6	0.0000	0.0193	0.0137	0.0059	0.0148	0.0060	0.0889	0.0000	0.0919	0.0000	0.0033	0.0740	0.0065	0.0000
MSLN	0.0000	0.0536	0.0586	0.0000	0.0148	0.1393	0.1502	0.0000	0.0249	0.1571	0.0576	0.1468	0.0000	0.0094
MTHFR	0.0000	0.0140	0.2133	0.0000	0.0400	0.0393	0.0463	0.0000	0.1256	0.0406	0.0027	0.0453	0.0095	0.0000
MUC1	0.0535	0.0929	0.0032	0.0061	0.0649	0.5842	0.0903	0.2777	0.1772	0.2964	0.1388	0.2699	0.5180	0.0000
MUC2	0.0000	0.0219	0.0125	0.0000	0.2677	1.1616	0.0161	0.0000	0.0173	0.0018	0.0000	0.0526	0.0000	0.0000
MUC4	0.0000	0.3099	0.4270	0.0035	0.1352	0.1016	0.1268	0.0000	0.2198	0.0443	0.3336	0.2033	0.0000	0.0147
MUC5AC	0.0000	0.1903	0.2662	0.0000	0.1500	0.0143	0.1385	0.0000	0.5114	0.0777	0.0118	0.1097	0.0000	0.0000
MYOD1	0.0000	0.0345	0.0064	0.0000	0.0359	0.0120	0.1814	0.0000	0.0446	0.0000	0.0276	0.0376	0.0035	0.0000
MYOG	0.0000	0.0217	0.0755	0.0059	0.0020	0.0333	0.0947	0.0000	0.1759	0.0000	0.0011	0.0228	0.0997	0.0000
NANOG	0.0000	0.0207	0.0311	0.0079	0.0975	0.0155	0.1539	0.0000	0.1042	0.0055	0.0000	0.0586	0.0000	0.0000
NAPSA	0.0000	0.0940	0.0983	0.0102	0.0449	0.0454	0.3890	0.0000	0.3190	0.0000	0.0000	1.0851	0.0042	0.0022
NCAM1	0.0161	0.0385	0.0786	0.5217	0.2480	0.0031	0.0604	0.0000	0.0083	0.0022	0.0000	0.0437	0.0660	0.0000
NCAM2	0.0294	0.1541	0.0382	0.0000	0.0480	0.2094	0.0676	0.0000	0.4229	0.0000	0.0000	0.1625	0.0466	0.0000

NKX2-2	0.0000	0.2202	0.0439	0.4077	0.0319	0.0222	0.1920	0.0000	0.0088	0.0000	0.0000	0.0601	0.0310	0.0000
NKX3-1	0.0715	0.1334	0.0299	0.0000	0.0489	0.2269	0.0418	0.0000	0.1014	0.0067	0.0048	0.1436	0.0000	0.0000
OSCAR	0.0000	0.0762	0.0949	0.0396	0.0145	0.1087	0.0906	0.0000	0.0190	0.0000	0.0000	0.0515	0.0000	0.0000
PAX2	0.0000	0.0091	0.0384	0.0000	0.0227	0.0384	0.1052	0.0000	0.0748	0.2851	0.0000	0.1045	0.0000	0.0000
PAX5	0.0000	0.0863	0.0813	0.0000	0.0260	0.0289	0.2066	0.0000	0.0915	0.0000	0.0000	0.0110	0.0256	0.0023
PAX8	0.0000	0.1905	0.4312	0.0000	0.1539	0.1731	1.6954	0.0000	0.3831	0.7741	0.0000	0.3878	0.0006	0.0082
PDPN	0.0000	0.0141	0.1592	0.4476	0.0048	0.0262	0.2675	0.0000	0.1346	0.0000	0.0000	0.0637	0.1012	0.0017
PDX1	0.0000	0.0993	0.0582	0.0000	0.0847	0.0691	0.0120	0.0000	0.1910	0.0000	0.0202	0.1244	0.0000	0.0000
PECAM1	0.0000	0.1201	0.1237	0.0000	0.0051	0.0367	0.0310	0.0000	0.1697	0.0504	0.0000	0.0164	0.0011	0.0000
PGR	0.0000	0.0619	0.1286	0.0000	0.3198	0.1078	0.5994	0.0000	0.0301	0.0000	0.0032	0.0448	0.0020	0.1911
PIP	0.0000	0.0909	0.3383	0.0000	0.0293	0.0208	0.1348	0.0000	0.0375	0.0072	0.0026	0.0842	0.0000	0.0000
PMEL	0.0000	0.0805	0.2466	0.0000	0.2023	0.0290	0.0776	0.0000	0.2113	0.0038	0.0297	0.0551	0.6758	0.0000
PMS2	0.0000	0.0404	0.0188	0.0000	0.0266	0.0101	0.0546	0.0000	0.1613	0.0000	0.0155	0.0196	0.0020	0.0000
POU5F1	0.0000	0.1802	0.0734	0.0000	0.0068	0.0667	0.0884	0.0000	0.0566	0.2956	0.1149	0.1029	0.1426	0.0000
PSAP	0.0153	0.2165	0.0039	0.0000	0.2756	0.0281	0.0901	0.0000	0.0982	0.0120	0.0000	0.0394	0.0000	0.0000
PTPRC	0.0000	0.0430	0.0243	0.0185	0.0000	0.0497	0.1087	0.0000	0.0321	0.0060	0.0000	0.0206	0.0055	0.0000
S100A10	0.0000	0.0535	0.1032	0.0048	0.1155	0.0099	0.0497	0.0000	0.0309	0.0598	0.0000	0.4226	0.0000	0.0067
S100A11	0.0000	0.0266	0.0222	0.2679	0.0665	0.0535	0.1391	0.0000	0.2227	0.0069	0.0095	0.0586	0.0137	0.0000
S100A12	0.0000	0.0118	0.1145	0.0000	0.1333	0.1050	0.0291	0.0000	0.1106	0.0000	0.0010	0.0800	0.0000	0.0000
S100A13	0.0000	0.0531	0.1346	0.0000	0.2296	0.0142	0.0090	0.0000	0.3664	0.2409	0.0097	0.3093	0.2785	0.0000
S100A14	0.0000	0.1249	0.2299	0.2962	0.0198	0.2156	0.0664	0.0000	0.0307	0.4307	0.0000	0.0213	0.3043	0.2359
S100A16	0.0000	0.0258	0.0146	0.0024	0.0054	0.0070	0.2035	0.0046	0.0380	0.0000	0.0000	0.0073	0.0000	0.0000
S100A1	0.0000	0.0617	0.3432	0.2453	0.1060	0.0155	0.0530	0.0000	0.0570	0.0082	0.0002	0.3935	0.2097	0.0000
S100A2	0.0000	0.2901	0.4465	0.0903	0.1006	0.1114	0.1342	0.0180	0.1053	0.0000	0.0680	0.0470	0.0117	0.2339
S100A4	0.0000	0.0947	0.0464	0.0483	0.0028	0.0979	0.0217	0.0000	0.0110	0.0032	0.0000	0.0296	0.0153	0.0183
S100A5	0.0464	0.0693	0.0477	0.0241	0.0479	0.0165	0.1167	0.0000	0.1373	0.0225	0.0000	0.0717	0.0227	0.0018
S100A6	0.0000	0.2004	0.2369	0.0000	0.1529	0.4517	0.3725	0.0000	0.0480	0.0000	0.1595	0.1261	0.0000	0.0153
S100A7A	0.0000	0.1159	0.0065	0.0000	0.0334	0.0696	0.0677	0.0000	0.0632	0.0000	0.0061	0.0250	0.0000	0.0000
S100A7L2	0.0000	0.0094	0.1057	0.0000	0.0290	0.0075	0.0166	0.0000	0.0077	0.0000	0.0000	0.0041	0.0000	0.0000
S100A7	0.0000	0.0148	0.0100	0.0000	0.0419	0.0515	0.1609	0.0000	0.2783	0.0000	0.0000	0.1521	0.0007	0.0000
S100A8	0.0000	0.0450	0.0116	0.0000	0.0080	0.0427	0.0198	0.0000	0.0256	0.0018	0.0029	0.0366	0.0000	0.0175
S100A9	0.0000	0.2209	0.0939	0.0000	0.0765	0.0773	0.2121	0.0020	0.2167	0.0000	0.0000	0.0603	0.0010	0.0322
S100B	0.0000	0.0517	0.0971	1.0716	0.2872	0.0174	0.0168	0.0000	0.3090	0.0480	0.0154	0.0283	1.2799	0.0000
S100PBP	0.0000	0.1183	0.0459	0.0002	0.0442	0.0178	0.0391	0.0000	0.0150	0.0044	0.0000	0.1418	0.0161	0.0000
S100P	0.0000	0.0464	0.1935	0.0000	0.0458	0.0154	0.2953	0.0000	0.0415	0.4360	0.0020	0.0287	0.1176	0.0031
S100Z	0.0000	0.0392	0.0013	0.0061	0.0019	0.0148	0.0261	0.0000	0.0333	0.0678	0.0000	0.1288	0.0000	0.0000
SALL4	0.0000	0.1235	0.1416	0.0314	0.1017	0.0255	0.1639	0.0000	0.1536	0.1856	0.0029	0.0184	0.0000	0.0155
SATB2	0.0000	0.2178	0.0032	0.0000	0.2461	0.5521	0.0431	0.0000	0.1301	0.0017	0.0588	0.0746	0.1050	0.0000
SDC1	0.0000	0.0448	0.0625	0.0024	0.0561	0.0818	0.0334	0.4088	0.0614	0.0000	0.0000	0.1180	0.0000	0.6138
SERPINA1	0.0158	0.5546	0.1814	0.0000	0.0515	0.0237	0.0520	0.0000	0.0987	0.0859	0.7962	0.0604	0.0000	0.0000
SERPINB5	0.0000	0.0840	0.2329	0.0000	0.0082	0.1128	0.0562	0.0000	0.5175	0.0280	0.0141	0.1436	0.0000	0.0018
SF1	0.0000	0.0445	0.0725	0.0000	0.0242	0.0260	0.0164	0.0000	0.0592	0.1009	0.0067	0.1398	0.0000	0.0015

SFTPA1	0.0000	0.1572	0.0461	0.0000	0.0110	0.0188	0.0331	0.0000	0.0953	0.0151	0.0000	0.2640	0.0028	0.0000
SMAD4	0.0000	0.0423	0.0369	0.0000	0.0093	0.0888	0.0668	0.0000	0.0800	0.0033	0.0081	0.0067	0.0000	0.0000
SMARCB1	0.0000	0.0753	0.0065	0.0325	0.3181	0.0016	0.2247	0.0000	0.0813	0.0096	0.0063	0.1316	0.0000	0.0333
SMN1	0.0000	0.1124	0.0081	0.0027	0.0768	0.0181	0.1144	0.0000	0.0492	0.0082	0.0000	0.0576	0.0000	0.0000
SOX2	0.0003	0.3363	0.3114	0.7907	0.0563	0.1969	0.0355	0.0000	0.3802	0.0220	0.0161	0.5792	0.0062	0.0000
SPN	0.0000	0.0141	0.0546	0.0000	0.0030	0.0777	0.0667	0.0000	0.2709	0.0000	0.0006	0.0173	0.0000	0.0398
SYP	0.1109	0.0444	0.0986	0.0000	0.0074	0.0356	0.0852	0.0000	0.1467	0.1603	0.0000	0.0204	0.0046	0.0000
TFE3	0.0000	0.1387	0.1111	0.0000	0.0183	0.0067	0.0179	0.0000	0.0119	0.0340	0.0000	0.0313	0.0034	0.0000
TFF1	0.0000	0.1821	0.2434	0.0000	0.0033	0.2416	0.0509	0.0000	0.4452	0.0000	0.0229	0.2230	0.0000	0.0000
TFF3	0.0000	0.0476	0.1606	0.0000	0.0381	0.3417	0.1866	0.0000	0.4172	0.0689	0.0000	0.0481	0.0021	0.0000
TG	0.0279	0.1321	0.0160	0.1140	0.0092	0.0808	0.0674	0.0000	0.0637	0.0481	0.0000	0.1287	0.0000	0.0008
TLE1	0.0000	0.1445	0.0225	0.0018	0.0051	0.0395	0.2590	0.0000	0.0294	0.0695	0.0000	0.1319	0.0032	0.0000
TMPRSS2	0.0297	0.1909	0.0829	0.0430	0.0078	0.1968	0.0803	0.0000	0.2937	0.0505	0.0000	0.2302	0.0000	0.0000
TNFRSF8	0.0004	0.0265	0.1215	0.0000	0.2457	0.0337	0.0043	0.0000	0.0157	0.0005	0.0054	0.1232	0.0020	0.0000
TP63	0.0000	0.0365	0.1117	0.0087	0.1018	0.0123	0.0739	0.0000	0.0123	0.0054	0.0000	0.0642	0.1038	0.1028
TPM1	0.0000	0.1078	0.0858	0.0045	0.0382	0.0673	0.0464	0.0000	0.2065	0.0011	0.0000	0.1372	0.1401	0.0021
TPM2	0.0000	0.0575	0.0205	0.0050	0.1451	0.0259	0.0845	0.0000	0.1216	0.0090	0.0149	0.0342	0.0000	0.0000
TPM3	0.0120	0.0484	0.0228	0.0048	0.0748	0.0085	0.0712	0.0000	0.0092	0.0519	0.0000	0.1855	0.0091	0.0082
TPM4	0.0000	0.0822	0.0866	0.0000	0.0337	0.0916	0.0518	0.0000	0.0468	0.0411	0.0549	0.1722	0.0000	0.0000
TPSAB1	0.0000	0.1863	0.0758	0.0028	0.2121	0.1570	0.0613	0.0018	0.3180	0.1164	0.0000	0.0876	0.0000	0.0000
TTF1	0.0000	0.0503	0.0094	0.0812	0.1321	0.0279	0.1320	0.0000	0.1492	0.0803	0.0215	0.0727	0.0215	0.0000
UPK2	0.0000	0.0412	0.0281	0.0222	0.1078	0.1170	0.0764	0.0000	0.1224	0.0000	0.0000	0.0776	0.0000	0.0000
UPK3A	0.0000	0.0213	0.1437	0.0017	0.0078	0.0162	0.2065	0.0000	0.0446	0.0000	0.0698	0.0076	0.1314	0.0000
UPK3B	0.0000	0.1889	0.2206	0.0169	0.1160	0.0398	0.0594	0.0000	0.0467	0.0148	0.0042	0.1143	0.0036	0.0000
VHL	0.0003	0.0806	0.0534	0.0000	0.2247	0.0285	0.4873	0.0000	0.0736	0.2955	0.0000	0.3369	0.0000	0.0067
VIL1	0.0000	0.5994	0.0240	0.0000	0.0848	0.5227	0.0238	0.0000	0.3881	0.0064	0.1221	0.0326	0.0682	0.0000
VIM	0.0000	0.0188	0.0328	0.0000	0.0033	0.0468	0.0369	0.0000	0.0438	0.0765	0.0000	0.0137	0.1803	0.2430
WT1	0.0000	0.0811	0.0466	0.0160	0.0391	0.0392	0.2561	0.0000	0.0696	0.0411	0.0000	0.1748	0.0000	0.0216

Table 118 *continued*

Transcript	Merk	Neu	OGCT	OFP	Panc	PM	PA	Ret	SP	SIA	SCC	TC	UC	Ute
ACVRL1	0.0000	0.0000	0.0000	0.2065	0.0367	0.0000	0.0000	0.0022	0.0000	0.0096	0.0034	0.0000	0.0587	0.0100
AFP	0.0000	0.0047	0.0000	0.0347	0.0163	0.0000	0.0000	0.0346	0.0000	0.0633	0.0672	0.0000	0.0249	0.0000
ALPP	0.0000	0.0000	0.0000	0.2427	0.0571	0.0000	0.0214	0.0000	0.2317	0.1172	0.0751	0.0000	0.0233	0.0000
AMACR	0.0000	0.0028	0.0033	0.1114	0.2357	0.0008	0.5918	0.0000	0.0000	0.0164	0.0335	0.0044	0.0899	0.0025
ANKRD30A	0.0000	0.0061	0.0000	0.0726	0.1040	0.0000	0.0000	0.0000	0.0064	0.0118	0.0134	0.0000	0.0109	0.0019
ANO1	0.0000	0.0183	0.0000	0.1417	0.7039	0.0000	0.0177	0.0074	0.1828	0.0138	0.1547	0.0052	0.1598	0.0055
ARG1	0.0000	0.1080	0.0000	0.1220	0.2156	0.0000	0.0000	0.0497	0.1198	0.2540	0.0613	0.2657	0.0133	0.0300
AR	0.0000	0.0181	0.0000	0.1520	0.0692	0.0000	0.1169	0.1206	0.0000	0.1860	0.4215	0.0031	0.0096	0.0465
BCL2	0.0000	0.0000	0.0000	0.0560	0.0404	0.0000	0.0140	0.0014	0.0321	0.0398	0.0403	0.0014	0.0029	0.0091
BCL6	0.0000	0.0100	0.0000	0.0155	0.0300	0.0027	0.0718	0.0330	0.0000	0.0157	0.0300	0.0032	0.0671	0.0623
CA9	0.0013	0.0612	0.0000	0.1736	0.0732	0.0321	0.0211	0.0000	0.0098	0.1940	0.0569	0.0237	0.0861	0.0000

CALB2	0.0000	0.0035	0.0000	0.0618	0.3098	0.5246	0.0076	0.0156	0.1907	0.1585	0.0587	0.2775	0.3746	0.0372
CALCA	0.0000	0.0206	0.0018	0.1032	0.0794	0.0000	0.0050	0.0015	0.0028	0.0181	0.1741	0.0000	0.0055	0.0000
CALD1	0.0000	0.0438	0.0000	0.0481	0.0228	0.0000	0.0002	0.0166	0.0000	0.0237	0.0778	0.0000	0.0352	0.0325
CCND1	0.0000	0.0316	0.0000	0.1941	0.0634	0.0000	0.0000	0.0017	0.0056	0.0445	0.0409	0.0799	0.0752	0.0000
CD1A	0.0000	0.0006	0.0000	0.0712	0.1698	0.0000	0.0036	0.0000	0.0000	0.0480	0.1672	0.0047	0.0610	0.0116
CD2	0.0000	0.0198	0.0000	0.0205	0.0681	0.0000	0.0032	0.0000	0.0040	0.0202	0.0112	0.0000	0.2658	0.0909
CD34	0.0000	0.0069	0.0000	0.0231	0.1297	0.0000	0.1084	0.2570	0.0005	0.0463	0.1436	0.0016	0.0352	0.0000
CD3G	0.0000	0.0333	0.0000	0.0154	0.0372	0.0000	0.0625	0.0000	0.0000	0.0306	0.4505	0.0077	0.2254	0.0069
CD5	0.0000	0.0224	0.0000	0.0271	0.3262	0.0000	0.0217	0.0035	0.0000	0.2452	0.0437	0.0189	0.1800	0.0177
CD79A	0.0000	0.0002	0.0000	0.0564	0.0607	0.0000	0.0000	0.0203	0.0088	0.0188	0.0938	0.0136	0.0361	0.4022
CD99L2	0.0000	0.0313	0.0000	0.1654	0.0522	0.0000	0.0119	0.0000	0.0000	0.2136	0.0335	0.0302	0.1242	0.0008
CDH17	0.0000	0.0270	0.0000	0.0926	0.1250	0.0000	0.0146	0.0076	0.0081	0.3786	0.0426	0.0000	0.0237	0.0687
CDH1	0.0000	0.0070	0.0000	0.0031	0.0312	0.0113	0.0772	0.1926	0.0074	0.0000	0.0790	0.1070	0.0024	0.1516
CDK4	0.0000	0.0000	0.0000	0.0402	0.0479	0.0000	0.0135	0.0780	0.0060	0.0515	0.1250	0.2140	0.1472	0.0444
CDKN2A	0.0000	0.0678	0.0000	0.0425	0.1363	0.0105	0.0475	0.0113	0.0061	0.1300	0.0548	0.0138	0.1118	0.0069
CDX2	0.0000	0.1367	0.0000	0.0507	0.1207	0.0000	0.0325	0.0176	0.0000	0.0253	0.0662	0.0000	0.0222	0.0000
CEACAM16	0.0000	0.0000	0.0000	0.0865	0.0625	0.0000	0.0025	0.0000	0.1820	0.0526	0.0256	0.0237	0.1766	0.0104
CEACAM18	0.0000	0.0270	0.0000	0.0307	0.1543	0.0000	0.0923	0.0095	0.1035	0.1317	0.0344	0.0488	0.0016	0.0045
CEACAM19	0.0000	0.0018	0.0000	0.1167	0.0660	0.0000	0.0045	0.0212	0.0000	0.0280	0.0753	0.0176	0.0388	0.0097
CEACAM1	0.0000	0.0000	0.0000	0.0246	0.0927	0.1300	0.1096	0.0563	0.0014	0.1391	0.1982	0.0111	0.0651	0.0554
CEACAM20	0.0000	0.0000	0.0000	0.0136	0.0637	0.0000	0.0028	0.0000	0.0000	0.0223	0.0393	0.0000	0.0000	0.0000
CEACAM21	0.0000	0.0000	0.0035	0.1164	0.0118	0.0000	0.1023	0.0000	0.0056	0.0265	0.0104	0.0000	0.0456	0.0000
CEACAM3	0.0000	0.1156	0.0000	0.2474	0.1011	0.0057	0.0373	0.0000	0.0020	0.0944	0.0497	0.0715	0.0567	0.0265
CEACAM4	0.0013	0.1420	0.0000	0.0370	0.0907	0.0000	0.0047	0.0000	0.0000	0.1055	0.0318	0.0463	0.1265	0.0000
CEACAM5	0.0473	0.1210	0.0000	0.2252	0.0651	0.0000	0.0792	0.0043	0.0000	0.3319	0.0687	0.2028	0.0849	0.0000
CEACAM6	0.0000	0.0044	0.0000	0.1199	0.1324	0.0000	0.1188	0.0062	0.0000	0.0081	0.1136	0.0340	0.1440	0.0000
CEACAM7	0.0000	0.0007	0.0000	0.0685	0.1338	0.0000	0.0011	0.0000	0.0000	0.0537	0.0276	0.0000	0.0443	0.0000
CEACAM8	0.0000	0.0085	0.0000	0.0469	0.0591	0.0000	0.0076	0.0000	0.0007	0.0485	0.1073	0.0000	0.0411	0.0019
CGA	0.0000	0.0132	0.0000	0.0208	0.1910	0.0000	0.0094	0.0076	0.0000	0.0873	0.0434	0.0477	0.0426	0.0000
CGB3	0.0000	0.0000	0.0000	0.0668	0.0102	0.0000	0.1259	0.0071	0.0000	0.1308	0.2238	0.0000	0.0368	0.0503
CNN1	0.0000	0.0065	0.0000	0.0826	0.0256	0.0000	0.1392	0.1850	0.0135	0.1274	0.2971	0.2199	0.1757	0.0918
COQ2	0.0000	0.0049	0.0000	0.0162	0.1601	0.0000	0.0000	0.0000	0.0000	0.0096	0.0972	0.0000	0.0268	0.0062
CPS1	0.0306	0.0010	0.0000	0.1042	0.2197	0.0030	0.1975	0.0849	0.0308	0.1777	0.0843	0.4173	0.4016	0.0000
CR1	0.0175	0.0010	0.0000	0.2003	0.0521	0.0000	0.0238	0.0206	0.0150	0.1249	0.1301	0.0029	0.0314	0.0092
CR2	0.0000	0.0000	0.0000	0.1221	0.1608	0.0000	0.0502	0.0000	0.0052	0.1074	0.0474	0.0000	0.0217	0.0000
CTNNB1	0.0000	0.0038	0.0000	0.0528	0.0185	0.0000	0.0000	0.0000	0.1967	0.0000	0.1189	0.0000	0.3425	0.0000
DES	0.0000	0.0555	0.0000	0.0907	0.2096	0.0000	0.0000	0.0014	0.0022	0.4895	0.1498	0.0000	0.3442	0.5577
DSC3	0.0000	0.1499	0.0000	0.1993	0.0164	0.0000	0.0430	0.0024	0.2247	0.1327	0.3182	0.0958	0.0009	0.0011
ENO2	0.0012	0.4094	0.0000	0.2069	0.0417	0.0000	0.0527	0.0019	0.6462	0.0198	0.0625	0.0171	0.0286	0.2003
ERBB2	0.2359	0.1385	0.0000	0.1432	0.1510	0.0000	0.0049	0.0000	0.2965	0.1034	0.0228	0.0380	0.0421	0.0895
ERG	0.0000	0.0572	0.0000	0.0488	0.0708	0.0000	0.0275	0.0107	0.0000	0.1162	0.0789	0.0044	0.0956	0.0495
ESR1	0.0000	0.0700	0.0000	0.2085	0.2562	0.0000	0.0145	0.0053	0.0000	0.2587	0.2922	0.0007	0.1219	0.3616

FLI1	0.0007	0.0119	0.0062	0.0702	0.0237	0.0091	0.0071	0.0048	0.0056	0.0931	0.0471	0.0126	0.0186	0.0910
FOXL2	0.0000	0.0000	0.6541	0.3268	0.0217	0.0000	0.0038	0.0068	0.0000	0.0073	0.1735	0.1298	0.0158	0.4519
FUT4	0.0000	0.0355	0.0000	0.2257	0.4461	0.0000	0.0217	0.0000	0.0000	0.0113	0.1870	0.0056	0.0874	0.0034
GATA3	0.0000	0.0087	0.0000	0.0255	0.7533	0.0000	0.0126	0.0035	0.0000	0.1591	0.0991	0.1194	1.3531	0.0416
GPC3	0.0000	0.0483	0.0000	0.1366	0.0427	0.0000	0.0030	0.0061	0.0000	0.1143	0.0288	0.0000	0.1322	0.0038
HAVCR1	0.0000	0.0244	0.0000	0.0296	0.0290	0.0008	0.0000	0.0000	0.0997	0.1009	0.1116	0.0356	0.0612	0.0017
HNF1B	0.0000	0.0097	0.0000	0.0412	0.2391	0.0000	0.0117	0.0000	0.1674	0.2912	0.1936	0.2745	0.1571	0.0000
IL12B	0.0000	0.0270	0.0000	0.1642	0.0112	0.0000	0.0545	0.0016	0.0086	0.0484	0.0191	0.0000	0.0067	0.0000
IMP3	0.0000	0.0000	0.0000	0.1021	0.0161	0.0000	0.0068	0.0000	0.0000	0.0256	0.1442	0.0083	0.0145	0.0110
INHA	0.0000	0.1020	0.0000	0.5386	0.0755	0.1400	0.0474	0.0000	0.0687	0.0125	0.0112	0.2668	0.0717	0.0000
ISL1	0.2415	0.5980	0.0000	0.1816	0.6570	0.0000	0.0000	0.0000	0.0000	0.0468	0.0848	0.0062	0.1594	0.0000
KIT	0.0000	0.0140	0.0000	0.0467	0.0867	0.0000	0.0043	0.1085	0.1652	0.0227	0.0778	0.0000	0.0080	0.0058
KLK3	0.0000	0.0140	0.0000	0.0130	0.0244	0.0000	1.2859	0.0000	0.0000	0.0032	0.0845	0.0000	0.0148	0.0000
KL	0.0000	0.0000	0.0000	0.1202	0.0208	0.0000	0.2215	0.0345	0.0000	0.0091	0.0269	0.0349	0.1833	0.0000
KRT10	0.0000	0.1224	0.0000	0.0549	0.1298	0.0000	0.0055	0.0177	0.0000	0.0952	0.0443	0.0044	0.0308	0.0076
KRT14	0.0000	0.0120	0.0000	0.0077	0.0418	0.0003	0.0028	0.0000	0.3191	0.0859	0.0383	0.0053	0.1801	0.0000
KRT15	0.0000	0.0241	0.0000	0.1212	0.0182	0.0000	0.0443	0.0081	0.0000	0.0737	0.1695	0.0000	0.0225	0.0000
KRT16	0.0000	0.0000	0.0000	0.0369	0.0679	0.0000	0.0000	0.0026	0.0163	0.0053	0.0550	0.0488	0.0050	0.0000
KRT17	0.0000	0.0183	0.0000	0.1493	0.0220	0.0000	0.0508	0.0000	0.0000	0.0417	0.5310	0.0329	0.1235	0.0010
KRT18	0.0000	0.0000	0.0000	0.1602	0.0248	0.0000	0.0772	0.6936	0.0110	0.1117	0.0600	0.0000	0.0102	0.7609
KRT19	0.0000	0.0000	0.0000	0.0251	0.1952	0.0013	0.0515	0.7039	0.0276	0.0514	0.0339	0.0085	0.2366	1.0412
KRT1	0.0000	0.0018	0.0031	0.0649	0.0446	0.0000	0.0021	0.0000	0.0167	0.0090	0.0199	0.0004	0.0298	0.0933
KRT20	0.0000	0.0000	0.0000	0.0395	0.0796	0.0000	0.0521	0.0000	0.0000	0.2969	0.3367	0.0000	0.5293	0.0015
KRT2	0.0000	0.0000	0.0000	0.0261	0.0074	0.0000	0.1371	0.0000	0.0000	0.0201	0.0433	0.0512	0.0236	0.0444
KRT3	0.0000	0.0000	0.0000	0.0489	0.1180	0.0006	0.0037	0.0000	0.0000	0.0072	0.0322	0.0000	0.0393	0.0129
KRT4	0.0000	0.0000	0.0000	0.0691	0.0339	0.0000	0.0000	0.0053	0.0107	0.0972	0.1146	0.0000	0.1128	0.0086
KRT5	0.0000	0.0000	0.0000	0.0525	0.0342	0.0464	0.0544	0.0000	0.0019	0.0574	0.4137	0.0000	0.0165	0.0000
KRT6A	0.0000	0.0000	0.0000	0.0507	0.0534	0.0000	0.0755	0.0000	0.0000	0.0051	0.5694	0.0000	0.0213	0.0000
KRT6B	0.0000	0.0011	0.0000	0.0278	0.2216	0.0000	0.0048	0.0042	0.0000	0.0341	0.1458	0.0000	0.0290	0.0903
KRT6C	0.0000	0.0000	0.0000	0.0387	0.2225	0.0000	0.0020	0.0000	0.0000	0.0400	0.1469	0.0000	0.0071	0.0000
KRT7	0.0660	0.0102	0.0000	0.0490	0.1859	0.0005	1.3765	0.0022	0.0544	0.0283	0.0844	0.0521	0.2697	0.0066
KRT8	0.0000	0.0000	0.1357	0.0468	0.1697	0.0000	0.0534	0.6236	0.0000	0.0915	0.0253	0.1412	0.0053	0.1662
LIN28A	0.0000	0.0780	0.0000	0.1663	0.0102	0.0000	0.0186	0.0000	0.0255	0.0894	0.0626	0.0028	0.0074	0.0043
LIN28B	0.0007	0.0527	0.0000	0.0413	0.0414	0.0000	0.0025	0.0000	0.0000	0.0229	0.0846	0.1007	0.0607	0.0000
MAGEA2	0.0000	0.0000	0.0000	0.0006	0.0882	0.0000	0.0000	0.0000	0.0009	0.0000	0.0079	0.0000	0.0031	0.0000
MDM2	0.0000	0.1009	0.0000	0.0494	0.1451	0.0000	0.0000	0.1194	0.0224	0.1082	0.0439	0.0000	0.0195	0.1168
MIB1	0.0000	0.0000	0.0000	0.0799	0.0341	0.0000	0.0075	0.0000	0.0000	0.0306	0.0208	0.0000	0.0021	0.0052
MITF	0.0000	0.0000	0.0000	0.1419	0.0700	0.0000	0.0864	0.0017	0.0000	0.0541	0.0143	0.0720	0.3510	0.2870
MLANA	0.0006	0.0000	0.0000	0.0667	0.0316	0.0000	0.0027	0.0000	0.0444	0.0496	0.0525	0.0053	0.1215	0.0470
MLH1	0.0000	0.0626	0.0000	0.0548	0.1467	0.0000	0.0000	0.0000	0.0000	0.0187	0.0212	0.0773	0.0245	0.1779
MME	0.0532	0.0052	0.0112	0.0410	0.0900	0.0000	0.0346	0.0004	0.0000	0.2221	0.0427	0.0781	0.1436	0.0163
MPO	0.0000	0.1720	0.0000	0.0319	0.0217	0.0000	0.0005	0.0000	0.0000	0.2111	0.0431	0.1047	0.0350	0.0061

MS4A1	0.0000	0.0173	0.0000	0.0720	0.0081	0.0000	0.0000	0.0113	0.0000	0.0174	0.0821	0.0029	0.0050	0.0000
MSH2	0.0000	0.0039	0.0000	0.0545	0.2342	0.0027	0.0000	0.0060	0.0035	0.0118	0.2956	0.0045	0.0144	0.0591
MSH6	0.0000	0.0347	0.1914	0.0060	0.0730	0.0000	0.0000	0.0000	0.0125	0.0258	0.1152	0.0385	0.0057	0.0000
MSLN	0.0000	0.0000	0.0000	0.2905	0.2293	0.0843	0.1757	0.0000	0.0000	0.0904	0.0835	0.0353	0.3326	0.3346
MTHFR	0.0000	0.0399	0.0000	0.0657	0.0602	0.0000	0.0020	0.0015	0.0000	0.0247	0.0902	0.0093	0.0718	0.0006
MUC1	0.0000	0.1051	0.1647	0.1800	0.0815	0.0000	0.2526	0.0000	0.0253	0.0179	0.0801	0.1233	0.5292	0.0276
MUC2	0.0000	0.0000	0.0000	0.0507	0.0817	0.0000	0.2307	0.0000	0.0000	0.4382	0.0224	0.0056	0.0018	0.0049
MUC4	0.0066	0.1878	0.0000	0.0428	0.1120	0.0000	0.0217	0.0000	0.0000	0.1516	0.0536	0.1056	0.0034	0.0801
MUC5AC	0.0000	0.0000	0.0000	0.1069	0.5233	0.0000	0.1067	0.0000	0.0000	0.0320	0.0637	0.0000	0.1855	0.0000
MYOD1	0.0000	0.0004	0.0000	0.1284	0.0361	0.0000	0.0000	0.0000	0.0000	0.0328	0.0178	0.0000	0.0752	0.0049
MYOG	0.0767	0.0000	0.0000	0.0218	0.0141	0.0000	0.0021	0.0000	0.0043	0.0015	0.0644	0.0000	0.0291	0.0873
NANOG	0.0000	0.0064	0.0000	0.0363	0.0361	0.0000	0.0000	0.0000	0.0000	0.0123	0.0411	0.0073	0.0478	0.0308
NAPSA	0.0000	0.0406	0.0000	0.0559	0.2030	0.0000	0.0200	0.0007	0.0022	0.1853	0.1043	0.0003	0.2322	0.0000
NCAM1	0.0000	0.6042	0.0000	0.1455	0.0044	0.0000	0.0000	0.0000	0.0000	0.1297	0.0456	0.0132	0.0253	0.6726
NCAM2	0.0000	0.0000	0.0000	0.1088	0.1730	0.0006	0.0543	0.0000	0.0000	0.1071	0.0958	0.0103	0.0727	0.0321
NKX2-2	0.0000	0.0469	0.0000	0.1041	0.1918	0.0000	0.0406	0.0000	0.0579	0.0976	0.0559	0.0000	0.0855	0.0838
NKX3-1	0.0000	0.0162	0.0000	0.2255	0.0636	0.0000	1.2703	0.0000	0.0000	0.0145	0.0570	0.0286	0.0659	0.0010
OSCAR	0.0000	0.0008	0.0000	0.0600	0.2009	0.0000	0.0099	0.0026	0.0000	0.0245	0.1075	0.1099	0.0620	0.0284
PAX2	0.0000	0.0103	0.0000	0.0552	0.0219	0.0000	0.0000	0.0000	0.0000	0.0737	0.0483	0.0000	0.0477	0.0000
PAX5	0.0000	0.0000	0.0000	0.0671	0.0196	0.0000	0.0542	0.0000	0.0040	0.0528	0.0503	0.0162	0.1061	0.0000
PAX8	0.0000	0.1138	0.0000	0.8760	0.0330	0.0000	0.0026	0.0000	0.0892	0.0869	0.1754	0.6914	0.2608	0.0000
PDPN	0.0000	0.0000	0.0000	0.1066	0.2313	0.1504	0.0037	0.0078	0.0000	0.1543	0.2600	0.0025	0.0932	0.0256
PDX1	0.0000	0.0127	0.0000	0.1495	0.8076	0.0000	0.0202	0.0000	0.0000	0.7265	0.0707	0.0316	0.0336	0.0032
PECAM1	0.0000	0.0141	0.0000	0.0918	0.0178	0.0000	0.0730	0.0072	0.0000	0.0082	0.0297	0.0000	0.0080	0.0256
PGR	0.0000	0.0154	0.1352	0.1223	0.0433	0.0000	0.0214	0.0096	0.0000	0.0230	0.0572	0.0000	0.0142	0.0000
PIP	0.0000	0.0091	0.0000	0.0373	0.0157	0.0000	0.0799	0.0098	0.5509	0.0078	0.0342	0.0141	0.1562	0.0000
PMEL	0.0000	0.0000	0.0000	0.1900	0.0832	0.0000	0.1445	0.0000	0.0000	0.2305	0.0862	0.0058	0.0520	0.0740
PMS2	0.0000	0.0471	0.0000	0.0221	0.1820	0.0000	0.0438	0.0000	0.0000	0.0560	0.1036	0.0000	0.0549	0.0000
POU5F1	0.0004	0.3770	0.0000	0.2549	0.1719	0.0000	0.0000	0.0028	0.0000	0.0305	0.0599	0.0425	0.0268	0.0211
PSAP	0.0000	0.0000	0.0000	0.0594	0.0153	0.0000	0.0000	0.0000	0.0061	0.0384	0.1554	0.0155	0.0005	0.0000
PTPRC	0.0000	0.0129	0.0000	0.1692	0.0172	0.0024	0.0061	0.0000	0.0000	0.1415	0.0390	0.0028	0.0000	0.1112
S100A10	0.0000	0.0263	0.0000	0.2405	0.0918	0.0000	0.1119	0.0054	0.0000	0.0692	0.0531	0.0230	0.2036	0.0346
S100A11	0.0000	0.1247	0.0011	0.0184	0.1784	0.0007	0.0295	0.0000	0.0000	0.0037	0.0163	0.0006	0.0173	0.0112
S100A12	0.0846	0.0066	0.0000	0.0844	0.0266	0.0000	0.0781	0.0000	0.0000	0.0582	0.0304	0.0000	0.0088	0.1121
S100A13	0.0000	0.0067	0.0000	0.3704	0.0017	0.0239	0.0681	0.0000	0.0000	0.0328	0.0461	0.0058	0.0091	0.0000
S100A14	0.0787	0.0124	0.0000	0.0590	0.1071	0.0000	0.0434	0.2697	0.0000	0.1100	0.2446	0.0683	0.1086	0.3884
S100A16	0.0000	0.0243	0.0000	0.0818	0.0216	0.0000	0.0600	0.0000	0.0047	0.0123	0.0207	0.0019	0.1370	0.0289
S100A1	0.0000	0.2747	0.0000	0.1272	0.0683	0.0000	0.0000	0.0000	0.3037	0.1091	0.4703	0.0000	0.0297	0.0107
S100A2	0.0000	0.0000	0.0000	0.0214	0.1344	0.0000	0.0271	0.0000	0.0027	0.1516	0.2694	0.2900	0.4107	0.0000
S100A4	0.0000	0.0068	0.0000	0.0840	0.2693	0.0000	0.0328	0.0000	0.0137	0.0158	0.0583	0.0000	0.1036	0.0168
S100A5	0.0000	0.0020	0.0000	0.0335	0.0678	0.0000	0.3275	0.0000	0.0000	0.0634	0.0096	0.0041	0.1003	0.0000
S100A6	0.0000	0.0127	0.0000	0.0136	0.0168	0.0000	0.0967	0.0000	0.0073	0.0402	0.2069	0.0200	0.0475	0.0000

S100A7A	0.0000	0.0000	0.0000	0.0492	0.1427	0.0004	0.0171	0.0000	0.0109	0.0029	0.0318	0.0021	0.0063	0.0115
S100A7L2	0.0000	0.0066	0.0000	0.0042	0.0012	0.0000	0.0000	0.0000	0.0000	0.0390	0.0553	0.0314	0.0173	0.0000
S100A7	0.0000	0.1408	0.0000	0.0500	0.0629	0.0000	0.0042	0.0000	0.0037	0.0085	0.0360	0.0000	0.0029	0.0000
S100A8	0.0000	0.0000	0.0000	0.0504	0.0777	0.0000	0.0043	0.0450	0.0082	0.1005	0.0850	0.0000	0.0119	0.0000
S100A9	0.0000	0.0436	0.0000	0.0086	0.0392	0.0000	0.0000	0.0082	0.0009	0.0330	0.0185	0.0047	0.0027	0.0000
S100B	0.0000	0.0000	0.0036	0.0204	0.0343	0.0000	0.0042	0.0272	0.0518	0.0473	0.0446	0.0082	0.0706	0.0833
S100PBP	0.0650	0.0176	0.0000	0.0800	0.0832	0.0000	0.0057	0.0142	0.0032	0.0051	0.0238	0.0204	0.0673	0.0144
S100P	0.0000	0.0000	0.0000	0.0740	0.2088	0.0000	0.0047	0.0218	0.0051	0.1975	0.0230	0.1375	0.3496	0.1993
S100Z	0.0000	0.1949	0.0000	0.0160	0.2012	0.0000	0.0125	0.0026	0.0000	0.0496	0.0178	0.0066	0.0035	0.0000
SALL4	0.0000	0.0000	0.0000	0.0322	0.2072	0.0000	0.0208	0.0000	0.1862	0.0444	0.0452	0.0292	0.3200	0.0245
SATB2	0.0000	0.0050	0.0000	0.0988	0.1879	0.0029	0.0332	0.0113	0.0128	0.0693	0.1365	0.0066	0.1447	0.1369
SDC1	0.0681	0.0167	0.2236	0.1215	0.0221	0.0000	0.1176	0.1562	0.0113	0.0265	0.3517	0.0279	0.0329	0.0632
SERPINA1	0.0000	0.0069	0.0076	0.1785	0.6933	0.0000	0.1383	0.0000	0.0000	0.3080	0.0627	0.0051	0.3476	0.0082
SERPINB5	0.0000	0.0607	0.0000	0.0683	0.1196	0.0000	0.0042	0.0012	0.0000	0.0982	0.2638	0.1166	0.0712	0.0000
SF1	0.0000	0.0000	0.0000	0.1115	0.1241	0.0163	0.0434	0.0000	0.0000	0.0401	0.0082	0.0047	0.0028	0.0000
SFTPA1	0.0000	0.0321	0.0028	0.1190	0.1051	0.0000	0.0945	0.0000	0.0000	0.2277	0.4403	0.0505	0.0514	0.0000
SMAD4	0.0000	0.0168	0.0000	0.0566	0.4264	0.0000	0.0020	0.0523	0.0181	0.0162	0.0363	0.0000	0.0314	0.0045
SMARCB1	0.0000	0.0000	0.0000	0.1221	0.2192	0.1813	0.0000	0.0000	0.0000	0.0136	0.0824	0.0183	0.0000	0.0000
SMN1	0.0000	0.0090	0.0000	0.0235	0.2683	0.0000	0.0000	0.0000	0.0000	0.1115	0.0403	0.0125	0.0218	0.0472
SOX2	0.0000	0.0342	0.0000	0.2216	0.2178	0.0000	0.0115	0.0031	0.0419	0.2305	0.6443	0.0000	0.1667	0.0869
SPN	0.0000	0.0223	0.0000	0.1472	0.1709	0.0000	0.0000	0.0000	0.0146	0.1605	0.0583	0.0211	0.0367	0.0265
SYP	0.0000	0.3155	0.0000	0.2023	0.0230	0.0087	0.0283	0.0007	0.0000	0.1538	0.0614	0.0493	0.0275	0.0117
TFE3	0.0000	0.0000	0.0000	0.3920	0.0098	0.0000	0.0210	0.0060	0.0000	0.0933	0.0856	0.0000	0.0137	0.0012
TFF1	0.0000	0.0045	0.0000	0.0313	0.2263	0.0000	0.0840	0.0061	0.2886	0.1426	0.0275	0.0008	0.1139	0.0141
TFF3	0.0000	0.3324	0.0000	0.1789	0.1254	0.0000	0.0000	0.0000	0.0110	0.1575	0.0444	0.1715	0.0229	0.0162
TG	0.0000	0.0457	0.0000	0.1462	0.0907	0.0000	0.0763	0.0000	0.0000	0.0046	0.0501	0.8319	0.0058	0.0026
TLE1	0.0000	0.0000	0.0000	0.3220	0.0808	0.0000	0.0184	0.0851	0.0000	0.2334	0.1047	0.1768	0.0664	0.0000
TMPRSS2	0.0475	0.0061	0.0000	0.1440	0.1280	0.0000	0.1206	0.0720	0.1013	0.0610	0.1099	0.0003	0.0443	0.0089
TNFRSF8	0.0000	0.0492	0.0000	0.0109	0.0088	0.0004	0.0728	0.0093	0.0000	0.0617	0.0232	0.0000	0.0062	0.0015
TP63	0.0000	0.0335	0.0000	0.0277	0.1223	0.0000	0.0000	0.0000	0.0061	0.0907	2.3082	0.0000	0.3923	0.0014
TPM1	0.0000	0.0000	0.0020	0.0425	0.2042	0.0000	0.0132	0.3712	0.5131	0.0215	0.1198	0.0391	0.0075	0.2254
TPM2	0.0000	0.0247	0.0000	0.0497	0.0282	0.0000	0.0093	0.0050	0.0111	0.0265	0.0889	0.0038	0.0689	0.0100
TPM3	0.0006	0.0528	0.0000	0.0773	0.0662	0.0000	0.0794	0.0713	0.0129	0.0567	0.2273	0.0725	0.0227	0.0079
TPM4	0.0000	0.2880	0.0000	0.1518	0.0796	0.0000	0.0521	0.2444	0.0015	0.1282	0.0779	0.0004	0.0386	0.1426
TPSAB1	0.0000	0.0428	0.0000	0.1971	0.1180	0.0012	0.0668	0.0114	0.0000	0.1520	0.1283	0.2829	0.0985	0.0155
TTF1	0.0000	0.0000	0.0000	0.0127	0.0491	0.0000	0.0088	0.0000	0.0000	0.0786	0.2237	0.0000	0.0194	0.0000
UPK2	0.0000	0.0000	0.0000	0.0039	0.0129	0.0000	0.0058	0.0000	0.0000	0.0826	0.0436	0.0000	0.5618	0.0000
UPK3A	0.0000	0.0727	0.0000	0.0806	0.0537	0.0000	0.2229	0.0736	0.0000	0.0270	0.0645	0.0960	0.2551	0.0062
UPK3B	0.0000	0.0000	0.0000	0.0668	0.0437	0.5605	0.0272	0.0017	0.0135	0.0289	0.0574	0.0268	0.0952	0.2858
VHL	0.0000	0.0393	0.0000	0.1045	0.0238	0.0000	0.0052	0.0000	0.0075	0.0042	0.0913	0.0059	0.2840	0.0023
VIL1	0.0000	0.1146	0.0000	0.1179	0.0235	0.0000	0.0000	0.0000	0.0000	0.0289	0.0364	0.0000	0.2484	0.1114
VIM	0.0000	0.0000	0.0000	0.0857	0.0377	0.0000	0.0413	0.0000	0.0012	0.0425	0.0817	0.2083	0.2505	0.0040

WT1	0.0000	0.0173	0.0000	2.0098	0.0094	0.3547	0.0022	0.0118	0.0000	0.0346	0.0731	0.0072	0.1587	0.0315

Table 119 – Importance of RNA Transcripts used to Classify Organ Type

Transcript	AG	Bla	Brain	Br	Colon	Eye	FGTP	Gast	HFN	Kid	LGC	Lung	Panc	Pros	Skin	SI	Thy
ACVRL1	.0003	.0671	.0000	.0475	.0222	.0000	.0056	.0236	.0064	.0680	.0876	.0352	.0320	.0005	.0272	.0094	.0000
AFP	.0000	.0096	.0000	.0369	.1508	.0000	.0130	.1900	.0214	.0000	.0740	.0188	.0423	.0019	.0028	.0427	.0012
ALPP	.0000	.0096	.0000	.0724	.1021	.0000	.1964	.0383	.0181	.0172	.0522	.0222	.1045	.0269	.0104	.0000	.0000
AMACR	.0000	.0913	.0000	.1646	.0941	.0005	.0430	.1599	.0887	.2368	.1110	.0666	.2646	.5598	.3141	.0064	.0000
ANKRD30A	.0000	.0124	.0000	.8385	.0095	.0000	.0209	.0134	.0004	.0000	.1418	.0822	.1093	.0000	.0045	.0000	.0000
ANO1	.0000	.1123	1.0334	.1658	.0384	.0000	.2532	.6185	.2232	.0825	.4571	.1535	.7984	.0207	.0738	.2189	.0014
ARG1	.0313	.0395	.0000	.0809	.1492	.0000	.1317	.0390	.0177	.0488	.0170	.0735	.1897	.0000	.0252	.0469	.3135
AR	.0000	.0745	.0679	.1416	.0317	.0000	.2628	.3634	.0504	.1697	.1404	.4098	.1246	.0766	.2539	.0690	.0000
BCL2	.0000	.0627	.0850	.0299	.0123	.3040	.2323	.1117	.0239	.0200	.1067	.0598	.0308	.0589	.0184	.0060	.0040
BCL6	.0000	.0723	.0279	.0000	.0422	.0002	.1007	.0607	.0158	.1668	.1525	.1039	.0186	.1279	.2406	.1593	.0000
CA9	.0000	.1180	.0000	.1187	.1010	.0007	.0292	.1173	.0200	.1638	.1019	.0117	.0125	.0181	.0406	.0452	.0608
CALB2	.0882	.3649	.0000	.0711	.0760	.0000	.2521	.0375	.0236	.0000	.1588	.0353	.2212	.0156	.0274	.1687	.2420
CALCA	.0000	.0092	.0000	.0622	.0957	.0000	.0353	.0744	.0032	.0953	.0859	.0437	.0637	.0021	.0768	.0072	.0000
CALD1	.0000	.0055	.0391	.0768	.0371	.0000	.1536	.0040	.0025	.0110	.1722	.1287	.0349	.0000	.0732	.2104	.0003
CCND1	.0000	.0979	.0147	.1192	.0074	.0056	.2440	.1178	.0452	.0208	.0268	.0110	.0890	.0000	.0288	.0589	.0851
CD1A	.0000	.0757	.0000	.0888	.0243	.0000	.0162	.2311	.0789	.0000	.0915	.0221	.1749	.0205	.0518	.0338	.0103
CD2	.0000	.2638	.0096	.0297	.1065	.0000	.0481	.0622	.0384	.0000	.0510	.0071	.0942	.0167	.0935	.0242	.0153
CD34	.0282	.0182	.0016	.0150	.1194	.0000	.0274	.3914	.0189	.1022	.0415	.0971	.0999	.1035	.1163	.0000	.0000
CD3G	.0000	.2669	.0157	.0464	.0414	.0000	.1717	.0928	.0025	.0000	.0031	.0387	.0419	.0224	.0874	.0018	.0000
CD5	.0000	.2324	.1592	.1878	.0535	.0000	.0275	.0993	.0954	.0000	.1891	.0497	.3574	.0052	.0345	.3299	.0062
CD79A	.0000	.0133	.0000	.0729	.0477	.0020	.0423	.1161	.0386	.0000	.1012	.0752	.0642	.0025	.1694	.0592	.0098
CD99L2	.0000	.0754	.0123	.1116	.0727	.0000	.1779	.0798	.1949	.0000	.0917	.3663	.0641	.0045	.0071	.0049	.0087
CDH17	.0000	.0423	.0033	.0032	.3831	.0000	.0184	.0422	.0172	.0000	.0189	.0817	.0842	.0108	.0334	.4462	.0000
CDH1	.1257	.0168	.0399	.1486	.0120	.0000	.1459	.3014	.0925	.7014	.0143	.0326	.0373	.0667	.0966	.0000	.0322
CDK4	.0000	.1171	.0018	.0056	.0590	.0000	.2757	.0669	.0363	.0000	.1529	.0802	.0494	.0161	.0046	.0000	.2172
CDKN2A	.0000	.1014	.0453	.2024	.1300	.0000	.4237	.0981	.0318	.4499	.1653	.1417	.1154	.0370	.0037	.0634	.0172
CDX2	.0000	.0502	.0047	.1807	1.3118	.0000	.1523	.7682	.0101	.0000	.0409	.0862	.1480	.0085	.0040	.3510	.0000
CEACAM16	.0000	.1401	.0000	.1643	.0981	.0000	.0547	.0539	.0290	.0096	.1304	.1034	.0742	.0072	.2789	.1652	.0050
CEACAM18	.0000	.0097	.0003	.0977	.1766	.0000	.0426	.0255	.0055	.0000	.0392	.0807	.1546	.0422	.0000	.1313	.0488
CEACAM19	.0000	.0328	.0000	.0222	.0298	.0000	.0437	.2109	.0297	.0378	.0833	.1299	.0743	.0132	.2811	.0099	.0167
CEACAM1	.0000	.1303	.5129	.0081	.1826	.0000	.0548	.0400	.1096	.0096	.0813	.2729	.0858	.0877	.1139	.0000	.0159
CEACAM20	.0000	.0022	.0000	.0018	.1326	.0000	.0038	.0505	.1120	.0046	.0392	.0026	.0285	.0000	.0114	.0000	.0000
CEACAM21	.0000	.0152	.0000	.0329	.0114	.0000	.1227	.0088	.0744	.0000	.1198	.0040	.0026	.0839	.0093	.0167	.0000
CEACAM3	.0000	.0312	.0059	.0372	.0454	.0000	.0089	.1434	.0223	.0000	.0909	.0587	.1765	.0244	.0084	.0121	.0584
CEACAM4	.0000	.0812	.0675	.1648	.0174	.0000	.0276	.0942	.0046	.0000	.0487	.0132	.1209	.0000	.0834	.1479	.0189
CEACAM5	.0000	.0332	.0000	.0755	.4657	.0000	.1099	.0082	.1680	.0825	.1855	.0166	.0626	.0518	.0388	.0260	.2552
CEACAM6	.0000	.1477	.0000	.0124	.0330	.0000	.1584	.3346	.0446	.0170	.0117	.3440	.1333	.0965	.0000	.0246	.0039
CEACAM7	.0000	.0128	.0000	.2111	.1943	.0000	.1543	.0694	.0782	.0037	.1400	.3624	.1242	.0151	.0259	.1387	.0000
CEACAM8	.0000	.0666	.0000	.0080	.1539	.0000	.1574	.0168	.2591	.0040	.0254	.1268	.1016	.0000	.0000	.0095	.0000
CGA	.0000	.0482	.0000	.0109	.0306	.0000	.0434	.0112	.0056	.0000	.0458	.0190	.1832	.0000	.0177	.0942	.1288

CGB3	.0000	.0477	.0885	.0198	.0598	.0000	.0676	.1499	.0030	.0000	.1153	.0650	.0147	.2017	.0542	.0268	.0000
CNN1	.0000	.2837	.0179	.1656	.1832	.0000	.0795	.0394	.1034	.0000	.2537	.2339	.0232	.0806	.1730	.2583	.2661
COQ2	.0000	.0445	.0060	.0623	.1028	.0002	.0235	.1307	.0422	.0538	.1192	.0157	.1701	.0072	.0956	.0000	.0000
CPS1	.0000	.4645	.0000	.0101	.1177	.0000	.1630	.0638	.0412	.1171	.0499	.0792	.2032	.3389	.0451	.0038	.3436
CR1	.0002	.0075	.0317	.0205	.1081	.0000	.1264	.0577	.0068	.0362	.0119	.0909	.0211	.0000	.1970	.1178	.0025
CR2	.0000	.0099	.0000	.0120	.0336	.0003	.0377	.0600	.0356	.0002	.0466	.0196	.1997	.0860	.0047	.0106	.0000
CTNNB1	.0000	.1319	.0000	.0328	.0840	.0043	.0529	.1220	.0080	.0000	.0696	.0631	.0404	.0000	.0105	.1604	.0098
DES	.0000	.4203	.0279	.2248	.1060	.0000	.3107	.2486	.0051	.0097	.1672	.1804	.2281	.0000	.1019	.2349	.0030
DSC3	.0000	.0068	.0118	.0430	.1329	.0000	.0392	.0577	.7147	.0027	.0996	.0414	.0225	.0057	.0000	.2462	.0833
ENO2	.0000	.0167	.0391	.0912	.0702	.0379	.0214	.3843	.2596	.2268	.2694	.1003	.0542	.0415	.0051	.0032	.0127
ERBB2	.0000	.0365	.0215	.0124	.1209	.0000	.1466	.1053	.1397	.1138	.0167	.2024	.1639	.0000	.0154	.0398	.0229
ERG	.0002	.0992	.0152	.0179	.2343	.0055	.0952	.0249	.0127	.0120	.0242	.0392	.0743	.0370	.0403	.0363	.0000
ESR1	.0000	.1535	.0652	.1127	.1408	.0000	1.0530	.0577	.1233	.0391	.4028	.1011	.1813	.0210	.1503	.0167	.0000
FLI1	.0000	.0665	.0074	.0187	.0942	.0000	.0424	.0080	.1055	.0145	.0456	.1075	.0187	.0317	.0157	.4217	.0358
FOXL2	.0000	.0094	.0131	.0225	.1601	.0000	.4227	.1110	.0621	.0000	.0669	.0549	.0137	.0024	.0297	.0452	.1166
FUT4	.0000	.1533	.0749	.0810	.2366	.0000	.0897	.5438	.0129	.0963	.0524	.1631	.3926	.0295	.0072	.1623	.0615
GATA3	.0000	1.3362	.0360	2.0010	.0265	.0000	.2732	.0478	.2203	.0386	.1597	.1885	.6680	.0035	.3548	.0047	.0887
GPC3	.0000	.0924	.1749	.0215	.1034	.0000	.1597	.0236	.0336	.0773	.1257	.0690	.0641	.0000	.0846	.0601	.0000
HAVCR1	.0000	.0285	.0000	.0259	.2369	.0017	.0156	.0702	.1647	.4680	.0909	.0878	.0346	.0000	.0055	.0016	.0163
HNF1B	.0000	.1637	.0266	.4322	.2227	.0008	.1474	.0309	.3677	.4912	.7119	.0808	.2556	.0061	.0959	.0171	.2405
IL12B	.0000	.0205	.0000	.0478	.0434	.0000	.1123	.0416	.1894	.0024	.0282	.1107	.0043	.0498	.0148	.0370	.0000
IMP3	.0000	.0818	.0000	.0050	.0307	.0000	.0080	.0336	.0100	.0000	.0504	.0384	.0222	.0000	.0195	.0000	.0000
INHA	.1494	.0375	.1251	.0282	.0321	.0000	.0473	.1673	.0870	.0000	.1546	.0468	.0852	.0294	.0331	.0017	.3150
ISL1	.0000	.2428	.0260	.1131	.0911	.0000	.0789	.2998	.0819	.0000	.0930	.2304	.6155	.0020	.0238	.0300	.0000
KIT	.0000	.0213	.0000	.1038	.0682	.0000	.1478	.1008	.0510	.0256	.0399	.1076	.1514	.0166	.0142	.0077	.0000
KLK3	.0000	.0610	.0000	.0352	.1028	.0000	.0257	.0090	.0512	.0152	.1014	.0322	.0469	1.2958	.0281	.0051	.0000
KL	.0000	.1684	.0000	.1550	.0225	.0000	.0553	.0273	.1720	.3120	.2054	.0375	.0267	.2279	.0025	.0000	.0359
KRT10	.0000	.0291	.1109	.0050	.1625	.0080	.0437	.0150	.0548	.0000	.0103	.2288	.1276	.0175	.0061	.0757	.0042
KRT14	.0000	.2083	.0115	.0979	.1050	.0000	.1055	.0955	.1525	.0024	.1009	.0884	.0272	.0000	.1471	.0062	.0000
KRT15	.0000	.0687	.1006	.5284	.0836	.0000	.2371	.0422	.2901	.0096	.0613	.1612	.0350	.0282	.1112	.0227	.0000
KRT16	.0000	.0089	.0331	.2914	.0147	.0000	.1705	.0346	.0179	.0007	.0354	.0804	.0616	.0000	.0611	.0371	.0580
KRT17	.0000	.0528	.0170	.0347	.1050	.0000	.0713	.0267	.0407	.0431	.1401	.0749	.0457	.0283	.0842	.0167	.0000
KRT18	.0000	.0043	.2272	.4277	.3549	.0000	.1155	.0070	.0830	.0004	.0609	.0817	.0206	.0776	.1036	.0018	.0000
KRT19	.0524	.2239	.0315	.0629	.1533	.0000	.0312	.0394	.0225	.0184	.0307	.1090	.1840	.0517	.3821	.0000	.0044
KRT1	.0000	.0547	.0000	.0268	.0407	.0000	.0190	.0299	.0197	.0000	.0246	.0396	.0360	.0133	.1066	.0117	.0000
KRT20	.0000	.5602	.0000	.1009	.6969	.0000	.0228	.1630	.0523	.0001	.0346	.2407	.0662	.1508	.0657	.3990	.0004
KRT2	.0000	.0174	.0000	.0222	.0340	.0005	.0429	.0963	.0930	.0452	.0181	.0410	.0107	.0947	.0243	.0202	.0438
KRT3	.0000	.0459	.0000	.0410	.0097	.0000	.0436	.0106	.0721	.0096	.0929	.0205	.1160	.0022	.0018	.0000	.0000
KRT4	.0000	.0579	.0000	.0604	.1359	.0000	.0581	.0740	.1764	.0000	.1881	.0467	.0230	.0158	.0114	.0309	.0000
KRT5	.0000	.0561	.0448	.2414	.0894	.0000	.3243	.0082	.7575	.0018	.2450	.0642	.0502	.0817	.0730	.0137	.0000
KRT6A	.0000	.0183	.0018	.0846	.1164	.0000	.0237	.0195	.0203	.0000	.0114	.3301	.0551	.0683	.0067	.0202	.0042
KRT6B	.0000	.0209	.0000	.2187	.3467	.0000	.0287	.0547	.0743	.0033	.0520	.0848	.2088	.0106	.0086	.1043	.0000
KRT6C	.0000	.0067	.0000	.0556	.0036	.0000	.0762	.1064	.0047	.0000	.0110	.0227	.1520	.0476	.0049	.0000	.0000
KRT7	.0000	.2521	.0628	.5254	1.2701	.0080	.0557	.0694	.0345	.2875	.2164	.3106	.1843	1.2860	.4042	.3030	.0339

KRT8	.0570	.0070	1.0342	.0194	.0289	.0005	.0726	.0753	.1716	.0324	.1153	.0806	.1772	.1102	.6755	.1144	.0822
LIN28A	.0000	.0072	.0000	.0096	.0637	.0000	.0120	.0076	.0156	.0000	.0260	.0175	.0343	.0261	.1665	.0280	.0000
LIN28B	.0000	.1592	.0000	.0351	.0450	.0000	.1485	.0676	.2085	.0000	.0138	.0315	.0429	.0041	.0147	.0000	.1655
MAGEA2	.0000	.0013	.0000	.0117	.0020	.0000	.0060	.0392	.0000	.0000	.0856	.0709	.0683	.0000	.0000	.0000	.0000
MDM2	.0000	.0140	.0020	.2969	.0579	.0000	.2265	.0276	.1408	.1983	.1261	.0509	.1656	.0000	.3251	.0574	.0000
MIB1	.0962	.0048	.0331	.0884	.1189	.0544	.0323	.0366	.1373	.0253	.0806	.0671	.0396	.0052	.0199	.0036	.0000
MITF	.0000	.3069	.0213	.0226	.0196	.3109	.0792	.0714	.0180	.0000	.0450	.1549	.0408	.1111	.1420	.1808	.0054
MLANA	.0000	.0648	.0041	.0475	.0192	.3318	.0533	.0368	.0555	.0234	.0977	.1835	.0200	.0072	.2699	.0143	.0161
MLH1	.0000	.0189	.0069	.0156	.1564	.0003	.0830	.0191	.1273	.0162	.0594	.2300	.1279	.0034	.0534	.0000	.0822
MME	.0000	.2636	.0013	.0735	.1515	.0000	.0462	.0055	.2608	.1049	.0880	.0335	.0956	.0654	.0839	.1181	.1127
MPO	.0000	.0352	.0000	.0071	.0438	.0000	.0034	.0363	.0201	.0108	.0795	.0499	.0263	.0000	.0029	.2622	.0509
MS4A1	.0000	.0071	.0102	.0584	.1582	.0003	.2448	.0095	.0386	.0113	.1348	.1566	.0104	.0027	.1812	.0078	.0001
MSH2	.0000	.0083	.3471	.0284	.0135	.0000	.2538	.0432	.0156	.0318	.0345	.0813	.1875	.0000	.0084	.0423	.0000
MSH6	.0000	.0000	.0098	.0012	.0104	.0000	.0526	.0790	.1828	.0000	.0206	.1600	.0389	.0056	.0105	.0000	.0148
MSLN	.0000	.3432	.0000	.0438	.1143	.0000	.1068	.0310	.0971	.1380	.0957	.0482	.2315	.1680	.0169	.0940	.0803
MTHFR	.0000	.0064	.0053	.2116	.0403	.0000	.0226	.1700	.0053	.0275	.0372	.1302	.0500	.0170	.0283	.0324	.0186
MUC1	.0000	.3594	.0728	.0028	.5746	.0000	.2050	.1341	.0888	.2678	.0567	.1148	.0732	.2098	.0722	.0115	.0312
MUC2	.0000	.0392	.0000	.0017	.8717	.0000	.0130	.0027	.0146	.0000	.0172	.0546	.0829	.1871	.0133	.5774	.0340
MUC4	.0000	.0522	.0179	.4349	.0926	.0006	.0528	.2242	.1497	.0215	.3392	.2554	.1277	.0737	.1638	.0050	.0487
MUC5AC	.0000	.2247	.0024	.2808	.0850	.0000	.0566	.3093	.2958	.0637	.1325	.1807	.4736	.0776	.0581	.0596	.0000
MYOD1	.0000	.1281	.0218	.0555	.0196	.0000	.0231	.0213	.0067	.0000	.0058	.0145	.0439	.0000	.0102	.0300	.0000
MYOG	.0000	.0302	.0000	.0768	.0186	.0000	.0094	.2205	.1699	.0250	.0118	.0649	.0165	.0028	.0306	.0000	.0014
NANOG	.0000	.0777	.0123	.0107	.0337	.0000	.0263	.0704	.0080	.0000	.0574	.0119	.0502	.0000	.0297	.0000	.0000
NAPSA	.0001	.2645	.0063	.1281	.0415	.0000	.1032	.1494	.0847	.0063	.0746	.9241	.1344	.0284	.0339	.0111	.0169
NCAM1	.0000	.0409	.3968	.0429	.0122	.0055	.0204	.0202	.0186	.0072	.0580	.0368	.0088	.0000	.1824	.0036	.0494
NCAM2	.0437	.0730	.0000	.0737	.1190	.0000	.0972	.4127	.1296	.0000	.1791	.3102	.1403	.0558	.0556	.1095	.0143
NKX2-2	.0000	.1005	.2205	.0522	.0990	.0000	.1576	.0511	.0114	.0000	.1899	.0210	.2672	.0444	.1354	.0048	.0000
NKX3-1	.0425	.0429	.0000	.0292	.1744	.0000	.0960	.1352	.0110	.0000	.1139	.1494	.0219	1.1378	.0109	.0042	.0231
OSCAR	.0000	.0124	.0034	.0532	.1362	.0000	.0294	.0562	.0392	.0016	.0739	.0732	.1713	.0084	.0677	.0391	.1180
PAX2	.0000	.0122	.0000	.0370	.0207	.0000	.1434	.0926	.0067	.2834	.0730	.1325	.0367	.0000	.0162	.0033	.0000
PAX5	.0000	.0924	.0000	.1044	.0086	.0006	.1276	.0185	.2914	.0000	.0805	.0118	.0179	.0557	.0000	.0511	.0056
PAX8	.0000	.3050	.0132	.3208	.0373	.0000	1.2795	.3209	.1479	.8966	.1523	.2109	.0231	.0065	.0731	.1650	.8590
PDPN	.0000	.0124	.6385	.1994	.1385	.0210	.1941	.2792	.0548	.0056	.0053	.0253	.1933	.0000	.0576	.0015	.0019
PDX1	.0000	.0366	.0060	.0316	.0984	.0000	.0538	.1423	.0072	.0078	.0506	.2131	.8132	.0085	.0013	.1270	.0295
PECAM1	.0002	.0141	.0000	.1046	.0353	.0000	.0067	.1972	.0374	.0463	.0920	.0147	.0234	.0973	.0252	.0923	.0000
PGR	.0000	.0186	.1330	.1311	.1656	.0000	.5083	.0444	.2894	.0000	.0100	.0978	.0183	.0296	.0437	.0100	.0000
PIP	.0000	.1526	.0000	.3285	.0380	.0057	.0558	.1931	.1178	.0073	.0483	.0620	.0254	.1123	.0396	.0000	.0155
PMEL	.0003	.0356	.0129	.1972	.1023	1.0156	.0518	.1773	.0228	.0080	.1240	.0124	.1000	.1675	.5473	.1542	.0027
PMS2	.0000	.0287	.0000	.0191	.0260	.0037	.1119	.1046	.0365	.0000	.0377	.0748	.1378	.0177	.0600	.0027	.0000
POU5F1	.0000	.0362	.0000	.0681	.0283	.0000	.1182	.0538	.0786	.2831	.2509	.1150	.2034	.0103	.0055	.0119	.0879
PSAP	.0563	.0265	.0000	.0065	.0869	.0063	.0702	.1636	.0091	.0077	.2201	.0257	.0072	.0003	.0305	.0359	.0162
PTPRC	.0000	.0058	.0000	.0337	.2122	.0000	.0800	.0318	.0066	.0000	.0523	.0629	.0387	.0336	.0000	.0720	.0021
S100A10	.0000	.2972	.0019	.1128	.0151	.1215	.1124	.0085	.0391	.0138	.0175	.4153	.0864	.1658	.1544	.0469	.0782
S100A11	.0000	.0113	.0106	.0099	.0300	.0000	.0426	.3009	.1101	.0000	.0155	.0579	.1451	.0015	.1747	.0000	.0174

S100A12	.0000	.0297	.0036	.0926	.1323	.0000	.0492	.0293	.0774	.0000	.0337	.0770	.0091	.0803	.0804	.0078	.0000
S100A13	.0000	.0057	.0066	.1174	.0270	.1525	.2538	.3404	.0622	.2862	.0851	.2209	.0091	.0197	.1541	.0093	.0106
S100A14	.0000	.0720	.8152	.1965	.2377	.0000	.0929	.0084	.1456	.4861	.1913	.0189	.1482	.0681	.0377	.0124	.0618
S100A16	.0000	.1208	.1491	.0259	.0510	.0310	.1116	.0267	.0073	.0000	.0420	.0424	.0161	.0580	.0579	.0000	.0007
S100A1	.0000	.0444	.1976	.4451	.0344	.0673	.0775	.1901	.1661	.0164	.0598	.4323	.0931	.0000	.1450	.2117	.0128
S100A2	.0001	.3483	.4600	.4888	.1843	.1423	.0662	.0832	.0175	.0000	.3213	.0589	.1294	.0129	.0093	.0260	.1894
S100A4	.0000	.0493	.1041	.0242	.0409	.0000	.0464	.0080	.0180	.0236	.0917	.0350	.2247	.0253	.0231	.0080	.0163
S100A5	.0000	.0429	.0000	.0424	.0227	.0000	.0761	.0986	.1627	.0165	.0511	.1205	.1296	.3310	.0247	.0553	.0053
S100A6	.0000	.1034	.0067	.2751	.2919	.0000	.0925	.0465	.2660	.0000	.1196	.0394	.0183	.0907	.0238	.0206	.0421
S100A7A	.0000	.0312	.0029	.0106	.0538	.0000	.0444	.0724	.0214	.0000	.0421	.0288	.1400	.0000	.0000	.0000	.0191
S100A7L2	.0000	.0166	.0022	.1401	.0685	.0000	.0074	.0299	.0164	.0000	.0000	.0042	.0000	.0086	.0000	.0000	.0433
S100A7	.0005	.0076	.0165	.0118	.0166	.0000	.1777	.2378	.0951	.0012	.0149	.0637	.0359	.0132	.0032	.0000	.0141
S100A8	.0000	.0114	.1244	.0143	.0796	.0000	.1051	.0029	.1445	.0000	.0538	.0194	.0946	.0195	.0000	.0236	.0000
S100A9	.0000	.0745	.0184	.0696	.0332	.0000	.1800	.2175	.0316	.0000	.2408	.0603	.0295	.0136	.0018	.0265	.0026
S100B	.0000	.1028	.9620	.1504	.0476	.0147	.0782	.2350	.2606	.0381	.0658	.0815	.0460	.0101	.8089	.0116	.0270
S100PBP	.0000	.0981	.0301	.0615	.0249	.0000	.0751	.0220	.0301	.0281	.0467	.0860	.1319	.0000	.0862	.0132	.0158
S100P	.0000	.2341	.0121	.1709	.1183	.0000	.1015	.0753	.0791	.4178	.0718	.0110	.0724	.0207	.0289	.0078	.2033
S100Z	.0000	.0187	.1509	.0003	.0101	.0022	.0343	.0934	.0089	.0189	.0111	.1308	.2410	.0419	.1333	.0241	.0153
SALL4	.0000	.4484	.0000	.1879	.0377	.0000	.2077	.0702	.2586	.1135	.0942	.0459	.1665	.0567	.0235	.0040	.1158
SATB2	.0000	.2100	.0196	.0157	.3127	.0036	.0687	.1100	.0978	.0070	.1929	.0649	.2148	.0420	.0683	.0284	.0033
SDC1	.0000	.0480	.0442	.0335	.0946	.0000	.0525	.1007	.0971	.0000	.0066	.0872	.0177	.0760	.0779	.1141	.0150
SERPINA1	.0297	.4227	.0000	.2262	.0950	.0000	.2388	.0393	.0243	.0568	.7522	.0195	.7488	.1644	.0341	.0653	.0039
SERPINB5	.0000	.0369	.0189	.1948	.1726	.0000	.0596	.4347	.0312	.0599	.0663	.0783	.0690	.0000	.0019	.0145	.3405
SF1	.0000	.0049	.0000	.0792	.0235	.0000	.0335	.0198	.0655	.1336	.0670	.0822	.1559	.0473	.1015	.1107	.0000
SFTPA1	.0000	.1543	.0051	.0297	.0753	.0000	.1514	.1391	.0353	.0000	.0969	.5577	.0979	.1310	.0365	.0295	.0244
SMAD4	.0000	.0259	.0000	.0259	.0948	.0000	.0713	.0336	.0542	.0000	.0119	.0468	.4014	.0205	.0936	.0000	.0138
SMARCB1	.0000	.0041	.0837	.0317	.1247	.0003	.3124	.0567	.0059	.0000	.0740	.0388	.1731	.0000	.0035	.0000	.0161
SMN1	.0000	.0294	.0000	.0241	.1636	.0015	.0893	.0755	.0065	.0067	.0227	.0686	.2914	.0048	.0977	.0000	.0104
SOX2	.0000	.2171	.6623	.3559	.2748	.0379	.1072	.3247	.0164	.0373	.3972	.6865	.2639	.0029	.0966	.0875	.0000
SPN	.0000	.0442	.0704	.0443	.0209	.0000	.0745	.4132	.1534	.0000	.0176	.0390	.1740	.0000	.0020	.1942	.0189
SYP	.1184	.0457	.0037	.0826	.0476	.0052	.0610	.1916	.1654	.1942	.0233	.0281	.0659	.0809	.0443	.0725	.0114
TFE3	.0000	.0803	.0000	.1118	.0113	.0000	.1354	.0475	.1683	.0202	.1734	.0574	.0120	.0297	.0134	.0206	.0000
TFF1	.0000	.1299	.0032	.2456	.1615	.0005	.1175	.2323	.1540	.0017	.0709	.1328	.2668	.1127	.0500	.1950	.0005
TFF3	.0000	.0279	.0000	.1382	.3563	.0000	.1708	.3722	.0261	.0318	.0719	.1564	.0725	.0019	.2413	.0547	.1485
TG	.0000	.0355	.0099	.0492	.0655	.0000	.0691	.1482	.0778	.0887	.1582	.0215	.0877	.0445	.0560	.0000	.8142
TLE1	.0000	.0385	.1665	.0147	.0724	.0000	.1913	.0174	.0494	.0407	.1724	.0918	.0440	.0458	.2932	.0053	.1212
TMPRSS2	.0000	.0226	.0087	.0828	.1775	.0000	.2887	.1526	.2659	.0407	.1977	.3973	.1369	.1683	.2548	.1761	.0000
TNFRSF8	.0000	.0113	.0137	.0889	.0461	.0000	.0310	.0119	.0652	.0000	.0268	.1567	.0085	.0960	.0070	.0082	.0014
TP63	.0000	.1924	.0006	.2707	.0365	.0000	.1571	.0534	.6012	.0000	.0126	.2757	.0482	.0188	.0035	.0479	.0000
TPM1	.0000	.0159	.0000	.1240	.0292	.0000	.0741	.3391	.0776	.0000	.0453	.0435	.0910	.0000	.2978	.0714	.0000
TPM2	.0000	.0435	.0047	.0348	.0418	.0000	.0327	.0658	.0844	.0159	.0844	.0294	.0107	.0116	.0418	.0531	.0000
TPM3	.0013	.0104	.0079	.0530	.0137	.0000	.0876	.0162	.0559	.0360	.0586	.1213	.0796	.0707	.0705	.0065	.1187
TPM4	.0000	.0306	.0039	.0407	.1157	.0006	.3221	.0346	.1068	.0346	.0870	.2280	.0772	.0650	.0380	.0007	.0055
TPSAB1	.0000	.0685	.0012	.0699	.1828	.0000	.0772	.1892	.0338	.1225	.1826	.0258	.1529	.0686	.0322	.0023	.2542

TTF1	.0002	.0150	.0000	.0049	.0467	.0000	.0502	.1130	.1137	.0795	.0534	.1594	.0845	.0078	.0320	.0128	.0000
UPK2	.0000	.4937	.0294	.0494	.0552	.0000	.0300	.0671	.1641	.0000	.0426	.0210	.0284	.0000	.0000	.1051	.0000
UPK3A	.0000	.2728	.0000	.1923	.0305	.0000	.0340	.1116	.1914	.0000	.0519	.0066	.0172	.2308	.0111	.0000	.0358
UPK3B	.0000	.1254	.0222	.1994	.0554	.0019	.0649	.0380	.0985	.0000	.2264	.0429	.0867	.0255	.0417	.0053	.0575
VHL	.0000	.2155	.0000	.0953	.0091	.0241	.1718	.0635	.0495	.2838	.0118	.4338	.0433	.0115	.0085	.0013	.0022
VIL1	.0000	.2557	.0000	.0205	.3151	.0000	.0469	.3934	.0105	.0000	.7444	.0218	.0261	.0000	.1729	.0023	.0000
VIM	.0000	.2238	.0137	.0638	.0562	.0287	.0547	.0598	.0266	.0709	.0205	.0273	.0512	.0000	.0065	.0421	.2279
WT1	.0000	.0189	.2166	.0572	.0610	.0166	.8319	.1361	.0467	.1979	.0161	.0840	.0163	.0118	.0000	.0108	.0432

Table 120 – RNA Transcripts used to Classify Histology

Transcript	Adeno	ACyC	AC	ACC	Astro	Carc	CS	Chol	CCC	DCIS	GBM	GIST	Gli	GCT	ILC
ACVRL1	0.0303	0.0000	0.0299	0.0000	0.0000	0.0827	0.0117	0.0849	0.0254	0.0643	0.0130	0.1231	0.0104	0.0000	0.1148
AFP	0.0097	0.0001	0.0192	0.0000	0.0000	0.0419	0.0264	0.0589	0.0430	0.1092	0.0732	0.0000	0.0110	0.0000	0.0242
ALPP	0.1621	0.0012	0.0367	0.0000	0.0000	0.0801	0.0955	0.0200	0.0438	0.1049	0.0224	0.0000	0.0323	0.0000	0.0068
AMACR	0.0431	0.0000	0.1815	0.0000	0.0391	0.0957	0.0739	0.0513	0.0544	0.2248	0.0691	0.0000	0.0197	0.0000	0.0738
ANKRD30A	0.0788	0.0000	0.0000	0.0000	0.0000	0.0646	0.0929	0.2001	0.0015	0.5130	0.0620	0.0000	0.0000	0.0000	0.3323
ANO1	0.0398	0.0144	0.0084	0.0000	0.0978	0.0730	0.1301	0.2250	0.0095	0.0309	0.0361	0.4708	0.0000	0.0000	0.0607
ARG1	0.0144	0.0000	0.0133	0.0311	0.0000	0.0591	0.1486	0.2801	0.1504	0.0684	0.0498	0.0000	0.0000	0.0000	0.0948
AR	0.0725	0.0000	0.0192	0.0000	0.1852	0.0345	0.1132	0.0710	0.0476	0.1823	0.1346	0.0000	0.0046	0.0000	0.2347
BCL2	0.0655	0.0067	0.0462	0.0000	0.0000	0.0823	0.0186	0.1332	0.1135	0.1671	0.0424	0.0000	0.0000	0.0000	0.0050
BCL6	0.0785	0.0000	0.0176	0.0000	0.0234	0.1209	0.0273	0.0588	0.0667	0.0772	0.3243	0.0000	0.0028	0.0000	0.2172
CA9	0.0485	0.0000	0.0204	0.0000	0.1205	0.0361	0.0124	0.0523	0.2053	0.0456	0.1995	0.0000	0.0072	0.0000	0.5629
CALB2	0.0304	0.0000	0.0394	0.0998	0.0389	0.0707	0.3244	0.2297	0.1158	0.2715	0.0038	0.0000	0.0000	0.0000	0.0000
CALCA	0.0611	0.0000	0.1202	0.0000	0.0000	0.0254	0.1765	0.0759	0.0249	0.0842	0.0938	0.0000	0.0896	0.0022	0.0022
CALD1	0.0704	0.0186	0.0855	0.0150	0.0247	0.0366	0.2868	0.0325	0.0644	0.0220	0.0130	0.0000	0.0000	0.0000	0.0385
CCND1	0.0283	0.0000	0.1805	0.0000	0.0151	0.0220	0.1704	0.1537	0.0896	0.0739	0.1834	0.0000	0.0086	0.0020	0.0000
CD1A	0.0826	0.0000	0.0207	0.0000	0.0021	0.0186	0.0642	0.1054	0.0014	0.0760	0.0065	0.0000	0.0000	0.0000	0.0629
CD2	0.0517	0.0171	0.0775	0.0000	0.0571	0.0381	0.0423	0.0094	0.0144	0.0879	0.0000	0.0000	0.0000	0.0000	0.0325
CD34	0.0620	0.0000	0.0245	0.0156	0.0000	0.0569	0.0266	0.1230	0.4295	0.0929	0.0294	0.0000	0.0197	0.0000	0.0420
CD3G	0.0755	0.0109	0.1986	0.0000	0.0000	0.0436	0.0356	0.0364	0.0268	0.0741	0.0156	0.0000	0.5012	0.0000	0.0069
CD5	0.0229	0.0000	0.0020	0.0006	0.0000	0.0203	0.1804	0.0810	0.0082	0.1923	0.0162	0.0000	0.0540	0.0000	0.0353
CD79A	0.0278	0.0000	0.0138	0.0000	0.0024	0.0307	0.0384	0.0068	0.0809	0.0982	0.0105	0.0000	0.0057	0.0000	0.2020
CD99L2	0.0447	0.0000	0.1820	0.0000	0.0008	0.1029	0.0336	0.1561	0.0940	0.0767	0.0144	0.0000	0.0070	0.0000	0.0408
CDH17	0.2193	0.0000	0.0227	0.0000	0.0648	0.1989	0.0473	0.0596	0.0393	0.1289	0.0817	0.0000	0.0238	0.0000	0.0769
CDH1	0.1336	0.0165	0.0070	0.1443	0.0031	0.2006	0.3718	0.0454	0.2874	0.2352	0.0000	0.0731	0.0700	0.0000	0.8042
CDK4	0.0521	0.0000	0.0000	0.0000	0.0070	0.0503	0.1631	0.2535	0.0440	0.0260	0.0119	0.0000	0.0064	0.0000	0.2456
CDKN2A	0.0356	0.0000	0.1996	0.0000	0.0064	0.0491	0.3736	0.2100	0.1382	0.3090	0.3358	0.0000	0.0060	0.0000	0.0259
CDX2	0.1164	0.0000	0.0048	0.0000	0.0037	0.0204	0.1191	0.0765	0.0449	0.1066	0.0049	0.0000	0.0000	0.0000	0.0097
CEACAM16	0.0387	0.0002	0.0609	0.0000	0.0283	0.1009	0.0115	0.0250	0.0479	0.0903	0.0223	0.0000	0.0000	0.0000	0.0031
CEACAM18	0.0532	0.0000	0.0050	0.0000	0.0091	0.0418	0.0232	0.0174	0.0000	0.1086	0.0000	0.0000	0.0000	0.0000	0.1954
CEACAM19	0.0363	0.0000	0.0000	0.0000	0.0035	0.0754	0.0971	0.0277	0.0663	0.0993	0.0211	0.0068	0.0273	0.0000	0.0245
CEACAM1	0.1527	0.0074	0.0044	0.0000	0.0022	0.0574	0.0788	0.0648	0.0977	0.0860	0.0928	0.0000	0.2759	0.0000	0.1013
CEACAM20	0.0377	0.0000	0.0000	0.0000	0.0153	0.0530	0.0281	0.0225	0.0200	0.1251	0.0000	0.0000	0.0000	0.0000	0.0000
CEACAM21	0.1119	0.0000	0.0614	0.0000	0.0148	0.0496	0.0103	0.0655	0.0594	0.0656	0.0020	0.0000	0.0000	0.0017	0.0100

CEACAM3	0.0126	0.0000	0.1095	0.0000	0.0083	0.0117	0.0954	0.0167	0.0958	0.0206	0.0041	0.0000	0.0140	0.0000	0.2264
CEACAM4	0.0585	0.0001	0.0748	0.0000	0.0067	0.0434	0.1052	0.1294	0.0256	0.3862	0.1093	0.0000	0.0291	0.0000	0.0356
CEACAM5	0.2644	0.0000	0.0878	0.0000	0.0000	0.2252	0.0000	0.0577	0.0176	0.0468	0.0020	0.0000	0.0000	0.0000	0.0503
CEACAM6	0.0695	0.0006	0.2272	0.0000	0.0512	0.0222	0.1479	0.0090	0.6500	0.1370	0.0667	0.0000	0.0000	0.0000	0.0035
CEACAM7	0.0710	0.0000	0.1835	0.0000	0.0064	0.0430	0.0792	0.0442	0.2010	0.1393	0.0925	0.0000	0.0783	0.0000	0.1301
CEACAM8	0.0413	0.0000	0.0370	0.0000	0.0420	0.0406	0.1021	0.0299	0.0129	0.1021	0.0362	0.0000	0.0187	0.0000	0.0646
CGA	0.0462	0.1722	0.1228	0.0000	0.0000	0.0225	0.0107	0.1993	0.0294	0.0683	0.0290	0.0000	0.0123	0.0000	0.1542
CGB3	0.0420	0.0000	0.0123	0.0000	0.0000	0.0239	0.0085	0.0442	0.0189	0.0653	0.1161	0.0000	0.1370	0.0000	0.0000
CNN1	0.0670	0.0000	0.0621	0.0000	0.2293	0.0791	0.0861	0.1975	0.1542	0.2504	0.0853	0.0000	0.0138	0.0000	0.0000
COQ2	0.0345	0.0000	0.0082	0.0000	0.0752	0.0552	0.2162	0.2841	0.0199	0.0996	0.0551	0.0000	0.0139	0.0000	0.0047
CPS1	0.1298	0.0000	0.1064	0.0000	0.0000	0.0567	0.0904	0.0732	0.1054	0.0776	0.0354	0.0000	0.1078	0.0000	0.0000
CR1	0.0440	0.0000	0.0282	0.0000	0.0167	0.0187	0.0309	0.0020	0.0299	0.2434	0.0791	0.0000	0.0171	0.0000	0.0014
CR2	0.0212	0.0000	0.0000	0.0000	0.0000	0.0638	0.0217	0.0080	0.0734	0.0369	0.0000	0.0000	0.0000	0.0000	0.0037
CTNNB1	0.0433	0.1378	0.0521	0.0000	0.0000	0.0610	0.0276	0.1112	0.0195	0.0428	0.0000	0.0000	0.0000	0.0000	0.0000
DES	0.0884	0.0000	0.0213	0.0000	0.0014	0.0470	0.2483	0.2429	0.0164	0.5792	0.0036	0.0000	0.0137	0.0000	0.0195
DSC3	0.0877	0.0799	0.0000	0.0000	0.0000	0.0274	0.2313	0.0449	0.0321	0.0867	0.0096	0.0000	0.0000	0.0000	0.0160
ENO2	0.0741	0.0143	0.0350	0.0000	0.0024	0.1365	0.0232	0.5293	0.0711	0.1637	0.0794	0.0000	0.0044	0.0000	0.1335
ERBB2	0.1005	0.0000	0.0258	0.0412	0.0198	0.0253	0.0315	0.0116	0.0427	0.0323	0.5524	0.0735	0.0824	0.0000	0.0120
ERG	0.0548	0.0000	0.2395	0.0000	0.0000	0.0462	0.3190	0.0179	0.0246	0.2301	0.1420	0.0000	0.0278	0.0000	0.0068
ESR1	0.0333	0.0009	0.0037	0.0000	0.0000	0.0646	0.0342	0.3642	0.0756	0.0098	0.1072	0.0000	0.0052	0.0000	0.0018
FLI1	0.0259	0.0000	0.0048	0.0000	0.0000	0.0392	0.0362	0.0407	0.0028	0.0791	0.1233	0.0000	0.0037	0.0057	0.0007
FOXL2	0.0762	0.0000	0.1145	0.0000	0.0000	0.0289	0.3640	0.0320	0.3600	0.0396	0.0366	0.0000	0.0377	0.6539	0.1327
FUT4	0.0743	0.0056	0.0634	0.0000	0.0415	0.0893	0.0346	0.4630	0.0605	0.0536	0.0348	0.0051	0.0079	0.0000	0.0000
GATA3	0.1572	0.0009	0.0036	0.0000	0.0000	0.7469	0.2166	0.2601	0.0235	1.4077	0.3759	0.0000	0.0000	0.0000	0.7803
GPC3	0.0279	0.0000	0.2881	0.0000	0.0000	0.0495	0.6239	0.0468	0.1615	0.0378	0.1123	0.0000	0.0234	0.0000	0.0876
HAVCR1	0.0483	0.0000	0.0144	0.0000	0.0153	0.0654	0.0202	0.0321	0.6898	0.2042	0.0000	0.0000	0.0000	0.0000	0.0000
HNF1B	0.3769	0.0000	0.0124	0.0000	0.0000	0.0706	0.0758	0.8381	0.6244	0.7232	0.0002	0.0000	0.0236	0.0000	0.0117
IL12B	0.0237	0.0011	0.0207	0.0000	0.0475	0.1833	0.0388	0.0322	0.0804	0.2427	0.0272	0.0000	0.0172	0.0000	0.0000
IMP3	0.0238	0.0011	0.0028	0.0000	0.0000	0.1225	0.0578	0.0152	0.0263	0.0331	0.0061	0.0016	0.0158	0.0000	0.0000
INHA	0.0326	0.0000	0.0000	0.1810	0.0000	0.0847	0.0851	0.2059	0.0505	0.1237	0.0081	0.0000	0.0000	0.0000	0.0110
ISL1	0.0755	0.0000	0.0028	0.0000	0.0000	0.0349	0.1421	0.1627	0.0118	0.2204	0.1602	0.0035	0.0029	0.0000	0.0507
KIT	0.0648	0.5111	0.0356	0.0000	0.1612	0.0937	0.2800	0.1377	0.0942	0.3399	0.0489	0.0893	0.0092	0.0000	0.0168
KLK3	0.1330	0.0000	0.1582	0.0000	0.0028	0.1167	0.0047	0.1333	0.0067	0.1049	0.0000	0.0000	0.0000	0.0000	0.0753
KL	0.0320	0.0000	0.0000	0.0000	0.0322	0.0506	0.0252	0.3774	0.0197	0.0605	0.0545	0.0000	0.0065	0.0000	0.1088
KRT10	0.0575	0.0000	0.0108	0.0000	0.0267	0.0209	0.0830	0.1563	0.1057	0.1905	0.3030	0.0000	0.0182	0.0000	0.0209
KRT14	0.0295	0.6176	0.1000	0.0000	0.0000	0.0191	0.0449	0.0046	0.0088	0.3260	0.0006	0.0000	0.0032	0.0000	0.0087
KRT15	0.0527	0.0000	0.3800	0.0000	0.0009	0.0292	0.0473	0.1310	0.0185	0.0913	0.4551	0.0000	0.0518	0.0000	0.0377
KRT16	0.0464	0.0000	0.1260	0.0000	0.0511	0.0344	0.0230	0.1396	0.2474	0.0920	0.0738	0.0000	0.0276	0.0000	0.0052
KRT17	0.1360	0.0000	0.0570	0.0000	0.3869	0.0497	0.3012	0.0759	0.0726	0.0562	0.0121	0.0000	0.0000	0.0000	0.0476
KRT18	0.1006	0.0001	0.0054	0.0000	0.0277	0.0447	0.0096	0.2984	0.0196	0.2394	1.2815	0.0018	0.0186	0.1076	0.0000
KRT19	0.0523	0.0000	0.3999	0.0569	0.0000	0.1013	0.1313	0.0238	0.0832	0.1517	0.4445	0.2812	0.0159	0.0000	0.0416
KRT1	0.0590	0.0000	0.0258	0.0000	0.0000	0.0290	0.0220	0.1220	0.0110	0.0128	0.0040	0.0000	0.0000	0.0000	0.0000
KRT20	0.0931	0.0000	0.0706	0.0000	0.0021	0.1631	0.0745	0.2072	0.0214	0.3478	0.1084	0.0000	0.0331	0.0000	0.0055
KRT2	0.0410	0.0000	0.0000	0.0000	0.0038	0.0948	0.1047	0.0125	0.1723	0.0517	0.0133	0.0000	0.0239	0.0000	0.0208

KRT3	0.0379	0.0000	0.0000	0.0000	0.0000	0.0202	0.0249	0.0456	0.2079	0.1026	0.1005	0.0013	0.0082	0.0000	0.0085
KRT4	0.0505	0.0009	0.0787	0.0000	0.0000	0.0499	0.2731	0.0584	0.0950	0.2321	0.0085	0.0000	0.0019	0.0000	0.0107
KRT5	0.3419	0.0000	0.0000	0.0000	0.0000	0.0573	0.0889	0.2456	0.0739	0.1943	0.1791	0.0000	0.0045	0.0000	0.2134
KRT6A	0.1105	0.0000	0.2033	0.0000	0.0000	0.0205	0.0541	0.0918	0.0059	0.0258	0.0872	0.0000	0.0064	0.0000	0.0206
KRT6B	0.0351	0.0000	0.0612	0.0000	0.0000	0.0470	0.6646	0.1217	0.0000	0.2434	0.0028	0.0000	0.0078	0.0000	0.0410
KRT6C	0.0131	0.0000	0.0714	0.0000	0.0000	0.0190	0.0745	0.1042	0.0116	0.0550	0.0000	0.0000	0.0000	0.0000	0.0117
KRT7	0.0993	0.0000	0.0313	0.0000	0.0000	0.1598	0.3404	0.3663	0.0671	0.2393	0.1495	0.0000	0.1437	0.0000	0.3083
KRT8	0.1448	0.0000	0.0008	0.0000	0.3103	0.0998	0.0099	0.0352	0.0267	0.1120	0.6446	0.2529	1.0337	0.0814	0.0243
LIN28A	0.0374	0.0000	0.1733	0.0000	0.0041	0.0323	0.0179	0.0100	0.0049	0.0343	0.0000	0.0000	0.0005	0.0000	0.0000
LIN28B	0.0357	0.0000	0.0093	0.0000	0.0179	0.0839	0.2837	0.0597	0.0123	0.0180	0.0029	0.0000	0.0227	0.0000	0.0061
MAGEA2	0.0035	0.0000	0.0197	0.0000	0.0000	0.0204	0.0069	0.1478	0.0000	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000
MDM2	0.0571	0.0000	0.0294	0.0000	0.0635	0.0405	0.0294	0.3571	0.0681	0.1443	0.0482	0.0000	0.1915	0.0000	0.0020
MIB1	0.0393	0.0184	0.0401	0.1948	0.0000	0.0171	0.1304	0.0378	0.1385	0.1610	0.0167	0.0000	0.2388	0.0000	0.0733
MITF	0.0699	0.0000	0.0173	0.0000	0.0013	0.3192	0.0583	0.2196	0.3497	0.1355	0.0262	0.0000	0.0000	0.0000	0.0183
MLANA	0.0447	0.0000	0.0127	0.0000	0.0179	0.0565	0.1727	0.0166	0.0494	0.0200	0.0566	0.0000	0.0248	0.0000	0.0527
MLH1	0.0607	0.0000	0.0142	0.0000	0.0000	0.0451	0.1695	0.4392	0.2528	0.0188	0.0000	0.0000	0.0110	0.0000	0.0000
MME	0.0285	0.0000	0.0186	0.0000	0.0015	0.0381	0.3911	0.0668	0.0968	0.5786	0.0026	0.0000	0.0009	0.0119	0.2762
MPO	0.0443	0.0000	0.0084	0.0000	0.0043	0.0538	0.0064	0.1377	0.0221	0.0417	0.0000	0.0000	0.0262	0.0000	0.0477
MS4A1	0.0791	0.0011	0.2588	0.0000	0.0000	0.0784	0.1161	0.0195	0.0032	0.1795	0.0705	0.0000	0.0429	0.0000	0.0398
MSH2	0.0443	0.0000	0.0045	0.0000	0.0937	0.0650	0.0930	0.1603	0.1040	0.0834	0.0324	0.0000	0.0000	0.0000	0.0000
MSH6	0.0980	0.0000	0.0087	0.0000	0.0595	0.0347	0.0549	0.0329	0.0048	0.0808	0.0000	0.0000	0.0017	0.1466	0.0150
MSLN	0.1086	0.0000	0.0503	0.0007	0.0053	0.0995	0.4299	0.1498	0.0399	0.1063	0.0000	0.0000	0.0123	0.0000	0.0145
MTHFR	0.0881	0.0000	0.0699	0.0000	0.0054	0.1041	0.0713	0.0333	0.0408	0.0240	0.0865	0.0000	0.0006	0.0000	0.0979
MUC1	0.2924	0.0000	0.0180	0.0347	0.4498	0.0514	0.4092	0.1764	0.0989	0.1107	0.1503	0.2889	0.0000	0.0000	0.4940
MUC2	0.0353	0.0000	0.0754	0.0000	0.0000	0.0332	0.0638	0.1168	0.0550	0.0935	0.0030	0.0000	0.0397	0.0000	0.0071
MUC4	0.0366	0.0000	0.0051	0.0000	0.0007	0.0656	0.0282	0.4620	0.0344	0.3633	0.0035	0.0000	0.0000	0.0000	0.3175
MUC5AC	0.2451	0.0001	0.0000	0.0000	0.0187	0.2406	0.0232	0.1563	0.0342	0.0897	0.0062	0.0000	0.0000	0.0000	0.0047
MYOD1	0.0305	0.0000	0.0210	0.0000	0.0029	0.0185	0.0467	0.0214	0.0648	0.2351	0.0000	0.0000	0.0004	0.0000	0.0149
MYOG	0.0455	0.0000	0.0067	0.0000	0.0000	0.0320	0.1141	0.0112	0.3825	0.0447	0.0083	0.0000	0.0023	0.0000	0.0000
NANOG	0.0626	0.0008	0.0000	0.0000	0.0366	0.0890	0.0342	0.0827	0.0213	0.1847	0.0063	0.0000	0.0050	0.0000	0.0068
NAPSA	0.0778	0.0000	0.3319	0.0000	0.0264	0.0897	0.2899	0.1382	0.5083	0.1269	0.0075	0.0000	0.0112	0.0000	0.1109
NCAM1	0.0416	0.0000	0.0090	0.0000	0.8230	0.0815	0.1464	0.0515	0.0815	0.3384	0.6458	0.0000	0.1516	0.0000	0.0333
NCAM2	0.0301	0.0001	0.1840	0.0000	0.0159	0.0380	0.0101	0.0125	0.0482	0.4548	0.0177	0.0000	0.5388	0.0000	0.1293
NKX2-2	0.0956	0.0001	0.0132	0.0000	0.0423	0.1316	0.0206	0.4682	0.0287	0.0153	0.8243	0.0000	0.0000	0.0000	0.0526
NKX3-1	0.0973	0.0000	0.0531	0.0928	0.0208	0.0685	0.0220	0.0607	0.1823	0.3601	0.0108	0.0000	0.0204	0.0000	0.3430
OSCAR	0.0590	0.0000	0.4226	0.0000	0.2128	0.0372	0.1323	0.0883	0.0846	0.0841	0.0027	0.0000	0.0058	0.0000	0.3083
PAX2	0.0508	0.0000	0.0000	0.0000	0.0012	0.0661	0.0235	0.0025	0.0700	0.0779	0.0022	0.0000	0.0000	0.0000	0.1699
PAX5	0.0361	0.0011	0.0453	0.0000	0.0000	0.1033	0.1375	0.0562	0.0045	0.0351	0.0478	0.0000	0.0164	0.0000	0.0013
PAX8	0.0266	0.0000	0.1035	0.0000	0.0000	0.0576	0.2124	0.0975	0.5638	0.4051	0.1016	0.0000	0.0060	0.0000	0.0566
PDPN	0.0517	0.0002	0.1428	0.0000	0.0000	0.2347	0.0552	0.0881	0.0134	0.0517	0.8837	0.0000	0.0921	0.0000	0.0036
PDX1	0.1379	0.0000	0.0300	0.0000	0.0000	0.0138	0.2562	0.0455	0.1878	0.0341	0.0240	0.0000	0.0000	0.0000	0.0476
PECAM1	0.0456	0.0000	0.0281	0.0000	0.0000	0.1047	0.1991	0.0221	0.0164	0.0408	0.0442	0.0000	0.0010	0.0000	0.0122
PGR	0.1144	0.0000	0.0000	0.0000	0.0814	0.0904	0.3056	0.0105	0.0577	0.0548	0.0138	0.0000	0.0000	0.0000	0.0277
PIP	0.0782	0.0000	0.1859	0.0000	0.0060	0.0669	0.0364	0.0588	0.0512	0.3791	0.0476	0.0000	0.0566	0.0000	0.0037

PMEL	0.0237	0.0000	0.0722	0.0004	0.0031	0.1230	0.0154	0.0278	0.0402	0.0637	0.1061	0.0000	0.0644	0.0000	0.0205
PMS2	0.0263	0.0000	0.0082	0.0000	0.0036	0.0330	0.0100	0.0652	0.1249	0.0776	0.0003	0.0000	0.0139	0.0000	0.0000
POU5F1	0.0513	0.0000	0.0469	0.0000	0.0253	0.0651	0.0310	0.2375	1.0489	0.0274	0.0899	0.0000	0.2486	0.0000	0.0000
PSAP	0.0563	0.0000	0.0986	0.0000	0.0014	0.0484	0.0258	0.0861	0.0767	0.0328	0.0000	0.0000	0.0013	0.0000	0.0006
PTPRC	0.0406	0.0000	0.0018	0.0000	0.0395	0.0291	0.0029	0.0682	0.0882	0.0180	0.0054	0.0008	0.0000	0.0000	0.0000
S100A10	0.0953	0.0007	0.0043	0.0007	0.0120	0.0737	0.0519	0.0085	0.0443	0.0282	0.0583	0.0010	0.0000	0.0000	0.0420
S100A11	0.0415	0.0000	0.0359	0.0000	0.0946	0.0492	0.0923	0.0226	0.0177	0.2103	0.1027	0.0000	0.0000	0.0009	0.0000
S100A12	0.0990	0.0000	0.2534	0.0000	0.0016	0.0337	0.0676	0.1337	0.1261	0.2927	0.0027	0.0000	0.0000	0.0000	0.0052
S100A13	0.0627	0.0000	0.0092	0.0000	0.0072	0.0473	0.0561	0.0384	0.0495	0.0449	0.0176	0.0037	0.0179	0.0000	0.0598
S100A14	0.0916	0.0000	0.0077	0.0000	0.0000	0.0551	0.0570	0.0609	0.3262	0.0332	0.3067	0.0000	0.0543	0.0000	0.0104
S100A16	0.0103	0.0000	0.0244	0.0000	0.0124	0.0251	0.1989	0.0028	0.0133	0.0157	0.0051	0.0045	0.0269	0.0000	0.0115
S100A1	0.1471	0.0000	0.0347	0.0000	0.2960	0.1011	0.0759	0.0283	0.1372	0.0820	0.0123	0.0011	0.0506	0.0000	0.7448
S100A2	0.1293	0.0000	0.0024	0.0000	0.0101	0.0448	0.4043	0.2608	0.0354	0.3199	0.0757	0.0000	0.0402	0.0000	0.0000
S100A4	0.0814	0.0018	0.0184	0.0000	0.4240	0.0280	0.2036	0.0107	0.0383	0.0648	0.0067	0.0000	0.0003	0.0000	0.0123
S100A5	0.0915	0.0000	0.0052	0.0000	0.0000	0.1135	0.0383	0.0445	0.1217	0.0388	0.0045	0.0000	0.0000	0.0000	0.3229
S100A6	0.0433	0.0778	0.0276	0.0000	0.0078	0.0550	0.4067	0.0420	0.1706	0.0491	0.0004	0.0000	0.0000	0.0000	0.0025
S100A7A	0.0955	0.0000	0.0000	0.0000	0.0000	0.0572	0.0462	0.0593	0.0674	0.0408	0.0196	0.0000	0.0000	0.0000	0.0525
S100A7L2	0.0353	0.0000	0.0000	0.0000	0.0000	0.0207	0.0056	0.0110	0.1647	0.1410	0.0474	0.0000	0.0000	0.0000	0.0014
S100A7	0.0833	0.0000	0.0596	0.0000	0.0000	0.0707	0.0636	0.1336	0.0364	0.1516	0.0000	0.0000	0.0000	0.0000	0.0062
S100A8	0.0547	0.0000	0.0036	0.0000	0.0000	0.1201	0.0045	0.1331	0.0457	0.1995	0.0874	0.0000	0.0071	0.0000	0.0051
S100A9	0.0607	0.0000	0.0135	0.0008	0.1144	0.0552	0.1603	0.1628	0.3308	0.0883	0.0865	0.0023	0.0113	0.0029	0.1154
S100B	0.0969	0.0000	0.0000	0.0000	1.2677	0.0487	0.1932	0.2718	0.0452	0.0153	1.3235	0.0000	0.8497	0.0020	0.0131
S100PBP	0.0573	0.0000	0.0105	0.0000	0.0020	0.0875	0.0399	0.0838	0.1370	0.1267	0.0091	0.0000	0.0000	0.0000	0.0000
S100P	0.0563	0.0000	0.0245	0.0000	0.0000	0.1691	0.0412	0.0962	0.3398	0.1459	0.0278	0.0000	0.0000	0.0000	0.0614
S100Z	0.0297	0.0000	0.0153	0.0000	0.0000	0.0196	0.1191	0.0282	0.3076	0.0134	0.0298	0.0000	0.0163	0.0000	0.0546
SALL4	0.0262	0.0000	0.0478	0.0000	0.1795	0.0298	0.0753	0.0297	0.0643	0.1220	0.1034	0.0000	0.0000	0.0000	0.0172
SATB2	0.0706	0.0000	0.0162	0.0000	0.0051	0.0423	0.0309	0.1550	0.0932	0.4879	0.0171	0.0000	0.2276	0.0000	0.0178
SDC1	0.0380	0.0006	0.0485	0.0003	0.1795	0.1022	0.0254	0.1856	0.0363	0.2517	0.1621	0.4088	0.4023	0.3116	0.0428
SERPINA1	0.1070	0.0000	0.2130	0.0000	0.0000	0.1024	0.2714	0.9927	0.0186	0.3578	0.0056	0.0000	0.0000	0.0011	0.2646
SERPINB5	0.0612	0.0000	0.0086	0.0000	0.0000	0.0605	0.0455	0.0930	0.1141	0.1290	0.0113	0.0000	0.0000	0.0000	0.1706
SF1	0.0271	0.0000	0.0000	0.0000	0.0000	0.0837	0.0073	0.1912	0.0991	0.0312	0.2400	0.0000	0.0029	0.0000	0.0095
SFTPA1	0.0546	0.0000	0.6110	0.0000	0.1626	0.0961	0.3220	0.3272	0.1281	0.2402	0.1506	0.0000	0.0000	0.0008	0.1089
SMAD4	0.0481	0.1555	0.0372	0.0000	0.0013	0.0814	0.0000	0.1728	0.0350	0.1275	0.0374	0.0000	0.0000	0.0000	0.0071
SMARCB1	0.0425	0.0000	0.0000	0.0000	0.0065	0.0810	0.1929	0.0100	0.0531	0.0912	0.1776	0.0000	0.0000	0.0000	0.0120
SMN1	0.0542	0.0003	0.0772	0.0000	0.1768	0.0509	0.0372	0.3121	0.0172	0.0351	0.0000	0.0000	0.0000	0.0000	0.0000
SOX2	0.0542	0.0001	0.2163	0.0000	0.8539	0.0592	0.1296	0.1575	0.0550	0.4843	0.8152	0.0000	0.3863	0.0000	0.3317
SPN	0.0240	0.0000	0.0039	0.0000	0.0026	0.1516	0.0569	0.0418	0.0289	0.1275	0.0449	0.0000	0.0405	0.0000	0.0276
SYP	0.0838	0.0000	0.1574	0.1257	0.0000	0.0658	0.0040	0.0746	0.2606	0.1050	0.0155	0.0000	0.6098	0.0000	0.0100
TFE3	0.0203	0.0000	0.0000	0.0000	0.0000	0.0098	0.0412	0.1226	0.0350	0.0896	0.0024	0.0000	0.0000	0.0000	0.0000
TFF1	0.0448	0.0000	0.0000	0.0000	0.0000	0.1024	0.0123	0.7223	0.0839	0.1383	0.0864	0.0000	0.0421	0.0000	0.0227
TFF3	0.1486	0.0001	0.0340	0.0000	0.1101	0.0959	0.0123	0.1150	0.0679	0.1779	0.0482	0.0049	0.0000	0.0000	0.6256
TG	0.0923	0.0000	0.1325	0.0000	0.0000	0.0819	0.0249	0.0615	0.0465	0.0063	0.0981	0.0000	0.0000	0.0000	0.0072
TLE1	0.0352	0.0000	0.0000	0.0000	0.0276	0.0495	0.1203	0.1772	0.0407	0.1247	0.0082	0.0000	0.0082	0.0016	0.0541
TMPRSS2	0.6698	0.0000	0.0000	0.0000	0.0628	0.1438	0.0027	0.4135	0.0487	0.0494	0.0522	0.0000	0.0000	0.0000	0.0068

TNFRSF8	0.0267	0.0000	0.0064	0.0000	0.0000	0.0290	0.0114	0.0934	0.0251	0.0364	0.0040	0.0000	0.0784	0.0000	0.0925
TP63	0.1645	0.0611	0.6474	0.0000	0.0004	0.0343	0.0290	0.0225	0.0170	0.1422	0.0203	0.0000	0.0000	0.0000	0.0000
TPM1	0.0811	0.0224	0.0156	0.0000	0.0401	0.0421	0.0915	0.1594	0.0846	0.0519	0.0831	0.0000	0.0137	0.0000	0.0101
TPM2	0.0292	0.0089	0.0279	0.0000	0.2139	0.0753	0.2048	0.0287	0.0740	0.0239	0.0061	0.0000	0.0000	0.0000	0.0000
TPM3	0.0646	0.3315	0.1448	0.0000	0.0037	0.0271	0.0915	0.0435	0.1476	0.2891	0.0445	0.0000	0.0235	0.0000	0.0117
TPM4	0.0898	0.0015	0.0308	0.0000	0.2819	0.0630	0.0354	0.0467	0.0585	0.1126	0.0038	0.0000	0.0072	0.0000	0.0104
TPSAB1	0.0366	0.0000	0.0804	0.0000	0.0000	0.1052	0.2333	0.0450	0.1244	0.2030	0.0252	0.0020	0.0000	0.0000	0.1027
TTF1	0.0242	0.0000	0.0763	0.0000	0.0080	0.0191	0.0685	0.0046	0.2690	0.1715	0.0785	0.0000	0.0133	0.0000	0.0036
UPK2	0.1191	0.0000	0.0033	0.0000	0.0588	0.0950	0.0166	0.0254	0.0105	0.1552	0.0215	0.0000	0.0000	0.0000	0.0628
UPK3A	0.0580	0.0000	0.0000	0.0000	0.0145	0.0630	0.0643	0.0643	0.0170	0.0860	0.2445	0.0000	0.0067	0.0000	0.0503
UPK3B	0.0462	0.0000	0.0441	0.0000	0.0000	0.0721	0.0469	0.2848	0.1285	0.2996	0.0280	0.0000	0.0380	0.0000	0.0516
VHL	0.0547	0.0000	0.2177	0.0000	0.0000	0.0370	0.0286	0.1825	0.0086	0.0334	0.0041	0.0000	0.0183	0.0000	0.0035
VIL1	0.0791	0.0000	0.0405	0.0000	0.0034	0.2266	0.1460	0.8138	0.1260	0.0962	0.0055	0.0000	0.0000	0.0000	0.0991
VIM	0.0264	0.0030	0.0154	0.0287	0.0069	0.0364	0.0376	0.0135	0.0362	0.1135	0.0432	0.0000	0.0094	0.0000	0.1413
WT1	0.0351	0.0000	0.1805	0.0000	0.0189	0.0552	0.1780	0.4010	0.3054	0.2016	0.0114	0.0000	0.0030	0.0000	0.0432

Table 120 ***continued***

Transcript	Lei	Lipo	Mel	Men	Merk	Meso	Neuro	NSCC	Oligo	Sarc	SerC	Serous	SCC	Sq
ACVRL1	0.0000	0.0194	0.1326	0.0000	0.0000	0.0000	0.0000	0.0702	0.0000	0.0771	0.0000	0.4134	0.0040	0.0337
AFP	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0.0253	0.0001	0.0000	0.0038	0.0198	0.0000	0.0648
ALPP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0892	0.0000	0.0037	0.0000	0.2362	0.0062	0.0440
AMACR	0.0000	0.0083	0.0000	0.0000	0.0000	0.0006	0.0021	0.0446	0.0000	0.0000	0.0182	0.0705	0.0106	0.0517
ANKRD30A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0413	0.2199	0.0001	0.0020	0.0061	0.0338	0.0000	0.0988
ANO1	0.0346	0.0000	0.0191	0.2936	0.0000	0.0000	0.0266	0.0683	0.0000	0.0035	0.0000	0.3164	0.1499	0.1244
ARG1	0.0000	0.0000	0.0540	0.0000	0.0000	0.0000	0.0820	0.1353	0.0000	0.0129	0.0371	0.2312	0.0000	0.0600
AR	0.1166	0.0000	0.1381	0.0104	0.0000	0.0000	0.0989	0.3680	0.0013	0.0611	0.0000	0.3377	0.0000	0.5690
BCL2	0.0000	0.0000	0.0118	0.0023	0.0000	0.0000	0.0024	0.1045	0.0098	0.0750	0.0031	0.0690	0.2242	0.0549
BCL6	0.0945	0.0000	0.0944	0.0137	0.0000	0.0000	0.0009	0.1674	0.0000	0.0081	0.0000	0.0433	0.0000	0.0086
CA9	0.0017	0.0000	0.0090	0.0000	0.0037	0.0218	0.0104	0.0924	0.0000	0.1524	0.0434	0.0773	0.1230	0.1082
CALB2	0.2303	0.0000	0.0005	0.0000	0.0000	0.5584	0.0008	0.0728	0.0000	0.0028	0.0020	0.0507	0.0324	0.0603
CALCA	0.0113	0.0000	0.0110	0.0087	0.0000	0.0000	0.0089	0.0900	0.0110	0.0156	0.0000	0.0275	0.1383	0.0353
CALD1	0.1347	0.0000	0.0000	0.0022	0.0000	0.0000	0.0000	0.0849	0.0000	0.2135	0.0026	0.0323	0.0000	0.0252
CCND1	0.0783	0.0005	0.0871	0.0379	0.0010	0.0000	0.0163	0.0786	0.0000	0.0278	0.0061	0.0941	0.0681	0.0925
CD1A	0.0080	0.0000	0.0195	0.0000	0.0000	0.0000	0.0000	0.0402	0.0000	0.0021	0.0130	0.0628	0.0456	0.0585
CD2	0.1357	0.0000	0.0781	0.0056	0.0000	0.0000	0.0239	0.0885	0.4549	0.0000	0.0016	0.0645	0.0235	0.0578
CD34	0.0239	0.0701	0.0000	0.0000	0.0000	0.0019	0.0130	0.0189	0.0016	0.0077	0.0022	0.1071	0.1177	0.1263
CD3G	0.0000	0.0003	0.0512	0.0000	0.0000	0.0000	0.0590	0.0867	0.0000	0.0790	0.0396	0.0868	0.0454	0.5591
CD5	0.0000	0.0000	0.0103	0.1699	0.0000	0.0000	0.0341	0.0347	0.0000	0.0020	0.0335	0.0627	0.0235	0.0750
CD79A	0.2340	0.0000	0.0969	0.0000	0.0000	0.0000	0.0000	0.1930	0.0334	0.0199	0.0000	0.1609	0.0175	0.0902
CD99L2	0.0032	0.0000	0.0209	0.0084	0.0000	0.0026	0.0029	0.0775	0.0343	0.0052	0.3332	0.1470	0.0261	0.0884
CDH17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0237	0.0704	0.0000	0.0186	0.0334	0.0384	0.0621	0.1226
CDH1	0.1206	0.2631	0.0000	0.1095	0.0000	0.0099	0.0000	0.0216	0.2687	0.0658	0.1951	0.1450	0.0053	0.0934
CDK4	0.0000	0.3028	0.0000	0.0000	0.0000	0.0006	0.0000	0.1002	0.0000	0.0002	0.0169	0.3539	0.0000	0.1079
CDKN2A	0.0000	0.0000	0.1460	0.0000	0.0000	0.0074	0.0324	0.1523	0.0000	0.1410	0.0978	0.5257	0.0393	0.0527

CDX2	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0088	0.0826	0.0010	0.0000	0.0219	0.2185	0.0013	0.0904
CEACAM16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.2136	0.0000	0.0016	0.0000	0.0791	0.0925	0.0515
CEACAM18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0073	0.0112	0.0415	0.0103	0.0077	0.0333	0.0223	0.0057	0.0827
CEACAM19	0.0617	0.0000	0.1690	0.0000	0.0000	0.0000	0.0619	0.0226	0.0000	0.1683	0.0056	0.1586	0.1520	0.1541
CEACAM1	0.0655	0.0004	0.0912	0.2840	0.0000	0.0387	0.0000	0.1772	0.1025	0.0060	0.1514	0.1488	0.0070	0.0627
CEACAM20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2582	0.0000	0.0044	0.0000	0.0307	0.0402	0.0383
CEACAM21	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000	0.0022	0.0596	0.0000	0.0089	0.0005	0.1190	0.0857	0.0604
CEACAM3	0.0000	0.0000	0.0107	0.0000	0.0000	0.0817	0.0578	0.1906	0.0000	0.0162	0.0000	0.2166	0.0070	0.0680
CEACAM4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0522	0.0429	0.0054	0.0000	0.0081	0.0275	0.0000	0.0212
CEACAM5	0.0000	0.0081	0.0028	0.0026	0.0147	0.0000	0.1568	0.0377	0.0000	0.0662	0.0711	0.1794	0.0455	0.0328
CEACAM6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0276	0.1025	0.0000	0.0069	0.0255	0.1754	0.0067	0.0508
CEACAM7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.2715	0.0000	0.0200	0.0000	0.0211	0.0000	0.0243
CEACAM8	0.0000	0.0007	0.0091	0.0000	0.0000	0.0000	0.0246	0.0523	0.0023	0.0235	0.0000	0.0688	0.0260	0.1095
CGA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0453	0.0756	0.0000	0.0000	0.0000	0.1266	0.1477	0.0620
CGB3	0.0000	0.0000	0.0748	0.0000	0.0000	0.0000	0.0430	0.0694	0.0000	0.0000	0.0128	0.0323	0.1818	0.1826
CNN1	0.4602	0.0000	0.0000	0.0000	0.0000	0.0000	0.0333	0.1607	0.0000	0.0000	0.0035	0.0938	0.0141	0.2457
COQ2	0.0199	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1271	0.0404	0.0000	0.0117	0.0425	0.0095	0.0577
CPS1	0.0615	0.0000	0.1500	0.0000	0.0603	0.0000	0.0096	0.0797	0.0000	0.0156	0.2381	0.2112	0.0068	0.1204
CR1	0.0067	0.0328	0.0000	0.0013	0.0295	0.0000	0.0087	0.0211	0.0000	0.0000	0.0369	0.0407	0.0000	0.1642
CR2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0648	0.0000	0.0408	0.0000	0.2135	0.0054	0.0319
CTNNB1	0.0004	0.0000	0.0195	0.0000	0.0000	0.0000	0.0031	0.2061	0.0000	0.0000	0.0025	0.0811	0.4604	0.1853
DES	0.2105	0.0000	0.0000	0.0000	0.0000	0.0000	0.0759	0.0584	0.0000	0.0169	0.0077	0.1431	0.0023	0.2380
DSC3	0.0021	0.0017	0.0212	0.0409	0.0000	0.0060	0.0189	0.0266	0.0001	0.0986	0.0000	0.3496	0.0000	0.4745
ENO2	0.1487	0.0014	0.0196	0.0000	0.0005	0.0000	0.3925	0.2998	0.0000	0.0869	0.0156	0.1923	0.0020	0.0446
ERBB2	0.1595	0.0000	0.0139	0.0000	0.2850	0.0000	0.2159	0.1602	0.0000	0.0000	0.0998	0.0337	0.0695	0.0392
ERG	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0189	0.0739	0.0181	0.0000	0.0000	0.0666	0.0000	0.1302
ESR1	0.0156	0.0027	0.0592	0.0011	0.0000	0.0000	0.2086	0.4605	0.0000	0.0164	0.0000	0.2626	0.0044	0.1409
FLI1	0.0000	0.0000	0.0007	0.0000	0.0000	0.0017	0.0043	0.1105	0.0000	0.0703	0.0009	0.0206	0.0145	0.0784
FOXL2	0.3188	0.0000	0.0000	0.0086	0.0000	0.0000	0.0000	0.1655	0.0048	0.0848	0.0222	0.2622	0.0000	0.1393
FUT4	0.0064	0.0000	0.0090	0.0000	0.0000	0.0000	0.0000	0.2052	0.0102	0.0115	0.0000	0.0738	0.0536	0.1795
GATA3	0.0000	0.0000	0.0000	0.0355	0.0000	0.0027	0.0000	0.2180	0.0000	0.0000	0.0086	0.0616	0.0000	0.2132
GPC3	0.0002	0.0004	0.0907	0.0000	0.0000	0.0000	0.0179	0.0852	0.0002	0.0000	0.0038	0.0770	0.0000	0.0689
HAVCR1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	0.1343	0.0000	0.0114	0.0008	0.0647	0.0820	0.2677
HNF1B	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0600	0.0007	0.0314	0.0169	0.2549	0.0000	0.3320
IL12B	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.0032	0.1805	0.0000	0.0000	0.1007	0.0838	0.0032	0.0147
IMP3	0.0335	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.0119	0.0000	0.0249	0.1609	0.2859	0.0025	0.2011
INHA	0.0026	0.0000	0.1065	0.0078	0.0000	0.0449	0.0543	0.2378	0.0313	0.0000	0.0021	0.0268	0.0710	0.0468
ISL1	0.0225	0.0000	0.0179	0.0000	0.2910	0.0000	0.6480	0.2721	0.0016	0.0000	0.0000	0.1192	0.6379	0.0354
KIT	0.0202	0.0039	0.0098	0.0025	0.0000	0.0000	0.0068	0.0719	0.0000	0.0059	0.0000	0.0714	0.5444	0.0694
KLK3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0116	0.1098	0.0000	0.0000	0.0000	0.1166	0.0390	0.0410
KL	0.0022	0.0009	0.0000	0.0007	0.0000	0.0000	0.0136	0.0578	0.0000	0.0000	0.0806	0.0659	0.1887	0.0594
KRT10	0.0000	0.0000	0.1388	0.2300	0.0025	0.0000	0.0289	0.1095	0.0000	0.0000	0.0346	0.0197	0.0045	0.0588
KRT14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0250	0.2027	0.0000	0.0085	0.0104	0.0400	0.0579	0.1112
KRT15	0.0000	0.0013	0.0106	0.0000	0.0000	0.0000	0.0298	0.0779	0.0186	0.1461	0.1244	0.2614	0.0476	0.0824

KRT16	0.0658	0.0000	0.0000	0.0628	0.0000	0.0000	0.0000	0.0400	0.0000	0.0000	0.0000	0.1296	0.0104	0.0396
KRT17	0.0025	0.0000	0.0662	0.0000	0.0000	0.0000	0.0051	0.0572	0.0021	0.0097	0.0000	0.1598	0.0181	0.8321
KRT18	0.7156	0.5117	0.1018	0.0000	0.0000	0.0000	0.0049	0.1243	0.7509	0.0054	0.0005	0.0210	0.0000	0.0879
KRT19	1.2857	0.2603	0.7118	0.0000	0.0000	0.0000	0.0560	0.0352	0.0000	0.8934	0.0009	0.0659	0.0677	0.1021
KRT1	0.0000	0.0000	0.0207	0.0000	0.0000	0.0000	0.0000	0.0879	0.0000	0.0370	0.0000	0.2108	0.0062	0.0187
KRT20	0.0000	0.0000	0.0000	0.0020	0.0000	0.0008	0.0000	0.0449	0.0036	0.0000	0.0000	0.0337	0.0586	0.2718
KRT2	0.1623	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.1053	0.0000	0.2684	0.0000	0.0523	0.0000	0.1150
KRT3	0.0212	0.0000	0.0000	0.0000	0.0000	0.0002	0.0049	0.1919	0.0010	0.0000	0.0014	0.1282	0.0000	0.0591
KRT4	0.0023	0.0000	0.0072	0.0079	0.0000	0.0000	0.0106	0.1192	0.0000	0.0000	0.0067	0.2677	0.0000	0.0307
KRT5	0.0000	0.0000	0.0000	0.0000	0.0000	0.1402	0.0000	0.1377	0.0000	0.0000	0.0238	0.1224	0.1361	0.8787
KRT6A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1167	0.0000	0.0000	0.0004	0.0457	0.1171	0.5259
KRT6B	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1034	0.0000	0.0000	0.0000	0.2588	0.0066	0.1718
KRT6C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0685	0.0000	0.0330	0.0000	0.1959	0.0000	0.1249
KRT7	0.0195	0.1825	0.0000	0.0083	0.0494	0.0006	0.0120	0.0605	0.0000	0.2594	0.0054	0.5886	0.0162	0.2365
KRT8	0.7388	0.0129	0.6362	0.5124	0.0000	0.0000	0.0116	0.0870	0.0000	0.0137	0.0064	0.1210	0.0000	0.0509
LIN28A	0.0000	0.0065	0.1182	0.0000	0.0000	0.0000	0.0313	0.0317	0.0000	0.0203	0.0066	0.1835	0.0043	0.0266
LIN28B	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0344	0.0430	0.0000	0.0000	0.0000	0.0736	0.0036	0.1618
MAGEA2	0.0000	0.0000	0.0000	0.0138	0.0000	0.0000	0.0000	0.2146	0.0000	0.0000	0.0000	0.0097	0.0025	0.0028
MDM2	0.0218	0.3254	0.0036	0.0294	0.0000	0.0000	0.0171	0.1187	0.0000	0.0032	0.0700	0.1588	0.0072	0.0718
MIB1	0.0000	0.0000	0.0108	0.0000	0.0000	0.0000	0.0000	0.0455	0.0000	0.0000	0.0285	0.0891	0.0040	0.0089
MITF	0.1166	0.0000	0.2020	0.0175	0.0000	0.0000	0.0316	0.1076	0.0000	0.0000	0.0378	0.0334	0.3685	0.0255
MLANA	0.0067	0.0000	0.4617	0.0000	0.0005	0.0000	0.0000	0.0703	0.0027	0.0006	0.0000	0.1913	0.0330	0.0778
MLH1	0.0773	0.0000	0.0000	0.0000	0.0000	0.0000	0.0149	0.0573	0.0229	0.0005	0.0154	0.1703	0.0063	0.0200
MME	0.0000	0.0132	0.0006	0.0038	0.0944	0.0000	0.0034	0.1307	0.0000	0.0780	0.5287	0.1239	0.1573	0.0488
MPO	0.0000	0.0000	0.0000	0.0121	0.0000	0.0000	0.1090	0.0260	0.0000	0.0039	0.0736	0.0854	0.0465	0.0205
MS4A1	0.0000	0.0003	0.0924	0.0000	0.0000	0.0000	0.0388	0.0339	0.0000	0.0048	0.0010	0.0097	0.0267	0.0285
MSH2	0.0042	0.0007	0.0000	0.2136	0.0000	0.0067	0.0000	0.0991	0.0037	0.0239	0.0013	0.0607	0.0933	0.2618
MSH6	0.0165	0.0000	0.0000	0.0000	0.0000	0.0000	0.0319	0.0930	0.0048	0.0028	0.0024	0.0959	0.0120	0.1485
MSLN	0.0011	0.0003	0.0390	0.0048	0.0005	0.1462	0.0000	0.3377	0.0000	0.0000	0.2129	0.4918	0.2586	0.0372
MTHFR	0.0008	0.0000	0.0619	0.0000	0.0000	0.0000	0.0534	0.0806	0.0000	0.0000	0.0039	0.0644	0.0538	0.1563
MUC1	0.0166	0.0000	0.5181	0.0000	0.0000	0.0000	0.2996	0.1200	0.0000	0.0000	0.0016	0.0753	0.4778	0.0987
MUC2	0.0000	0.0000	0.0058	0.0000	0.0000	0.0080	0.0000	0.2272	0.0001	0.0081	0.0000	0.1580	0.0071	0.1316
MUC4	0.0105	0.0000	0.0000	0.0184	0.0053	0.0000	0.1225	0.0448	0.0000	0.0564	0.0143	0.1906	0.5281	0.1882
MUC5AC	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0085	0.0686	0.0000	0.0041	0.0000	0.1796	0.0208	0.0524
MYOD1	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.1587	0.0000	0.0480	0.0000	0.0310	0.0159	0.0153
MYOG	0.0286	0.0000	0.0519	0.0000	0.0744	0.0000	0.0084	0.1007	0.0000	0.2284	0.0000	0.0937	0.0000	0.0954
NANOG	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.0052	0.1241	0.0000	0.0245	0.0302	0.1074	0.0000	0.0590
NAPSA	0.0000	0.0000	0.0036	0.0047	0.0004	0.0000	0.0748	0.0731	0.0000	0.0024	0.1033	0.1671	0.0175	0.0281
NCAM1	0.1329	0.0008	0.0514	0.0000	0.0000	0.0000	0.5313	0.2375	0.8634	1.0584	0.0003	0.0514	1.5638	0.0364
NCAM2	0.0000	0.0000	0.0456	0.0000	0.0000	0.0000	0.0175	0.1092	0.0062	0.0237	0.1308	0.0401	0.0045	0.1502
NKX2-2	0.0109	0.0037	0.0122	0.0000	0.0000	0.0000	0.0891	0.0926	0.0000	0.3744	0.0181	0.1279	0.3525	0.0191
NKX3-1	0.0126	0.0000	0.0000	0.0000	0.0000	0.0000	0.0107	0.0656	0.0069	0.0176	0.2486	0.0740	0.0146	0.0173
OSCAR	0.0000	0.0071	0.0072	0.0000	0.0000	0.0000	0.0126	0.1076	0.0000	0.0319	0.1949	0.0401	0.0000	0.1076
PAX2	0.0000	0.0000	0.0003	0.0003	0.0000	0.0000	0.0000	0.1114	0.0000	0.0037	0.0000	0.1480	0.0207	0.0752

PAX5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0109	0.0048	0.0026	0.0000	0.0328	0.5490	0.1451
PAX8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2204	0.2207	0.0014	0.0833	0.0000	1.4219	0.0000	0.2317
PDPN	0.1577	0.1071	0.1112	0.0014	0.0000	0.2774	0.0000	0.0653	0.0172	0.0021	0.0496	0.1240	0.0099	0.1429
PDX1	0.0000	0.0000	0.0049	0.0000	0.0000	0.0019	0.0079	0.0181	0.0000	0.0044	0.0420	0.0515	0.0000	0.0471
PECAM1	0.0030	0.0000	0.0013	0.0000	0.0000	0.0000	0.0140	0.0596	0.0000	0.0000	0.0032	0.1528	0.0616	0.0700
PGR	0.0143	0.0038	0.0021	0.2152	0.0000	0.0000	0.0277	0.0757	0.0000	0.0000	0.0085	0.1129	0.0000	0.1692
PIP	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.0011	0.2079	0.0000	0.0069	0.0000	0.1061	0.1434	0.0904
PMEL	0.0000	0.0000	0.8212	0.0000	0.0000	0.0000	0.0000	0.0754	0.0000	0.0512	0.0081	0.1625	0.0066	0.1642
PMS2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0362	0.0717	0.0000	0.0000	0.1479	0.0439	0.0069	0.2477
POU5F1	0.0000	0.0000	0.1686	0.0000	0.0000	0.0000	0.0668	0.0951	0.0000	0.0524	0.2000	0.0356	0.0037	0.0889
PSAP	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0954	0.0000	0.0000	0.0064	0.0877	0.0087	0.1666
PTPRC	0.0312	0.0007	0.0192	0.0000	0.0000	0.0053	0.0471	0.2771	0.0000	0.0000	0.0101	0.0394	0.0298	0.0298
S100A10	0.0360	0.0054	0.0027	0.0524	0.0000	0.0000	0.1669	0.0953	0.0000	0.0000	0.0263	0.0565	0.5088	0.0466
S100A11	0.0048	0.0000	0.0021	0.0000	0.0000	0.0015	0.4565	0.0661	0.4309	0.0000	0.2571	0.0551	0.3458	0.0141
S100A12	0.0000	0.0063	0.0000	0.0000	0.0470	0.0000	0.0000	0.1326	0.0007	0.0000	0.1065	0.0747	0.1572	0.0311
S100A13	0.0000	0.0000	0.3703	0.0000	0.0000	0.0000	0.0000	0.0789	0.0031	0.0054	0.0000	0.2269	0.0530	0.0504
S100A14	0.1648	0.0037	0.4983	0.3337	0.0468	0.0000	0.0065	0.0342	0.1434	0.4994	0.4276	0.2245	0.0048	0.1856
S100A16	0.0096	0.0000	0.0000	0.0000	0.0000	0.0052	0.0319	0.0602	0.0000	0.0000	0.0404	0.3255	0.0000	0.0306
S100A1	0.0197	0.0000	0.0740	0.0000	0.0000	0.0000	0.3546	0.3587	0.0009	0.0408	0.0114	0.0937	0.0130	0.4877
S100A2	0.0007	0.0000	0.0049	0.1196	0.0000	0.0000	0.0000	0.1330	0.0088	0.0000	0.0274	0.0863	0.0095	0.1500
S100A4	0.0061	0.0000	0.0194	0.0416	0.0000	0.0000	0.1067	0.1375	0.2105	0.0000	0.0883	0.0472	0.0224	0.0687
S100A5	0.2135	0.0000	0.0000	0.0003	0.0000	0.0000	0.0095	0.1069	0.0000	0.0071	0.1755	0.3122	0.0849	0.0309
S100A6	0.0000	0.0000	0.0028	0.0176	0.0000	0.0000	0.0211	0.0941	0.0000	0.0000	0.0000	0.0275	0.2425	0.2987
S100A7A	0.0030	0.0000	0.0000	0.0000	0.0000	0.0019	0.0000	0.1654	0.0000	0.0021	0.0262	0.0538	0.0094	0.0455
S100A7L2	0.0088	0.0000	0.0000	0.0000	0.0000	0.0000	0.0110	0.0095	0.0000	0.0000	0.0000	0.0351	0.0000	0.1266
S100A7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1054	0.0370	0.0034	0.1035	0.0451	0.0240	0.0201	0.0404
S100A8	0.0000	0.0000	0.0100	0.0227	0.0000	0.0000	0.0022	0.0855	0.0000	0.0000	0.0158	0.0895	0.0423	0.1287
S100A9	0.0212	0.0059	0.0029	0.0231	0.0000	0.0000	0.0141	0.0342	0.0000	0.0000	0.0260	0.1034	0.0029	0.0356
S100B	0.0497	0.0074	1.2133	0.0000	0.0000	0.0000	0.0134	0.1238	0.0000	0.0251	0.0010	0.0817	0.0020	0.0271
S100PBP	0.0004	0.0000	0.0041	0.0000	0.0314	0.0000	0.0264	0.0240	0.1020	0.0509	0.0058	0.0677	0.0165	0.0468
S100P	0.1138	0.0000	0.0135	0.0000	0.0000	0.0000	0.0088	0.1531	0.0000	0.1384	0.0000	0.2549	0.0792	0.0417
S100Z	0.0044	0.0000	0.0000	0.0000	0.0000	0.0000	0.2346	0.2556	0.0000	0.0293	0.0546	0.0849	0.0647	0.0274
SALL4	0.0507	0.0000	0.0072	0.0184	0.0478	0.0000	0.0000	0.0931	0.0625	0.0000	0.0000	0.1662	0.0420	0.0445
SATB2	0.2218	0.0002	0.1597	0.0000	0.0000	0.0119	0.0651	0.0424	0.0000	0.2507	0.2480	0.4029	0.0038	0.1155
SDC1	0.0622	0.0060	0.0000	0.5929	0.0000	0.0000	0.1322	0.1158	0.1000	0.0191	0.0238	0.3000	0.0297	0.3134
SERPINA1	0.0000	0.0006	0.0000	0.0002	0.0000	0.0000	0.0081	0.1930	0.0000	0.0000	0.0000	0.2772	0.0000	0.1166
SERPINB5	0.0000	0.0000	0.0000	0.0019	0.0000	0.0000	0.0174	0.0932	0.0000	0.1004	0.0000	0.1800	0.0829	0.3867
SF1	0.0047	0.0000	0.0062	0.0014	0.0000	0.0023	0.0000	0.1650	0.0000	0.0000	0.0125	0.1431	0.0000	0.0197
SFTPA1	0.0000	0.0000	0.0076	0.0000	0.0000	0.0000	0.0270	0.3428	0.0008	0.0000	0.2125	0.1150	0.0059	0.2155
SMAD4	0.0272	0.0000	0.0000	0.0000	0.0150	0.0000	0.0116	0.2866	0.0000	0.0000	0.0496	0.1447	0.0127	0.0617
SMARCB1	0.0000	0.0000	0.0000	0.0701	0.0000	0.2646	0.0000	0.0166	0.0000	0.0000	0.0000	0.0312	0.0049	0.0798
SMN1	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000	0.0250	0.0541	0.0003	0.0000	0.0157	0.0584	0.2638	0.0639
SOX2	0.0607	0.0042	0.0777	0.0000	0.0000	0.0000	0.0509	0.3111	0.0095	0.0209	0.0380	0.2204	0.0025	0.7663
SPN	0.0000	0.0006	0.0000	0.0227	0.0000	0.0000	0.0087	0.0644	0.0000	0.0000	0.0061	0.0449	0.0101	0.0201

SYP	0.0414	0.0013	0.0020	0.0000	0.0014	0.0000	0.3135	0.0395	0.3229	0.0545	0.0297	0.0218	0.2181	0.0676
TFE3	0.0015	0.0000	0.0049	0.0075	0.0000	0.0000	0.0065	0.0676	0.0000	0.0609	0.0029	0.0983	0.0146	0.1474
TFF1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0096	0.1063	0.0276	0.0209	0.0071	0.1115	0.0952	0.1028
TFF3	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.2867	0.2256	0.0000	0.0066	0.0000	0.2560	0.1633	0.0155
TG	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1004	0.0071	0.0119	0.0023	0.2005	0.0956	0.1166
TLE1	0.0052	0.0000	0.0000	0.0030	0.0000	0.0168	0.0000	0.0810	0.0000	0.0000	0.0122	0.1071	0.0034	0.0873
TMPRSS2	0.0147	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4196	0.0000	0.1294	0.0000	0.0587	0.0000	0.2092
TNFRSF8	0.0000	0.0000	0.0046	0.0074	0.0000	0.0002	0.0000	0.0272	0.0000	0.0070	0.0186	0.0668	0.0006	0.0338
TP63	0.0087	0.0000	0.1029	0.0828	0.0000	0.0000	0.1021	0.2985	0.0000	0.0084	0.0688	0.0563	0.0073	2.1955
TPM1	0.2399	0.0034	0.2265	0.0024	0.0000	0.0000	0.0000	0.0414	0.0000	0.0578	0.0000	0.1404	0.0000	0.0940
TPM2	0.2544	0.0000	0.0000	0.0280	0.0000	0.0000	0.0355	0.1050	0.0386	0.0359	0.0000	0.0472	0.0000	0.0962
TPM3	0.0006	0.0000	0.0091	0.0103	0.0000	0.0000	0.0094	0.1137	0.0000	0.0083	0.0768	0.0791	0.0185	0.1827
TPM4	0.3360	0.0658	0.0000	0.0000	0.0000	0.0000	0.0246	0.1235	0.0004	0.0074	0.0028	0.1710	0.0015	0.1585
TPSAB1	0.0000	0.0000	0.0039	0.0000	0.0000	0.0000	0.0054	0.0588	0.0000	0.0016	0.0000	0.0877	0.1779	0.2889
TTF1	0.0000	0.0000	0.0267	0.0093	0.0000	0.0000	0.0027	0.0819	0.0342	0.0000	0.0515	0.0738	0.0969	0.2675
UPK2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.0354	0.0579	0.0000	0.0058	0.0145	0.0888	0.0697
UPK3A	0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0772	0.0381	0.0008	0.0000	0.0000	0.0576	0.0211	0.0987
UPK3B	0.0014	0.0018	0.0055	0.0000	0.0000	0.5617	0.0000	0.0308	0.0000	0.0000	0.0022	0.0295	0.0004	0.1637
VHL	0.0000	0.0008	0.0000	0.0000	0.0000	0.0000	0.0599	0.1707	0.0000	0.0000	0.0686	0.0794	0.0631	0.0949
VIL1	0.0021	0.0000	0.0832	0.0000	0.0000	0.0000	0.0138	0.0637	0.0000	0.0055	0.0115	0.1072	0.0339	0.0583
VIM	0.0000	0.0000	0.1933	0.2832	0.0000	0.0000	0.0000	0.1175	0.0301	0.0000	0.4466	0.0938	0.0036	0.0684
WT1	0.0063	0.0017	0.0011	0.0099	0.0000	0.0771	0.0034	0.0333	0.0000	0.1347	0.0000	2.1030	0.0205	0.0966

As noted, the transcripts provided in **Tables 117-120** can be used in the systems and processes outlined in **FIGs. 4A-B**. For example, the disclosure provides a method for classifying a biological sample **400**, **410**, the method comprising: obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample **401**, **411**; obtaining, as desired, by one or more computers, second data representing another initial classification for the biological sample that were previously determined based on DNA sequences of the biological sample **416** (see, e.g., **Tables 2-16** and related text); providing, by one or more computers, at least a portion of the first data and the second data as an input to a dynamic voting engine **406**, **415** that has been trained to predict a target biological sample classification based on processing of multiple initial biological sample classifications; processing, by one or more computers, the provided input data through the dynamic voting engine; obtaining, by one or more computers, output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the provided input data; and determining, by one or more computers, a target biological sample classification for the biological sample based on the obtained output data **407**, **417**. In some embodiments, obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample comprises: obtaining data representing a cancer type classification for the biological sample based the RNA sequences of the biological

sample **403**, **412** (see, e.g., **Table 118** and related text); obtaining data representing an organ from which the biological sample originated based on the RNA sequences of the biological sample **404**, **413** (see, e.g., **Table 119** and related text); and obtaining data representing a histology for the biological sample based on the RNA sequences of the biological sample **405**, **414** (see, e.g., **Table 120** and related text), and wherein providing at least a portion of the first data and the second data as an input to the dynamic voting engine **406**, **415** comprises: providing the obtained data representing the cancer type **403**, **412**, the obtained data representing the organ from which the biological sample originated **404**, **413**, the obtained data representing the histology **405**, **414**, and the second data as an input to the dynamic voting engine **406**, **415**. In some embodiments, the dynamic voting engine **406**, **415** comprises one or more machine learning model. In some embodiments, previously determining an initial classification for the biological sample based on DNA sequences of the biological sample comprises **416**: receiving, by one or more computers, a biological signature representing the biological sample that was obtained from a cancerous neoplasm in a first portion of a body, wherein the model includes a cancerous biological signature for each of multiple different types of cancerous biological samples, wherein each of the cancerous biological signatures include at least a first cancerous biological signature representing a molecular profile of a cancerous biological sample from the first portion of one or more other bodies and a second cancerous biological signature representing a molecular profile of a cancerous biological sample from a second portion of one or more other bodies; performing, by one or more computers and using a pairwise-analysis model, pairwise analysis of the biological signature using the first cancerous biological signature and the second cancerous biological signature; generating, by one or more computers and based on the performed pairwise analysis, a likelihood that the cancerous neoplasm in the first portion of the body was caused by cancer in a second portion of the body; and storing, by one or more computers, the generated likelihood in a memory device.

Relatedly, the disclosure also a method comprising: (a) obtaining a biological sample from a subject having a cancer; (b) performing at least one assay on the sample to assess one or more biomarkers, thereby obtaining a biosignature for the sample; (c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises at least one pre-determined biosignature indicative of at least one attribute, and wherein the at least one attribute of the cancer is selected from the group comprising primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof; (d) processing, by one or more computers, the provided biosignature through the model; and (e) outputting from the model a prediction of the at least one attribute of the cancer. The assays may comprise next generation sequencing of DNA and RNA, e.g., as described in **Example 1**. The assays can be performed to measure the same inputs as those used to train the models, e.g., based on **Tables 2-116** and/or **Tables 118-120**. Therefore the data for the sample from the subject can be processed to determine the attribute. For example, the models may be trained using data for DNA analysis of groups of genes

selected from **Tables 123-125** and/or **Tables 128-129**, or selections thereof. For example, the models may also be trained using data for RNA analysis of groups of genes selected from **Table 117**, or selections thereof. The biomarkers within the models thereby provide predetermined biosignatures. Then the assays performed on the samples for the subject can query those same biomarkers within the predetermined biosignatures. As a non-limiting example, predetermined biosignatures trained to predict a cancer or disease type may be according to **Table 118**, predetermined biosignatures trained to predict an organ type may be according to **Table 119**, and/or predetermined biosignatures trained to predict a histology may be according to **Table 120**. Following this example, a sample from a subject would then be assayed in order to determine a biosignature comprising the genes in **Table 118**, **Table 119**, and or **Table 120**. Accordingly, the sample biosignature can be processed by the models comprising the corresponding predetermined biosignatures.

As a further illustration of the method of predicting the at least one attribute of a cancer, the disclosure provides a method such as outlined in **FIGs. 4A-B 400**, **410** comprising: (a) obtaining a biological sample from a subject having a cancer, wherein the biological sample comprises a tumor sample, bodily fluid, or other obtainable sample such as described herein; (b) performing at least one assay to assess one or more biomarkers in the biological sample to obtain a biosignature for the sample, e.g., performing DNA analysis by sequencing genomic DNA from the biological sample **416**, wherein the DNA analysis can be performed for selections of the genes in **Tables 2-116**; and/or performing RNA analysis by sequencing messenger RNA transcripts from the biological sample **410**, **411**, wherein the RNA analysis is performed for selections of the genes in **Table 117** or **Tables 118-120**; (c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises a plurality of intermediate models, wherein the plurality of intermediate models comprises: (1) an first intermediate model trained to process DNA data using the predetermined biosignatures according to **Tables 2-116** (**416**); (2) a second intermediate model trained to process RNA data using predetermined biosignatures according to **Table 118** (**403**, **412**); (3) a third intermediate model trained to process RNA data using predetermined biosignatures according to **Table 119** (**403**, **412**); and (4) a fourth intermediate model trained to process RNA data using the predetermined biosignatures according to **Table 120** (**404**, **413**); (d) processing, by one or more computers, the provided biosignature through each of the plurality of intermediate models in part (c), providing the output of each of the plurality of intermediate models into a final predictor model, e.g. dynamic voting module **415**, and processing by one or more computers, the output of each of the plurality of intermediate models through the final predictor model; and (e) outputting from the final predictor model a prediction of the at least one attribute of the cancer **417**. As described herein, the attribute is related to a tissue characteristic, such as TOO, and can be output at a desired level of granularity. In some embodiments, the predicted at least one attribute of the cancer is a tissue-of-origin selected from the group consisting of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal

adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, uterine sarcoma, and a combination thereof. As desired, the models can be trained to output the TOO at different levels of granularity as described herein. See, e.g., the disease types and organ groups denoted in **Tables 2-116** and related discussion.

The predicted at least one attribute of the cancer may be compared to a threshold. For example, the prediction or classification provided by the systems and methods herein may comprise a probability, likelihood, or similar statistical measure that indicates a confidence level in the predicted attribute. Such confidence level may be determined for each potential attribute. See, e.g., discussion in **Example 3** and in the exemplar reports in **Examples 4-5**. The confidence in the prediction may be particularly important when assisting in treatment decision making for cancer patients. As desired, the disclosure contemplates additional clinical testing or review to confirm or not the predicted attribute.

The disclosure further provides a system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described in the paragraphs above. The disclosure also provides a non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described in the paragraphs above.

Advantageously, the systems and methods provided herein can be performed using the molecular profiling data that is used to help guide treatment selection for cancer patients. See, e.g., **Example 1**. The predicted attributes may help provide a diagnosis of a CUP sample, or provide a quality check and potentially adjusted diagnosis for any profiled sample. The latter may be particularly desirable to verify the origin of a metastatic sample, or other remote sample such as a blood sample or other bodily fluid. Thus, the systems and methods provided herein provide an efficient means to help improve treatment of cancer patients.

Example 3 provides further details and demonstration of RNA and panomic classifiers **400** and **410**.

Report

In an embodiment, the methods as described herein comprise generating a molecular profile report. The report can be delivered to the treating physician or other caregiver of the subject whose cancer has been profiled. The report can comprise multiple sections of relevant information, including without limitation: 1) a list of the biomarkers that were profiled (i.e., subject to molecular testing); 2) a description of the molecular profile comprising characteristics of the genes and/or gene products as determined for the subject; 3) a treatment associated with the characteristics of the genes and/or gene

products that were profiled; and 4) and an indication whether each treatment is likely to benefit the patient, not benefit the patient, or has indeterminate benefit. The list of the genes in the molecular profile can be those presented herein. See, e.g., **Example 1**. The description of the biomarkers assessed may include such information as the laboratory technique used to assess each biomarker (e.g., RT-PCR, FISH/CISH, PCR, FA/RFLP, NGS, etc) as well as the result and criteria used to score each technique. By way of example, the criteria for scoring a CNV may be a presence (i.e., a copy number that is greater or lower than the "normal" copy number present in a subject who does not have cancer, or statistically identified as present in the general population, typically diploid) or absence (i.e., a copy number that is the same as the "normal" copy number present in a subject who does not have cancer, or statistically identified as present in the general population, typically diploid) The treatment associated with one or more of the genes and/or gene products in the molecular profile can be determined using a biomarker-treatment association rule set such as in **Tables 2-116**, **Tables 117-120**, ISNM1, or **Tables 121-130** herein or any of International Patent Publications WO/2007/137187 (Int'l Appl. No. PCT/US2007/069286), published November 29, 2007; WO/2010/045318 (Int'l Appl. No. PCT/US2009/060630), published April 22, 2010; WO/2010/093465 (Int'l Appl. No. PCT/US2010/000407), published August 19, 2010; WO/2012/170715 (Int'l Appl. No. PCT/US2012/041393), published December 13, 2012; WO/2014/089241 (Int'l Appl. No. PCT/US2013/073184), published June 12, 2014; WO/2011/056688 (Int'l Appl. No. PCT/US2010/054366), published May 12, 2011; WO/2012/092336 (Int'l Appl. No. PCT/US2011/067527), published July 5, 2012; WO/2015/116868 (Int'l Appl. No. PCT/US2015/013618), published August 6, 2015; WO/2017/053915 (Int'l Appl. No. PCT/US2016/053614), published March 30, 2017; WO/2016/141169 (Int'l Appl. No. PCT/US2016/020657), published September 9, 2016; and WO2018175501 (Int'l Appl. No. PCT/US2018/023438), published September 27, 2018; each of which publications is incorporated by reference herein in its entirety. Such biomarker-treatment associations can be updated over time, e.g., as associations are refuted or as new associations are discovered. The indication whether each treatment is likely to benefit the patient, not benefit the patient, or has indeterminate benefit may be weighted. For example, a potential benefit may be a strong potential benefit or a lesser potential benefit. Such weighting can be based on any appropriate criteria, e.g., the strength of the evidence of the biomarker-treatment association, or the results of the profiling, e.g., a degree of over- or underexpression.

Various additional components can be added to the report as desired. In preferred embodiments, the report comprises a section detailing results of tissue classification, e.g., as described for determining one or more of a primary tumor local, cancer category, cancer/disease type, organ type, and/or histology. See, e.g., **FIGs. 7E**, **8C**. Such attribute can be provided at a desired level of granularity, e.g., at a level that may alter treatment if the predicted attribute differs from the original attribution. See, e.g., **FIGs. 6AH-AL** and related discussion.

In some embodiments, the report comprises a list having an indication of whether a presence, level or state of an assessed biomarker is associated with an ongoing clinical trial. The report may include identifiers for any such trials, e.g., to facilitate the treating physician's investigation of potential enrollment of the subject in the trial. In some embodiments, the report provides a list of evidence supporting the association of the assessed biomarker with the reported treatment. The list can contain citations to the evidentiary literature and/or an indication of the strength of the evidence for the particular biomarker-treatment association. In some embodiments, the report comprises a description of the genes and gene products that were profiled. The description of the genes in the molecular profile can comprise without limitation the biological function and/or various treatment associations.

The molecular profiling report can be delivered to the caregiver for the subject, e.g., the oncologist or other treating physician. The caregiver can use the results of the report to guide a treatment regimen for the subject. For example, the caregiver may use one or more treatments indicated as likely benefit in the report to treat the patient. Similarly, the caregiver may avoid treating the patient with one or more treatments indicated as likely lack of benefit in the report.

In some embodiments of the method of identifying at least one therapy of potential benefit, the subject has not previously been treated with the at least one therapy of potential benefit. The cancer may comprise a metastatic cancer, a recurrent cancer, or any combination thereof. In some cases, the cancer is refractory to a prior therapy, including without limitation front-line or standard of care therapy for the cancer. In some embodiments, the cancer is refractory to all known standard of care therapies. In other embodiments, the subject has not previously been treated for the cancer. The method may further comprise administering the at least one therapy of potential benefit to the individual. Progression free survival (PFS), disease free survival (DFS), or lifespan can be extended by the administration.

Exemplary reports are provided herein in **FIGs. 7** and **8**, which are detailed in **Examples 4** and **5**, respectively.

The report can be computer generated, and can be a printed report, a computer file or both. The report can be made accessible via a secure web portal.

In an aspect, the disclosure provides use of a reagent in carrying out the methods as described herein as described above. In a related aspect, the disclosure provides of a reagent in the manufacture of a reagent or kit for carrying out the methods as described herein as described herein. In still another related aspect, the disclosure provides a kit comprising a reagent for carrying out the methods as described herein as described herein. The reagent can be any useful and desired reagent. In preferred embodiments, the reagent comprises at least one of a reagent for extracting nucleic acid from a sample, and a reagent for performing next-generation sequencing.

The disclosure also provides systems for performing molecular profiling and generating a report comprising results and analysis thereof. In an aspect, the disclosure provides a system for

identifying at least one therapy associated with a cancer in an individual, comprising: (a) at least one host server; (b) at least one user interface for accessing the at least one host server to access and input data; (c) at least one processor for processing the inputted data; (d) at least one memory coupled to the processor for storing the processed data and instructions for: i) accessing a molecular profile, e.g., according to **Example 1**; and ii) identifying, based on the status of various biomarkers within the molecular profile, at least one therapy with potential benefit for treatment of the cancer; and (e) at least one display for displaying the identified therapy with potential benefit for treatment of the cancer. In some embodiments, the system further comprises at least one memory coupled to the processor for storing the processed data and instructions for identifying, based on the generated molecular profile according to the methods above, at least one therapy with potential benefit for treatment of the cancer; and at least one display for display thereof. The system may further comprise at least one database comprising references for various biomarker states, data for drug/biomarker associations, or both. The at least one display can be a report provided by the present disclosure.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope as described herein described in the claims.

Example 1: Molecular Profiling

Comprehensive molecular profiling provides a wealth of data concerning the molecular status of patient samples. We have performed such profiling on well over 100,000 tumor patients from practically all cancer lineages using various profiling technologies. To date, we have tracked the benefit or lack of benefit from treatments in over 20,000 of these patients. Our molecular profiling data can thus be compared to patient benefit to treatments to identify additional biomarker signatures that predict the benefit to various treatments in additional cancer patients. We have applied this "next generation profiling" (NGP) approach to identify biomarker signatures that correlate with patient benefit (including positive, negative, or indeterminate benefit) to various cancer therapeutics.

The general approach to NGP is as follows. Over several years we have performed comprehensive molecular profiling of tens of thousands of patients using various molecular profiling techniques. As further outlined in **FIG. 2C**, these techniques include without limitation next generation sequencing (NGS) of DNA to assess various attributes **2301**, gene expression and gene fusion analysis of RNA **2302**, IHC analysis of protein expression **2303**, and ISH to assess gene copy number and chromosomal aberrations such as translocations **2304**. We currently have matched patient clinical outcomes data for over 20,000 patients of various cancer lineages **2305**. We use cognitive computing approaches **2306** to correlate the comprehensive molecular profiling results against the actual patient outcomes data for various treatments as desired. Clinical outcome may be determined

using the surrogate endpoint time-on-treatment (TOT) or time-to-next-treatment (TTNT or TNT). See, e.g., Roever L (2016) Endpoints in Clinical Trials: Advantages and Limitations. Evidence Based Medicine and Practice 1: e111. doi:10.4172/ebmp.1000e111. The results provide a biosignature comprising a panel of biomarkers **2307**, wherein the biosignature is indicative of benefit or lack of benefit from the treatment under investigation. The biosignature can be applied to molecular profiling results for new patients in order to predict benefit from the applicable treatment and thus guide treatment decisions. Such personalized guidance can improve the selection of efficacious treatments and also avoid treatments with lesser clinical benefit, if any.

Table 121 lists numerous biomarkers we have profiled over the past several years. As relevant molecular profiling and patient outcomes are available, any or all of these biomarkers can serve as features to input into the cognitive computing environment to develop a biosignature of interest. The table shows molecular profiling techniques and various biomarkers assessed using those techniques. The listing is non-exhaustive, and data for all of the listed biomarkers will not be available for every patient. It will further be appreciated that various biomarker have been profiled using multiple methods. As a non-limiting example, consider the EGFR gene expressing the Epidermal Growth Factor Receptor (EGFR) protein. As shown in **Table 121**, expression of EGFR protein has been detected using IHC; EGFR gene amplification, gene rearrangements, mutations and alterations have been detected with ISH, Sanger sequencing, NGS, fragment analysis, and PCR such as qPCR; and EGFR RNA expression has been detected using PCR techniques, e.g., qPCR, and DNA microarray. As a further non-limiting example, molecular profiling results for the presence of the EGFR variant III (EGFRvIII) transcript has been collected using fragment analysis (e.g., RFLP) and sequencing (e.g., NGS).

Table 122 shows exemplary molecular profiles for various tumor lineages. Data from these molecular profiles may be used as the input for NGP in order to identify one or more biosignatures of interest. In the table, the cancer lineage is shown in the column "Tumor Type." The remaining columns show various biomarkers that can be assessed using the indicated methodology (i.e., immunohistochemistry (IHC), in situ hybridization (ISH), or other techniques). As explained above, the biomarkers are identified using symbols known to those of skill in the art. Under the IHC column, "MMR" refers to the mismatch repair proteins MLH1, MSH2, MSH6, and PMS2, which are each individually assessed using IHC. Under the WES column "DNA Alterations," "CNA" refers to copy number alteration, which is also referred to herein as copy number variation (CNV). Under the WES column "Genomic Signatures," "MSI" refers to microsatellite instability; "TMB" refers to tumor mutational burden, which may be referred to as tumor mutational load or TML; "LOH" refers to loss of heterozygosity; and "FOLFOX" refers to a predictor of FOLFOX response in metastatic colorectal adenocarcinoma as described in Int'l Patent Publication WO2020113237, titled "NEXT-GENERATION MOLECULAR PROFILING" and based on Int'l Patent Application No. PCT/US2019/064078, filed December 2, 2019, which publication is hereby incorporated by reference

in its entirety. Whole transcriptome sequencing (WTS) is used to assess all RNA transcripts in the specimen and can detect, inter alia, fusions and variant transcripts. Under the column "Other," abbreviations include EBER for Epstein-Barr encoding region; and HPV is human papilloma virus. One of skill will appreciate that molecular profiling technologies may be substituted as desired and/or interchangeable. For example, other suitable protein analysis methods can be used instead of IHC (e.g., alternate immunoassay formats), other suitable nucleic acid analysis methods can be used instead of ISH (e.g., that assess copy number and/or rearrangements, translocations and the like), and other suitable nucleic acid analysis methods can be used instead of fragment analysis. Similarly, FISH and CISH are generally interchangeable and the choice may be made based upon probe availability and the like. **Tables 123-125** and **128-129** present panels of genomic analysis and genes that have been assessed using Next Generation Sequencing (NGS) analysis of DNA such as genomic DNA. Whole exome sequencing (WES) can be used to analyze the genomic DNA. One of skill will appreciate that other nucleic acid analysis methods can be used instead of NGS analysis, e.g., other sequencing (e.g., Sanger), hybridization (e.g., microarray, Nanostring) and/or amplification (e.g., PCR based) methods. The biomarkers listed in **Tables 126-127** can be assessed by RNA sequencing, such as WTS. Using WTS, any fusions, splice variants, or the like can be detected. **Tables 126-127** list biomarkers with commonly detected alterations in cancer.

Nucleic acid analysis may be performed to assess various aspects of a gene. For example, nucleic acid analysis can include, but is not limited to, mutational analysis, fusion analysis, variant analysis, splice variants, SNP analysis and gene copy number/amplification. Such analysis can be performed using any number of techniques described herein or known in the art, including without limitation sequencing (e.g., Sanger, Next Generation, pyrosequencing), PCR, variants of PCR such as RT-PCR, fragment analysis, and the like. NGS techniques may be used to detect mutations, fusions, variants and copy number of multiple genes in a single assay. Unless otherwise stated or obvious in context, a "mutation" as used herein may comprise any change in a gene or genome as compared to wild type, including without limitation a mutation, polymorphism, deletion, insertion, indels (i.e., insertions or deletions), substitution, translocation, fusion, break, duplication, loss, amplification, repeat, or copy number variation. Different analyses may be available for different genomic alterations and/or sets of genes. For example, **Table 123** lists attributes of genomic stability that can be measured with NGS, **Table 124** lists various genes that may be assessed for point mutations and indels, **Table 125** lists various genes that may be assessed for point mutations, indels and copy number variations, **Table 126** lists various genes that may be assessed for gene fusions via RNA analysis, e.g., via WTS, and similarly **Table 127** lists genes that can be assessed for transcript variants via RNA. Molecular profiling results for additional genes can be used to identify an NGP biosignature as such data is available.

Table 121 – Molecular Profiling Biomarkers

Technique	Biomarkers
IHC	ABL1, ACPP (PAP), Actin (ACTA), ADA, AFP, AKT1, ALK, ALPP (PLAP-1), APC, AR, ASNS, ATM, BAP1, BCL2, BCRP, BRAF, BRCA1, BRCA2, CA19-9, CALCA, CCND1 (BCL1), CCR7, CD19, CD276, CD3, CD33, CD52, CD80, CD86, CD8A, CDH1 (ECAD), CDW52, CEACAM5 (CEA; CD66e), CES2, CHGA (CGA), CK 14, CK 17, CK 5/6, CK1, CK10, CK14, CK15, CK16, CK19, CK2, CK3, CK4, CK5, CK6, CK7, CK8, COX2, CSF1R, CTL4A, CTLA4, CTNNB1, Cytokeratin, DCK, DES, DNMT1, EGFR, EGFR H-score, ERBB2 (HER2), ERBB4 (HER4), ERCC1, ERCC3, ESR1 (ER), F8 (FACTOR8), FBXW7, FGFR1, FGFR2, FLT3, FOLR2, GART, GNA11, GNAQ, GNAS, Granzyme A, Granzyme B, GSTP1, HDAC1, HIF1A, HNF1A, HPL, HRAS, HSP90AA1 (HSPCA), IDH1, IDO1, IL2, IL2RA (CD25), JAK2, JAK3, KDR (VEGFR2), KI67, KIT (cKIT), KLK3 (PSA), KRAS, KRT20 (CK20), KRT7 (CK7), KRT8 (CYK8), LAG-3, MAGE-A, MAP KINASE PROTEIN (MAPK1/3), MDM2, MET (cMET), MGMT, MLH1, MPL, MRP1, MS4A1 (CD20), MSH2, MSH4, MSH6, MSI, MTAP, MUC1, MUC16, NFKB1, NFKB1A, NFKB2, NGF, NOTCH1, NPM1, NRAS, NY-ESO-1, ODC1 (ODC), OGFR, p16, p95, PARP-1, PBRM1, PD-1, PDGF, PDGFC, PDGFR, PDGFRA, PDGFRA (PDGFR2), PDGFRB (PDGFR1), PD-L1, PD-L2, PGR (PR), PIK3CA, PIP, PMEL, PMS2, POLA1 (POLA), PR, PTEN, PTGS2 (COX2), PTPN11, RAF1, RARA (RAR), RB1, RET, RHOH, ROS1, RRM1, RXR, RXRB, S100B, SETD2, SMAD4, SMARCB1, SMO, SPARC, SST, SSTR1, STK11, SYP, TAG-72, TIM-3, TK1, TLE3, TNF, TOP1 (TOPO1), TOP2A (TOP2), TOP2B (TOPO2B), TP, TP53 (p53), TRKA/B/C, TS, TUBB3, TXNRD1, TYMP (PDECGF), TYMS (TS), VDR, VEGFA (VEGF), VHL, XDH, ZAP70
ISH (CISH/FISH)	1p19q, ALK, EML4-ALK, EGFR, ERCC1, HER2, HPV (human papilloma virus), MDM2, MET, MYC, PIK3CA, ROS1, TOP2A, chromosome 17, chromosome 12
Pyrosequencing	MGMT promoter methylation
Sanger sequencing	BRAF, EGFR, GNA11, GNAQ, HRAS, IDH2, KIT, KRAS, NRAS, PIK3CA
NGS	See genes and types of testing in **Tables 122-129**, MSI, TMB, LOH

	WES, WTS
Fragment Analysis	ALK, EML4-ALK, EGFR Variant III, HER2 exon 20, ROS1, MSI
PCR	ALK, AREG, BRAF, BRCA1, EGFR, EML4, ERBB3, ERCC1, EREG, hENT-1, HSP90AA1, IGF-1R, KRAS, MMR, p16, p21, p27, PARP-1, PGP (MDR-1), PIK3CA, RRM1, TLE3, TOPO1, TOPO2A, TS, TUBB3
Microarray	ABCC1, ABCG2, ADA, AR, ASNS, BCL2, BIRC5, BRCA1, BRCA2, CD33, CD52, CDA, CES2, DCK, DHFR, DNMT1, DNMT3A, DNMT3B, ECGF1, EGFR, EPHA2, ERBB2, ERCC1, ERCC3, ESR1, FLT1, FOLR2, FYN, GART, GNRH1, GSTP1, HCK, HDAC1, HIF1A, HSP90AA1 (HSPCA), IL2RA, HSP90AA1, KDR, KIT, LCK, LYN, MGMT, MLH1, MS4A1, MSH2, NFKB1, NFKB2, OGFR, PDGFC, PDGFRA, PDGFRB, PGR, POLA1, PTEN, PTGS2, RAF1, RARA, RRM1, RRM2, RRM2B, RXRB, RXRG, SPARC, SRC, SSTR1, SSTR2, SSTR3, SSTR4, SSTR5, TK1, TNF, TOP1, TOP2A, TOP2B, TXNRD1, TYMS, VDR, VEGFA, VHL, YES1, ZAP70

Table 122 – Molecular Profiles

Tumor Type	IHC	Whole Exome Sequencing (WES)		Whole Transcriptome Sequencing (WTS)	Other
		DNA alterations	Genomic Signatures	RNA	
Bladder	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Breast	AR, ER, Her2/Neu, MMR, PD-L1, PR, PTEN	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	Her2, TOP2A (CISH)
Cancer of Unknown Primary - Female	AR, ER, HER2, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Cancer of Unknown Primary - Male	AR, HER2, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Cervical	ER, MMR, PD-L1, PR	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Cholangiocarcinoma/ Hepatobiliary	Her2/Neu, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	Her2 (CISH)
Colorectal and Small Intestinal	Her2/Neu, MMR, PD-L1, PTEN	Mutation, Indels, CNA	MSI, TMB,	Fusions, Variant Transcripts	

			LOH, FOLFOX		
Endometrial	ER, MMR, PD-L1, PR, PTEN	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Esophageal	Her2/Neu, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	EBER (CISH)
Gastric/GEJ	Her2/Neu, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	EBER, Her2 (CISH)
GIST	MMR, PD-L1, PTEN	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Glioma	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	MGMT Methylation (Pyrosequencing)
Head & Neck	MMR, p16, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	EBER, HPV (CISH), reflex to confirm p16 result
Kidney	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Lymphoma / Leukemia		Mutation, Indels, CNA	TMB	Fusions, Variant Transcripts	
Melanoma	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Merkel Cell	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Neuroendocrine	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Non-Small Cell Lung	ALK, MMR, PD-L1, PTEN	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Ovarian	ER, MMR, PD-L1, PR	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Pancreatic	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Prostate	AR, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Salivary Gland	AR, Her2/Neu, MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Sarcoma	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	

Small Cell Lung	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Thyroid	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Uterine Serous	ER, Her2/Neu, MMR, PD-L1, PR, PTEN	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	Her2 (CISH)
Vulvar Cancer (SCC)	ER, MMR, PD-L1, PR, TRK A/B/C	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	
Other Tumors	MMR, PD-L1	Mutation, Indels, CNA	MSI, TMB, LOH	Fusions, Variant Transcripts	

Table 123 – Genomic Stability Testing (DNA)

Microsatellite Instability (MSI)	Tumor Mutational Burden (TMB)	Loss of Heterozygosity (LOH)

5

Table 124 – Point Mutations and Indels (DNA)

ABI1	CDK12	FOXO3	INHBA	MSN	PHOX2B	SSX1
ABL1	CDKN2B	FOXO4	IRS2	MTCP1	PIK3CG	STAG2
ACKR3	CDKN2C	FSTL3	JUN	MUC1	PLAG1	TAL1
AKT1	CEBPA	GATA1	KAT6A (MYST3)	MUTYH	PMS1	TAL2
AMER1 (FAM123B)	CHCHD7	GATA2	KAT6B	MYCL (MYCL1)	POU5F1	TBL1XR1
AR	CNOT3	GNA11	KCNJ5	NBN	PPP2R1A	TCEA1
ARAF	COL1A1	GPC3	KDM5C	NDRG1	PRF1	TCL1A
ATP2B3	COX6C	HEY1	KDM6A	NKX2-1	PRKDC	TERT
ATRX	CRLF2	HIST1H3B	KDSR	NONO	RAD21	TFE3
BCL11B	DDB2	HIST1H4I	KLF4	NOTCH1	RECQL4	TFPT
BCL2	DDIT3	HLF	KLK2	NRAS	RHOH	THRAP3

BCL2L2	DNM2	HMGN2P46	LASP1	NUMA1	RNF213	TLX3
BCOR	DNMT3A	HNF1A	LMO1	NUTM2B	RPL10	TMPRSS2
BCORL1	EIF4A2	HOXA11	LMO2	OLIG2	SEPT5	UBR5
BRD3	ELF4	HOXA13	MAFB	OMD	SEPT6	VHL
BRD4	ELN	HOXA9	MAX	P2RY8	SFPQ	WAS
BTG1	ERCC1	HOXC11	MECOM	PAFAH1B2	SLC45A3	ZBTB16
BTK	ETV4	HOXC13	MED12	PAK3	SMARCA4	ZRSR2
C15orf65	FAM46C	HOXD11	MKL1	PATZ1	SOCS1	
CBLC	FANCF	HOXD13	MLLT11	PAX8	SOX2	
CD79B	FEV	HRAS	MN1	PDE4DIP	SPOP	
CDH1	FOXL2	IKBKE	MPL	PHF6	SRC	

Table 125 – Point Mutations, Indels and Copy Number Variations (DNA)

ABL2	CCND2	ERBB4 (HER4)	GSK3B	MLF1	PML	SRGAP3
ACSL3	CCND3	ERC1	H3F3A	MLH1	PMS2	SRSF2
ACSL6	CCNE1	ERCC2	H3F3B	MLLT1	POLE	SRSF3
ADGRA2	CD274 (PDL1)	ERCC3	HERPUD1	MLLT10	POT1	SS18
AFDN	CD74	ERCC4	HGF	MLLT3	POU2AF1	SS18L1
AFF1	CD79A	ERCC5	HIP1	MLLT6	PPARG	STAT3
AFF3	CDC73	ERG	HMGA1	MNX1	PRCC	STAT4
AFF4	CDH11	ESR1	HMGA2	MRE11	PRDM1	STAT5B
AKAP9	CDK4	ETV1	HNRNPA2B1	MSH2	PRDM16	STIL
AKT2	CDK6	ETV5	HOOK3	MSH6	PRKAR1A	STK11
AKT3	CDK8	ETV6	HSP90AA1	MSI2	PRRX1	SUFU
ALDH2	CDKN1B	EWSR1	HSP90AB1	MTOR	PSIP1	SUZ12
ALK	CDKN2A	EXT1	IDH1	MYB	PTCH1	SYK
APC	CDX2	EXT2	IDH2	MYC	PTEN	TAF15
ARFRP1	CHEK1	EZH2	IGF1R	MYCN	PTPN11	TCF12
ARHGAP26	CHEK2	EZR	IKZF1	MYD88	PTPRC	TCF3
ARHGEF12	CHIC2	FANCA	IL2	MYH11	RABEP1	TCF7L2

ARID1A	CHN1	FANCC	IL21R	MYH9	RAC1	TET1
ARID2	CIC	FANCD2	IL6ST	NACA	RAD50	TET2
ARNT	CIITA	FANCE	IL7R	NCKIPSD	RAD51	TFEB
ASPSCR1	CLP1	FANCG	IRF4	NCOA1	RAD51B	TFG
ASXL1	CLTC	FANCL	ITK	NCOA2	RAF1	TFRC
ATF1	CLTCL1	FAS	JAK1	NCOA4	RALGDS	TGFBR2
ATIC	CNBP	FBXO11	JAK2	NF1	RANBP17	TLX1
ATM	CNTRL	FBXW7	JAK3	NF2	RAP1GDS1	TNFAIP3
ATP1A1	COPB1	FCRL4	JAZF1	NFE2L2	RARA	TNFRSF14
ATR	CREB1	FGF10	KDM5A	NFIB	RB1	TNFRSF17
AURKA	CREB3L1	FGF14	KDR (VEGFR2)	NFKB2	RBM15	TOP1
AURKB	CREB3L2	FGF19	KEAP1	NFKBIA	REL	TP53
AXIN1	CREBBP	FGF23	KIAA1549	NIN	RET	TPM3
AXL	CRKL	FGF3	KIF5B	NOTCH2	RICTOR	TPM4
BAP1	CRTC1	FGF4	KIT	NPM1	RMI2	TPR
BARD1	CRTC3	FGF6	KLHL6	NSD1	RNF43	TRAF7
BCL10	CSF1R	FGFR1	KMT2A (MLL)	NSD2	ROS1	TRIM26
BCL11A	CSF3R	FGFR1OP	KMT2C (MLL3)	NSD3	RPL22	TRIM27
BCL2L11	CTCF	FGFR2	KMT2D (MLL2)	NT5C2	RPL5	TRIM33
BCL3	CTLA4	FGFR3	KNL1	NTRK1	RPN1	TRIP11
BCL6	CTNNA1	FGFR4	KRAS	NTRK2	RPTOR	TRRAP
BCL7A	CTNNB1	FH	KTN1	NTRK3	RUNX1	TSC1
BCL9	CYLD	FHIT	LCK	NUP214	RUNX1T1	TSC2
BCR	CYP2D6	FIP1L1	LCP1	NUP93	SBDS	TSHR
BIRC3	DAXX	FLCN	LGR5	NUP98	SDC4	TTL
BLM	DDR2	FLI1	LHFPL6	NUTM1	SDHAF2	U2AF1
BMPR1A	DDX10	FLT1	LIFR	PALB2	SDHB	USP6
BRAF	DDX5	FLT3	LPP	PAX3	SDHC	VEGFA
BRCA1	DDX6	FLT4	LRIG3	PAX5	SDHD	VEGFB
BRCA2	DEK	FNBP1	LRP1B	PAX7	SEPT9	VTI1A
BRIP1	DICER1	FOXA1	LYL1	PBRM1	SET	WDCP

BUB1B	DOT1L	FOXO1	MAF	PBX1	SETBP1	WIF1
CACNA1D	EBF1	FOXP1	MALT1	PCM1	SETD2	WISP3
CALR	ECT2L	FUBP1	MAML2	PCSK7	SF3B1	WRN
CAMTA1	EGFR	FUS	MAP2K1 (MEK1)	PDCD1 (PD1)	SH2B3	WT1
CANT1	ELK4	GAS7	MAP2K2 (MEK2)	PDCD1LG2 (PDL2)	SH3GL1	WWTR1
CARD11	ELL	GATA3	MAP2K4	PDGFB	SLC34A2	XPA
CARS	EML4	GID4 (C17orf39)	MAP3K1	PDGFRA	SMAD2	XPC
CASP8	EMSY	GMPS	MCL1	PDGFRB	SMAD4	XPO1
CBFA2T3	EP300	GNA13	MDM2	PDK1	SMARCB1	YWHAE
CBFB	EPHA3	GNAQ	MDM4	PER1	SMARCE1	ZMYM2
CBL	EPHA5	GNAS	MDS2	PICALM	SMO	ZNF217
CBLB	EPHB1	GOLGA5	MEF2B	PIK3CA	SNX29	ZNF331
CCDC6	EPS15	GOPC	MEN1	PIK3R1	SOX10	ZNF384
CCNB1IP1	ERBB2 (HER2/NEU)	GPHN	MET	PIK3R2	SPECC1	ZNF521
CCND1	ERBB3 (HER3)	GRIN2A	MITF	PIM1	SPEN	ZNF703

Table 126 – Gene Fusions (RNA)

ABL	FGR	MAML2	NTRK2	RELA
AKT3	FGFR1	MAST1	NTRK3	RET
ALK	FGFR2	MAST2	NUMBL	ROS1
ARHGAP26	FGFR3	MET	PDGFRA	RSPO2
AXL	ERG	MSMB	PDGFRB	RSPO3
BCR	ESR1	MUSK	PIK3CA	TERT
BRAF	ETV1	MYB	PKN1	TFE3
BRD3	ETV4	NOTCH1	PPARG	TFEB
BRD4	ETV5	NOTCH2	PRKCA	THADA
EGFR	ETV6	NRG1	PRKCB	TMPRSS2
EWSR1	INSR	NTRK1	RAF1	

Table 127 – Variant Transcripts

AR-V7	EGFR vIII	MET Exon 14 Skipping

Abbreviations used in this **Example** and throughout the specification, e.g., IHC: immunohistochemistry; ISH: in situ hybridization; CISH: colorimetric in situ hybridization; FISH: fluorescent in situ hybridization; NGS: next generation sequencing; PCR: polymerase chain reaction; CNA: copy number alteration; CNV: copy number variation; MSI: microsatellite instability; TMB: tumor mutational burden.

With whole exome sequencing (WES) and whole transcriptome sequencing (WTS), quantitative sequencing data is available for practically all known genes and transcripts. For example, WES and WTS may query 22,000 or more sequences of interest. In addition to the genes in **Tables 124-125**, **Tables 128-129** provide additional selections of genes of interest, e.g. genes most commonly associated with cancer, that may be of particular interest in molecular profiling cancer samples.

Table 128 – Point Mutations and Indels (DNA)

ABL1	CDK12	HDAC	MAX	PMS1	SDHAF2
AIP	CXCR4	HIST1H3B	MED12	POLD1	SETD2
AKT1	DNMT3A	HIST1H3C	MPL	PPP2R1A	SMARCA4
AMER1	EPHA2	HNF1A	MSH3	PPP2R2A	SOCS1
AR	FANCB	HOXB13	MST1R	PRKACA	SPOP
ARAF	FANCF	HRAS	MUTYH	PRKDC	SRC
ATRX	FANCI	KDM5C	NBN	RABL3	TERT
B2M	FANCM	KDM6A	NOTCH1	RAD51B	TMEM127
BCL2	FAT1	KDR	NRAS	RAD51C	VHL
BCOR	FOXL2	LYN	NTHL1	RAD51D	XRCC1
BTK	FYN	LZTR1	PARP1	RAD54L	YES1
CD79B	GLI2	MAPK1	PHOX2B	RHOA	
CDH1	GNA11	MAPK3	PIK3CB	SDHA	

Table 129 – Point Mutations, Indels and Copy Number Variations (DNA)

ALK	CDK4	FANCA	H3F3A	MITF	PIK3R1	SMAD2
APC	CDK6	FANCC	H3F3B	MLH1	PIM1	SMAD4
ARID1A	CDKN1B	FANCD2	IDH1	MRE11	PMS2	SMARCB1
ARID2	CDKN2A	FANCE	IDH2	MSH2	POLE	SMARCE1
ASXL1	CHEK1	FANCG	IRF4	MSH6	POT1	SMO
ATM	CHEK2	FANCL	JAK1	MTOR	PPARG	SPEN
ATR	CIC	FAS	JAK2	MYCN	PRDM1	STAT3
BAP1	CREBBP	FBXW7	JAK3	MYD88	PRKAR1A	STK11

BARD1	CSF1R	FGFR1	KEAP1	NF1	PTCH1	SUFU
BCL9	CTNNA1	FGFR2	KIT	NF2	PTEN	TNFAIP3
BLM	CTNNB1	FGFR3	KMT2A	NFE2L2	PTPN11	TNFRSF14
BMPR1A	CYLD	FGFR4	KMT2C	NFKBIA	RAD50	TP53
BRAF	DDR2	FH	KMT2D	NPM1	RAF1	TSC1
BRCA1	DICER1	FLCN	KRAS	NSD1	RB1	TSC2
BRCA2	EGFR	FLT1	LCK	NTRK1	RET	U2AF1
BRIP1	EP300	FLT3	MAP2K1	NTRK2	RNF43	WRN
CARD11	ERBB2	FLT4	MAP2K2	NTRK3	ROS1	WT1
CBFB	ERBB3	FUBP1	MAP2K4	PALB2	RUNX1	
CCND1	ERBB4	GATA3	MAP3K1	PBRM1	SDHB	
CCND2	ERCC2	GNA13	MEF2B	PDGFRA	SDHC	
CCND3	ESR1	GNAQ	MEN1	PDGFRB	SDHD	
CDC73	EZH2	GNAS	MET	PIK3CA	SF3B1	

The precise molecular profiles in this **Example** have been and are adjusted over time, including without limitation reasons such as the development of new and updated technologies, biomarker tests and companion diagnostics, and new or updated evidence for biomarker – treatment associations. Thus, for some patient molecular profiles gathered in the past, data for various biomarkers tested with other methods than those in **Tables 122-129** is available and can be used for NGP.

Table 130 presents a view of associations between the biomarkers assessed and various therapeutic agents. Such associations can be determined by correlating the biomarker assessment results with drug associations from sources such as the NCCN, literature reports and clinical trials. The column headed "Agent" provides candidate agents (e.g., drugs or biologics) or biomarker status. In some cases, the agent comprises clinical trials that can be matched to a biomarker status. In some cases, multiple biomarkers are associated with an agent or group of agents. Platform abbreviations are as used throughout the application, e.g., IHC: immunohistochemistry; CISH: colorimetric in situ hybridization; NGS: next generation sequencing; PCR: polymerase chain reaction; CNA: copy number alteration. Tumor Type abbreviations include: TNBC: triple negative breast cancer; NSCLC: non-small cell lung cancer; CRC: colorectal cancer; GEJ: gastroesophageal junction, EBDA: extrahepatic bile duct adenocarcinoma. Biomarker abbreviations include: HRR: Homologous Recombination Repair, which includes the genes ATM, BARD1, BRCA1, BRCA2, BRIP1, CDK12, CHEK1, CHEK2, FANCL, PALB2, RAD51B, RAD51C, RAD51D, RAD54L; MSI: microsatellite instability; MSS: microsatellite stable; MMR: mismatch repair; TMB: tumor mutational burden.

Agents for biomarker PD-L1 identify specific antibodies used in detection assays in the parentheticals.

Table 130 – Biomarker – Treatment Associations

Biomarker	Technology / Alteration	Agent
ALK	IHC, RNA fusion	crizotinib, ceritinib, alectinib, brigatinib (NSCLC), lorlatinib (NSCLC)
	DNA mutation	resistance to crizotinib, alectinib
AR	IHC	bicalutamide, leuprolide (salivary gland tumors)
		enzalutamide, bicalutamide (TNBC)
ATM	DNA mutation	carboplatin, cisplatin, oxaliplatin
		olaparib (prostate)
BRAF	DNA mutation	vemurafenib, dabrafenib, cobimetinib, trametinib
		vemurafenib +(cetuximab or panitumumab)+irinotecan (CRC)
		encorafenib + binimetinib (melanoma)
		dabrafenib+trametinib (anaplastic thyroid and NSCLC)
		atezolizumab + cobimetinib + vemurafenib (melanoma)
		cetuximab + encorafenib (CRC)
		cetuximab, panitumumab with BRAF and or MEK inhibitors (CRC)
BRCA1/2	DNA mutation	carboplatin, cisplatin, oxaliplatin
		niraparib (ovarian, prostate), olaparib (breast, cholangiocarcinoma, ovarian, pancreatic, prostate), rucaparib (ovarian, pancreatic, prostate), talazoparib (breast), veliparib combination (pancreatic)
		resistance to olaparib, niraparib, rucaparib with reversion mutation
EGFR	DNA mutation	afatinib (NSCLC)
		afatinib + cetuximab (T790M; NSCLC)
		erlotinib, gefitinib (NSCLC and CUP)
		osimertinib, dacomitinib (NSCLC)
ER	IHC	endocrine therapies
		everolimus, temsirolimus (breast)
		palbociclib, ribociclib, abemaciclib (breast)
ERBB2	IHC, CISH, DNA	trastuzumab, lapatinib, neratinib (breast), pertuzumab, T-

(HER2)	mutation, CNA	DM1, fam-trastuzumab deruxtecan-nxki, tucatinib
	DNA mutation	T-DM1 (NSCLC)
ER/PR/ERBB2 (HER2)	IHC, CISH	sacituzumab govitecan (TNBC)
ESR1	DNA mutation	exemestane + everolimus, fulvestrant, palbociclib combination therapy (breast)
		resistance to aromatase inhibitors (breast)
FGFR2/3	DNA mutation, RNA fusion	erdafitinib (urothelial bladder), pemigatinib (cholangiocarcinoma)
HRR	DNA mutation	olaparib (prostate)
IDH1	DNA mutation	temozolomide (high grade glioma)
		ivosidenib (cholangiocarcinoma and EBDA)
KIT	DNA mutation	imatinib
		regorafenib, sunitinib (both GIST)
KRAS	DNA mutation	resistance to cetuximab, panitumumab (CRC)
		resistance to erlotinib/gefitinib (NSCLC)
		resistance to trastuzumab, lapatinib, pertuzumab (CRC)
MET	RNA exon skipping, DNA exon skipping, CNA	cabozantinib, crizotinib (NSCLC)
MGMT	Pyrosequencing (Methylation)	temozolomide (high grade glioma)
MMR Deficiency	IHC, DNA mutation	pembrolizumab
MSI		pembrolizumab, nivolumab (CRC, small bowel adenocarcinoma), nivolumab+ipilimumab (CRC, small bowel adenocarcinoma)
MMR Proficiency	IHC, DNA mutation	pembrolizumab + lenvatinib (endometrial)
MSS		
NRAS	DNA mutation	resistance to cetuximab, panitumumab (CRC)
		resistance to trastuzumab, lapatinib, pertuzumab (CRC)
NTRK1/2/3	RNA fusion	entrectinib, larotrectinib
	DNA mutation	resistance to entrectinib, larotrectinib
PALB2	DNA mutation	olaparib (pancreatic and prostate), veliparib combination

		(pancreatic)
PDGFRA	DNA mutation	imatinib , avapritinib (GIST), sunitinib
PD-L1	IHC	pembrolizumab (22c3 TPS in NSCLC; 22c3 CPS in cervical, GEJ/gastric, head & neck, urothelial and non-urothelial bladder, vulvar)
		atezolizumab (SP142 IC urothelial bladder cancer and SP142 IC & TC NSCLC)
		pembrolizumab + chemotherapy (22c3 CPS in TNBC)
		atezolizumab + nab-paclitaxel (SP142 IC in TNBC)
		nivolumab/ipilimumab combination (28-8 NSCLC)
		avelumab (non-urothelial bladder and Merkel cell)
PIK3CA	DNA mutation	alpelisib + fulvestrant (breast)
POLE	DNA mutation	pembrolizumab (endometrial and CRC)
PR	IHC	endocrine therapies
RET	RNA fusion	cabozantinib, vandetanib, selpercatinib, pralsetinib (NSCLC)
	DNA mutation	vandetanib, cabozantinib, selpercatinib (thyroid); resistance to vandetanib, cabozantinib
ROS1	IHC, RNA fusion	crizotinib, ceritinib, entrectinib, lorlatinib (NSCLC)
TMB	DNA mutation	pembrolizumab
TOP2A	CISH	doxorubicin, liposomal doxorubicin, epirubicin (all breast)

Example 2: Genomic Prevalence Score (GPS) using a DNA NGS Panel to Predict Tumor Types

This Example describes the development of a Genomic Prevalence Score system (which may also be referred to herein as GPS; Genomic Profiling Similarity; Molecular Disease Classifier; MDC) to predict tumor type of a biological sample using a next generation sequencing panel to assess genomic DNA. This Example further applies GPS to the prediction of tumor types for an expanded specimen cohort, with closer analysis of Carcinoma of Unknown Primary (CUP; aka Cancer of Unknown Primary).

Current standard histological diagnostic tests are not able to determine the origin of metastatic cancer in as many as 10% of patients[1], leading to a diagnosis of cancer of unknown primary (CUP). The lack of a definitive diagnosis can result in administration of suboptimal treatment regimens and poor outcomes. Gene expression profiling has been used to identify the tissue of origin but suffers from a number of inherent limitations. These limitations impair performance in identifying tumors

with low neoplastic percentage in metastatic sites which is where identification is often most needed[2]. The GPS system provided herein was developed using data for genomic DNA sequencing of a 592 gene panel (see description in **Example 1**, with panel comprises of biomarkers in **Tables 123-125**) coupled with a machine learning platform to aid in the diagnosis of cancer. The algorithm created was trained on 34,352 cases and tested on 15,473 unambiguously diagnosed cases. The performance of the algorithm was then assessed on 1,662 CUP cases. The GPS accurately predicted the tumor type in the labeled data set with sensitivity, specificity, PPV, and NPV of 90.5%, 99.2%, 90.5% and 99.2% respectively. Performance was consistent regardless of the percentage of tumor nuclei or whether or not the specimen had been obtained from a site of metastasis. Pathologic re-evaluation of selected discordant cases resulted in confirmation of GPS results and clinical utility. Moreover, all genomic markers essential for therapy selection are assessed in this assay, maximizing the clinical utility for patients within a single test.

Introduction

Carcinoma of Unknown Primary (CUP) represents a clinically challenging heterogeneous group of metastatic malignancies in which a primary tumor remains elusive despite extensive clinical and pathologic evaluation. Approximately 2-4% of cancer diagnoses worldwide comprise CUP[3]. In addition, some level of diagnostic uncertainty with respect to an exact tumor type classification is a frequent occurrence across oncologic subspecialties. Efforts to secure a definitive diagnosis can prolong the diagnostic process and delay treatment initiation. Furthermore, CUP is associated with poor outcome which might be explained by use of suboptimal therapeutic intervention. Immunohistochemical (IHC) testing is the gold standard method to diagnose the site of tumor origin, especially in cases of poorly differentiated or undifferentiated tumors. Assessing the accuracy in challenging cases and performing a meta-analysis of these studies reported that IHC analysis had an accuracy of 66% in the characterization of metastatic tumors[4-9]. Since therapeutic regimes are highly dependent upon diagnosis, this represents an important unmet clinical need. To address these challenges, assays aiming at tissue-of-origin (TOO) identification based on assessment of differential gene expression have been developed and tested clinically. However, integration of such assays into clinical practice is hampered by relatively poor performance characteristics (from 83% to 89%[11-14]) and limited sample availability. For example, a recent commercial RNA-based assay has a sensitivity of 83% in a test set of 187 tumors and confirmed results on only 78% of a separate 300 sample validation set[14]. This may, at least in part, be a consequence of limitations of typical RNA-based assays in regards to normal cell contamination, RNA stability, and dynamics of RNA expression. Nevertheless, initial clinical studies demonstrate possible benefit of matching treatments to tumor types predicted by the assay[15]. With increasing availability of comprehensive molecular profiling assays, in particular next-generation DNA sequencing, genomic features have been incorporated in CUP treatment strategies[16]. While this approach rarely supports unambiguous identification of the TOO, it does reveal targetable molecular alterations in some of the patients[16].

In this Example, we pursued a different strategy of TOO identification by using a novel machine-learning approach as provided herein to build TOO classifiers based on data from a large NGS genomic DNA panel that assesses hundreds of gene sequences and various attributes thereof (see **Example 1**) and has been broadly used in clinical treatment of cancer patients. This computational classification system identified TOO at an accuracy significantly exceeding that of previously published technologies. Moreover, the 592-gene NGS assay simultaneously determines the GPS and presence of underlying genetic abnormalities that guide treatment selection (see **Example 1**), thus generating substantially increased clinical utility in a single test.

Methodology

Study Design

GPS can be used with patients previously diagnosed with cancer in various settings, including without limitation as a confirmatory or quality control (QC) measure for every case wherein molecular profiling is performed. GPS may also be particularly useful in guiding treatment of cases having a diagnosis of cancer of unknown primary (CUP) or any cases having an uncertain diagnosis. From a database of cases that have profiled with the 592-gene NGS assay, we selected 55,780 cases with a pathology report available. This study was performed with IRB approval. This data set was split into three cohorts: 34,352 cases with an unambiguous diagnosis; 15,473 cases with an unambiguous diagnosis reserved as an independent validation set; and 1,662 CUP cases. All cases were de-identified prior to analysis.

The general study design **500** is shown in **FIG. 5A**. Starting with the 34,352 cases with an unambiguous diagnosis, the machine learning algorithms were trained **501** using 27,439 samples at a training cohort and 6,913 samples were used for validation. Once models were trained and optimized, the algorithm was locked **502**. The 15,473 cases with an unambiguous diagnosis were used as an independent validation set **503**. 1,662 CUP cases **504** were used to assess classification and prospective validation **505** was performed with over 10,000 clinical cases.

592 NGS Panel

Next generation sequencing (NGS) was performed on genomic DNA isolated from formalin-fixed paraffin-embedded (FFPE) tumor samples using the NextSeq platform (Illumina, Inc., San Diego, CA). Matched normal tissue was not sequenced. A custom-designed SureSelect XT assay was used to enrich 592 whole-gene targets (Agilent Technologies, Santa Clara, CA). The particular targets are listed in **Tables 123-125** above. All variants were detected with > 99% confidence based on allele frequency and amplicon coverage, with an average sequencing depth of coverage of > 500 and an analytic sensitivity of 5%. Prior to molecular testing, tumor enrichment was achieved by harvesting targeted tissue using manual microdissection techniques. Genetic variants identified were interpreted by board-certified molecular geneticists and categorized as 'pathogenic,' 'presumed pathogenic,' 'variant of unknown significance,' 'presumed benign,' or 'benign,' according to the American College of Medical Genetics and Genomics (ACMG) standards. When assessing mutation frequencies

of individual genes, 'pathogenic,' and 'presumed pathogenic' were counted as mutations while 'benign', 'presumed benign' variants and 'variants of unknown significance' were excluded.

Tumor Mutation Load (TML) was measured (592 genes and 1.4 megabases [MB] sequenced per tumor) by counting all non-synonymous missense mutations found per tumor that had not been previously described as germline alterations. The threshold to define TML-high was greater than or equal to 17 mutations/MB and was established by comparing TML with MSI by fragment analysis in CRC cases, based on reports of TML having high concordance with MSI in CRC.

Microsatellite Instability (MSI) was examined using over 7,000 target microsatellite loci and compared to the reference genome hg19 from the University of California, Santa Cruz (UCSC) Genome Browser database. The number of microsatellite loci that were altered by somatic insertion or deletion was counted for each sample. Only insertions or deletions that increased or decreased the number of repeats were considered. Genomic variants in the microsatellite loci were detected using the same depth and frequency criteria as used for mutation detection. MSI-NGS results were compared with results from over 2,000 matching clinical cases analyzed with traditional PCR-based methods. The threshold to determine MSI by NGS was determined to be 46 or more loci with insertions or deletions to generate a sensitivity of > 95% and specificity of > 99%.

Copy number alteration (CNA, also referred to as copy number variation or CNV herein) was tested using the NGS panel and was determined by comparing the depth of sequencing of genomic loci to a diploid control as well as the known performance of these genomic loci. Calculated gains of 6 copies or greater were considered amplified.

For further description of the 592 NGS panel and MSI and TML calling, see **Example 1**; and International Patent Publication WO 2018/175501 A1, published September 27, 2018 and based on Int'l Patent Application PCT/US2018/023438 filed March 20, 2018, which is incorporated by reference herein in its entirety.

Machine Learning

The GPS system was built using an artificial intelligence platform leveraging the framework provided herein, which uses multiple models to vote against one another to determine a final result. See, e.g., **FIGs. 1F-1G** and accompanying text. A set of 115 distinct tumor site and histology classes were used to generate subpopulations of patients, stratified by primary location (e.g., prostate) and histology (e.g., adenocarcinoma), and combined as "disease type" or "cancer type" (e.g., prostate adenocarcinoma). The 115 disease/cancer types included: adrenal cortical carcinoma; anus squamous carcinoma; appendix adenocarcinoma, NOS; appendix mucinous adenocarcinoma; bile duct, NOS, cholangiocarcinoma; brain astrocytoma, anaplastic; brain astrocytoma, NOS; breast adenocarcinoma, NOS; breast carcinoma, NOS; breast infiltrating duct adenocarcinoma; breast infiltrating lobular carcinoma, NOS; breast metaplastic carcinoma, NOS; cervix adenocarcinoma, NOS; cervix carcinoma, NOS; cervix squamous carcinoma; colon adenocarcinoma, NOS; colon carcinoma, NOS; colon mucinous adenocarcinoma; conjunctiva malignant melanoma, NOS; duodenum and ampulla

adenocarcinoma, NOS; endometrial adenocarcinoma, NOS; endometrial carcinosarcoma; endometrial endometrioid adenocarcinoma; endometrial serous carcinoma; endometrium carcinoma, NOS; endometrium carcinoma, undifferentiated; endometrium clear cell carcinoma; esophagus adenocarcinoma, NOS; esophagus carcinoma, NOS; esophagus squamous carcinoma; extrahepatic cholangio, common bile, gallbladder adenocarcinoma, NOS; fallopian tube adenocarcinoma, NOS; fallopian tube carcinoma, NOS; fallopian tube carcinosarcoma, NOS; fallopian tube serous carcinoma; gastric adenocarcinoma; gastroesophageal junction adenocarcinoma, NOS; glioblastoma; glioma, NOS; gliosarcoma; head, face or neck, NOS squamous carcinoma; intrahepatic bile duct cholangiocarcinoma; kidney carcinoma, NOS; kidney clear cell carcinoma; kidney papillary renal cell carcinoma; kidney renal cell carcinoma, NOS; larynx, NOS squamous carcinoma; left colon adenocarcinoma, NOS; left colon mucinous adenocarcinoma; liver hepatocellular carcinoma, NOS; lung adenocarcinoma, NOS; lung adenosquamous carcinoma; lung carcinoma, NOS; lung mucinous adenocarcinoma; lung neuroendocrine carcinoma, NOS; lung non-small cell carcinoma; lung sarcomatoid carcinoma; lung small cell carcinoma, NOS; lung squamous carcinoma; meninges meningioma, NOS; nasopharynx, NOS squamous carcinoma; oligodendroglioma, anaplastic; oligodendroglioma, NOS; ovary adenocarcinoma, NOS; ovary carcinoma, NOS; ovary carcinosarcoma; ovary clear cell carcinoma; ovary endometrioid adenocarcinoma; ovary granulosa cell tumor, NOS; ovary high-grade serous carcinoma; ovary low-grade serous carcinoma; ovary mucinous adenocarcinoma; ovary serous carcinoma; pancreas adenocarcinoma, NOS; pancreas carcinoma, NOS; pancreas mucinous adenocarcinoma; pancreas neuroendocrine carcinoma, NOS; parotid gland carcinoma, NOS; peritoneum adenocarcinoma, NOS; peritoneum carcinoma, NOS; peritoneum serous carcinoma; pleural mesothelioma, NOS; prostate adenocarcinoma, NOS; rectosigmoid adenocarcinoma, NOS; rectum adenocarcinoma, NOS; rectum mucinous adenocarcinoma; retroperitoneum dedifferentiated liposarcoma; retroperitoneum leiomyosarcoma, NOS; right colon adenocarcinoma, NOS; right colon mucinous adenocarcinoma; salivary gland adenoid cystic carcinoma; skin melanoma; skin melanoma; skin merkel cell carcinoma; skin nodular melanoma; skin squamous carcinoma; skin trunk melanoma; small intestine adenocarcinoma; small intestine gastrointestinal stromal tumor, NOS; stomach gastrointestinal stromal tumor, NOS; stomach signet ring cell adenocarcinoma; thyroid carcinoma, anaplastic, NOS; thyroid carcinoma, NOS; thyroid papillary carcinoma of thyroid; tonsil, oropharynx, tongue squamous carcinoma; transverse colon adenocarcinoma, NOS; urothelial bladder adenocarcinoma, NOS; urothelial bladder carcinoma, NOS; urothelial bladder squamous carcinoma; urothelial carcinoma, NOS; uterine endometrial stromal sarcoma, NOS; uterus leiomyosarcoma, NOS; uterus sarcoma, NOS; uveal melanoma; vaginal squamous carcinoma; vulvar squamous carcinoma. Note that NOS, or "Not Otherwise Specified," is a subcategory in systems of disease/disorder classification such as ICD-9, ICD-10, or DSM-IV, and is generally but not exclusively used where a more specific diagnosis was not made.

For training the GPS, all 115 disease types were trained against each other in a pairwise comparison approach using the training set to generate 6555 model signatures, where each signature is built to differentiate between a pair of disease types. The signatures were generated using Gradient Boosted Forests and applied a voting module approach as described herein.

The models were validated using the test cases. Each test case was processed individually through all 6555 signatures, thereby providing a pairwise analysis between every disease type for every case. The results are analyzed in a 115 x 115 matrix where each column and each row is a single disease type and the cell at the intersection is the probability that a case is one disease type or the other. The probabilities for each disease type are summed for each column which results in 115 disease types with their probability sums. These disease types are ranked by their probability sums.

The disease types were then used to determine a final probability for each case belonging to a superset of 15 distinct organ groups, which include the following: Colon; Liver, Gall Bladder, Ducts; Brain; Breast; Female Genital Tract and Peritoneum (FGTP); Esophagus; Stomach; Head, Face or Neck, not otherwise specified (NOS); Kidney; Lung; Pancreas; Prostate; Skin/Melanoma; and Bladder. For each case, each of these organs can be assigned a probability which will be used to make the primary origin prediction(s). **Tables 2-116** above list selections of features that contribute to the disease type predictions, where each row in the table represents a feature ranked by Importance. As noted, the titles of **Tables 2-116** indicate how the 115 disease types relate to the 15 organ groups, as the tables are titled in the format "disease type – organ group." As an example, the title heading of **Table 2** is "Adrenal Cortical Carcinoma – Adrenal Gland," indicating that the disease type is adrenal cortical carcinoma, which is placed within the organ group is adrenal gland.

FIG. 5B shows an example 115x115 matrix generated for a test case of prostate origin (i.e., Primary Site: Prostate Gland; Histology: Adenocarcinoma). In the figure, the X and Y legends are the 115 disease types listed above. Each row is the probability of a "negative" call (probability < 0.5) and each column is the probability of a positive call, as noted above. The shaded squares in the matrix represent probability scores ≥ 0.98. The arrow indicates disease type "prostate adenocarcinoma." The probability sum for this case for prostate was 114.3 out of a possible 115.

Further details can be found in Abraham J., et al. Genomic Profiling Similarity, Int'l Patent Publication WO2020146554, which publication is herein incorporated by reference in its entirety.

Results

Retrospective Validation

Using the machine learning approach, a probability was assigned to each case that the case was from one of the 15 distinct organ groups. The probability may be referred to as the GPS Score. Of the 15,473 cases with an unambiguous diagnosis used as an independent validation set (see **FIG. 5A 503**), 6229 cases that had a GPS Score of >0.95. Of those, 98.4% were concordant with the case-assigned result. The 98.4% concordance exceeded our acceptance criteria for validating the GPS Scores >0.95. This criteria was greater than 95% accuracy when presenting a score >0.95. The GPS

Score had extremely high performance when assigning scores of 0 to organ groups (i.e., probability of the tumor sample being from that organ group is determined by GPS as zero). The percentage of the time that a tumor type that does not match the case was given a zero GPS Score (12270/12279) was 99.92%.

FIG. 5C shows the Scores for the 6229 cases with GPS Scores > 0.95 plotted against the probability of match for each sample. The resulting correlation coefficient of 0.990 indicates GPS Score is highly correlated to accuracy.

Analytical sensitivity of the GPS Score was determined by evaluating performance relative to two distinct parameters: (1) tumor percentage, and (2) average read depth per sample. To evaluate tumor percentage, accuracy of the GPS relative to the case-assigned organ type was determined. **FIG. 5D** shows a correlation chart for the data grouped into ranges of 20-49%, 50-80% and >80% tumor content. The figure indicates that the GPS Score is insensitive to tumor percentage. **FIG. 5E** shows a correlation chart for the data used to evaluate read depth. The accuracy of the GPS Score relative to the case-assigned organ type was determined with classification of read depths between 300-500X and >500X. As with tumor percentage, the figure indicates that the GPS Score was insensitive to read depth. In both cases, the correlation coefficient according to Pearson's r remained greater than 98% for each data grouping.

We also found that the GPS Score was robust to metastasis. **Table 131** shows performance metrics on subsets of the test data from a primary site (N = 8,437), metastatic site (6,690), and samples with low (9,492) and high tumor percentages (5,945).

Table 131 – Performance metrics of assay with noted characteristics

	Sensitivity	Specificity	PPV	NPV	Accuracy	Call Rate
Primary	90.9%	98.0%	91.1%	98.9%	97.6%	97.3%
Metastatic	89.0%	97.9%	89.3%	98.2%	96.9%	97.6%
20-50% Tumor	90.3%	98.2%	90.6%	98.5%	97.5%	97.1%
>50% Tumor	90.3%	98.2%	90.6%	98.5%	97.5%	97.1%

The performance held across multiple tumor types. **Table 132** shows performance metrics and cohort sizes of subsets of the independent test dataset where the primary tumor site was known. FGTP represents female genital tract and peritoneum.

Table 132 – Performance metrics of assay across tumor types

Tumor Type	Train N	Test N	Sensitivity	Specificity	PPV	NPV	Accuracy	Call Rate
Head, Face, Neck	299	144	45.4%	100.0%	96.4%	99.6%	99.6%	82.6%

Melanoma	976	402	85.0%	99.9%	94.3%	99.6%	99.5%	96.3%
FGTP	8,872	4,115	93.4%	98.3%	95.4%	97.6%	97.0%	98.8%
Prostate	785	477	96.1%	99.8%	94.7%	99.9%	99.7%	96.6%
Brain	1,554	479	93.3%	99.8%	93.5%	99.8%	99.6%	96.0%
Colon	5,805	2,532	94.5%	98.5%	92.9%	98.9%	97.9%	98.9%
Kidney	426	178	84.1%	99.9%	91.7%	99.8%	99.8%	88.2%
Bladder	447	304	60.6%	99.9%	89.4%	99.3%	99.1%	91.8%
Breast	3,324	1,386	90.9%	98.7%	87.9%	99.1%	98.0%	98.3%
Lung	7,744	3,540	96.0%	95.4%	86.3%	98.7%	95.5%	98.2%
Pancreas	1,637	708	83.7%	99.3%	84.6%	99.2%	98.5%	98.3%
Gastroesophageal	1,521	743	72.0%	99.3%	82.6%	98.6%	98.0%	93.8%
Liver, Gallbladder, Ducts	734	364	57.7%	99.7%	82.2%	99.0%	98.8%	92.6%

The GPS Score had extremely high performance when assigning scores of 0 to organ groups (i.e., probability of the tumor sample being from that organ group is determined by GPS as less than 0.001). Of the 15,473 validation cases evaluated, 12,279 had a GPS Score of 0 for one or more organ types. The percentage of the time that a tumor type that did not match the case was given a zero GPS Score (12270/12279) was 99.92%, which exceeded our acceptance criteria for validating the GPS Zero% scores. The criteria was greater than 99.9% accuracy when presenting a score of 0. Thus, the zero score was highly accurate. There were only nine cases that had a GPS Score of 0 for the case-assigned organ result case.

Table 133 shows performance metrics of the GPS algorithm on the independent test set of 15,473 cases as compared to other methods currently available. In the table and those below, "Sensitivity" is the probability of getting a positive test result for tumors with the tumor type and therefore relates to the potential of GPS to recognize the tumor type; "Specificity" is the probability of a negative result in a subject without the tumor type and therefore relates to the GPS' ability to recognize subjects without the tumor type, i.e. to exclude the tumor type; Positive Predictive Value ("PPV") is the probability of having the tumor type of interest in a subject with positive result for that tumor type, and therefore PPV represents a proportion of patients with positive test result in total of subjects with positive result; NPV is the probability of not having the tumor type in a subject with a negative test result, and therefore provides a proportion of subjects without the tumor type with a negative test result in total of subjects with negative test results; Accuracy represents the proportion of true positives and true negatives in the text population; and Call Rate is the proportion of samples for which GPS is able to provide a prediction.

Table 133 – Performance of GPS on Validation Set

Assay	Overall Accuracy	PPV	NPV	Sensitivity / PPA	Specificity / NPA	Call Rate	N
MDC/GPS	98.4%	90.5%	99.2%	90.5%	99.2%	97.5%	15,473
Cancer Genetics Tissue of Origin	94.1%[18]	NR	NR	88.5% [17]	99.1% [17]	89% [18]	462[17] 36[18]
CancerTYPE ID[2]	NR	83%	99%	83%	99%	78%	187
Gamble AR, 1993[19]	NR	NR	NR	64%	NR	100%	90
Brown, RW, 1997[20]	NR	NR	NR	66%	NR	87%	128
Dennis, JL, 2005[21]	NR	NR	NR	67%	NR	100%	452
Park SY, 2007[22]	NR	NR	NR	65%	NR	78%	374

Prospective Validation

A target of 10,000 prospective samples were evaluated by the GPS Score platform based on clinical samples incoming for molecular profiling using the 592 NGS gene panel. The GPS Score for an organ group was >0.95 for 2857 cases. Of those, 54 cases had a GPS Score which differed from the organ group listed on the incoming case (i.e., as listed by the ordering physician) and were flagged for further pathological review. Pathologists reviewed those 54 cases, plus an additional 12 cases with GPS scores ≤0.95 and requested by the pathologist for various reasons (Score close to 0.95, suspicious IHC findings, etc). There was a 43.9% (29/66) response from pathology review that the results obtained via the GPS system were considered "reasonable." The pathology review resulted in changes to the tumor type from what was originally reported from the ordering physician for 11 cases. The results of this evaluation exceeded our acceptance criteria for validating the capability of the GPS Score to provide evidence to support a new diagnosis. This acceptance criteria was whether pathologists consider the information reasonable in greater than 25% of the cases and the information results in any change in diagnosis that may affect patient treatment. In these cases, a change in tumor origin may affect such treatment. Thus, automated flagging of discordant tumor type by GPS may positively influence the course of treatment of a substantial number of patients.

Analysis of CUP

Validation of a CUP assay at the individual patient level is a fundamentally difficult as the "truth" may be unknown. However, population based methods can be used to gain greater insight into the performance of the GPS classifier and generally validate its performance. To accomplish this, we compared the frequency of mutations across known patient populations to the frequency in the predicted group. For example, the frequency of BRAF mutations in colon cancer in the known patient cohort is 10.3% and is 4.8% in all non-colon cancer patients. The frequency of BRAF in the CUP cases that the classifier called colon is 10.3% and is 4.9% in the CUP cases the classifier called as non-colon. In this way we can show that the population of CUP cases that are classified as a specific cancer type matches the population of each specific tumor type. A subset of markers we used in this manner are shown in **Table 134**, demonstrating the similarities of the GPS predicted CUP populations to the actual populations. The data for correlation of between the frequencies for the predicted CUP cases and the training set show that the predicted populations most closely resemble the actual population with the exception of brain cancer, which, without being bound by theory, may be due to small sample size, with only 17 CUP cases predicted to be brain. These data together show that the GPS can classify CUP at the population level into classes consistent with other molecular characteristics of the tumors.

Table 134 – Frequencies of variants detected or observed medians among notable biomarkers per tumor type

		Of This Tumor Type		Not Of This Tumor Type	
Marker	**Tumor Type**	**Train + Test***	**CUP****	**Train + Test***	**CUP****
BRAF	Colon	10.3%	10.3%	4.8%	4.9%
BRAF	Lung	6.2%	6.3%	5.6%	5.7%
BRAF	Melanoma	39.1%	38.4%	4.8%	4.9%
BRCA1	Breast	7.0%	7.1%	6.4%	6.4%
BRCA1	FGTP	8.6%	8.6%	5.7%	5.8%
BRCA1	Melanoma	9.9%	10.3%	6.4%	6.4%
BRCA1	Prostate	4.1%	4.2%	6.5%	6.5%
cKIT	Gastroesophageal	5.8%	5.5%	3.4%	3.4%
cKIT	Lung	4.3%	4.3%	3.3%	3.3%
EGFR	Brain	17.6%	17.2%	6.5%	6.5%
EGFR	Lung	16.1%	15.4%	4.3%	4.4%
KRAS	Colon	50.0%	49.1%	16.4%	16.6%
KRAS	Lung	26.4%	26.1%	20.8%	20.7%
KRAS	Pancreas	84.2%	83.3%	19.0%	18.8%
PIK3CA	Breast	31.5%	31.1%	13.5%	13.5%
PIK3CA	FGTP	21.3%	21.1%	13.1%	13.0%

PIK3CA	Lung	6.3%	6.6%	17.8%	17.7%
TP53	Head and Neck	45.4%	45.4%	61.8%	61.1%
TP53	Melanoma	28.2%	29.9%	62.6%	61.9%

* Represents the observed value among the known tumor type of the combined training and testing datasets.

** Represents the observed value among CUP cases predicted to be of the tumor type in each row.

Cancer of unknown primary remains a substantial problem for both clinicians and patients, diagnosis can be aided with the GPS algorithms provided herein. The tumor type predictors can render a histologic diagnosis to CUP cases that can inform treatment and potentially improve outcomes. Our NGS analysis of tumors (see **Example 1**) and GPS provided here return both diagnostic and therapeutic information that optimize patient treatment strategy from a single test. This method provides a substantial improvement over the current standard of multiple tests that require more tissue.

References (as indicated by superscripted numbers in the text of the Example)

1. Haskell CM, et al. Metastasis of unknown origin. Curr Probl Cancer. 1988 Jan-Feb;12(1):5-58. Review. PubMed PMID: 3067982.

2. Erlander MG, et al. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. J Mol Diagn. 2011 Sep;13(5):493-503. doi: 10.1016/j.jmoldx.2011.04.004. Epub 2011 Jun 25.

3. Varadhachary. New Strategies for Carcinoma of Unknown Primary: the role of tissue of origin molecular profiling. Clin Cancer Res. 2013 Aug 1;19(15):4027-33. DOI: 10.1158/1078-0432.CCR-12-3030

4. Brown RW, et al. Immunohistochemical identification of tumor markers in metastatic adenocarcinoma: a diagnostic adjunct in the determination of primary site. Am J Clin Pathol 1997, 107:12e19

5. Dennis JL, et al. Markers of adenocarcinoma characteristic of the site of origin: development of a diagnostic algorithm. Clin Cancer Res 2005, 11:3766e3772

6. Gamble AR, et al. Use of tumour marker immunoreactivity to identify primary site of metastatic cancer. BMJ 1993, 306:295e298

7. Park SY, et al. Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. Arch Pathol Lab Med 2007, 131:1561e1567

8. DeYoung BR, Wick MR. Immunohistologic evaluation of metastatic carcinomas of unknown origin: an algorithmic approach. Semin Diagn Pathol 2000, 17:184e193

9. Anderson GG, Weiss LM. Determining tissue of origin for metastatic cancers: meta-analysis and literature review of immunohistochemistry performance. Appl Immunohistochem Mol Morphol 2010, 18:3e8

10. Erlander MG, et al. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. J Mol Diagn 2011, 13:493e503

11. Pillai R, et al. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. J Mol Diagn 2011, 13:48e56

12. Rosenwald S, et al. Validation of a microRNA-based qRT-PCR test for accurate identification of tumor tissue origin. Mod Pathol 2010, 23:814e823

13. Kerr SE, et al. Multisite validation study to determine performance characteristics of a 92-gene molecular cancer classifier. Clin Cancer Res 2012, 18:3952e3960

14. Kucab JE, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019 May 2;177(4):821-836.e16. doi: 10.1016/j.cell.2019.03.001. Epub 2019 Apr 11. PubMed PMID: 30982602; PubMed Central PMCID: PMC6506336.

15. Hainsworth JD, et al, Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. J Clin Oncol. 2013 Jan 10;31(2):217-23. doi: 10.1200/JCO.2012.43.3755. Epub 2012 Oct 1.

16. Ross JS, et al. Comprehensive Genomic Profiling of Carcinoma of Unknown Primary Site New Routes to Targeted Therapies. JAMA Oncol. 2015;1(1):40-49. doi:10.1001/jamaoncol.2014.216

Example 3: Machine learning analysis using genomic and transcriptomic profiles to accurately predict tumor attributes

This disclosure provides a machine learning based classifiers to predict the origin of a tumor sample, or TOO (tissue-of-origin), and related attributes based on analysis of genomic DNA (*see, e.g.,* **Example 2**) and based on analysis of transcriptome analysis. *See, e.g.,* **FIG. 4A**, **Tables 117-120**, and accompanying description. As noted herein, DNA and RNA each have advantages and disadvantages as biological analytes. Without being bound by theory, we hypothesized that a combination of genomic DNA analysis with RNA transcriptome analysis may provide optimal results. Advanced machine learning analysis may take advantage of the strengths of each analyte while curtailing the weaknesses. We term this combined classifier a "panomic" predictor. This Example details this panomic classifier, which may be referred to as "MI GPSai" in this Example.

Cancer of Unknown Primary (CUP) occurs in 3–5% of patients when standard histological diagnostic tests are unable to determine the origin of metastatic cancer. Typically, a CUP diagnosis is treated empirically and has poor outcome, with median overall survival less than one year. Gene expression profiling alone has been used to identify the tissue of origin (TOO) but struggles with low neoplastic percentage in metastatic sites which is where identification is often most needed. This

Example provides a "Genomic Prevalence Score," or "GPS," which uses DNA sequencing and whole transcriptome data coupled with machine learning to aid in the diagnosis of cancer. The system implementing the GPS, termed "MI GPSai," was trained on genomic data from 34,352 cases and genomic and transcriptomic data from 23,137 cases and was validated on 19,555 cases. MI GPSai predicted the tumor type in the labeled data set with an accuracy of over 94% on 93% of cases while deliberating amongst 21 possible categories of cancer: breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma. When also considering the second highest prediction, the accuracy increased to 97%. Additionally, MI GPSai rendered a prediction for 71.7% of CUP cases. Pathologist evaluation of discrepancies between submitted diagnosis and MI GPSai predictions resulted in change of diagnosis in 41.3% of the time. MI GPSai provides clinically meaningful information in a large proportion of CUP cases and inclusion of MI GPSai in clinical routine could improve diagnostic fidelity. Moreover, all genomic markers essential for therapy selection are assessed in this assay, maximizing the clinical utility for patients within a single test.

Introduction

Carcinoma of Unknown Primary (CUP) represents a clinically challenging heterogeneous group of metastatic malignancies in which a primary tumor remains elusive despite extensive clinical and pathologic evaluation. CUPs comprise approximately 3–5% of cancer diagnoses worldwide [1] and efforts to secure a definitive diagnosis can prolong the diagnostic process and delay treatment initiation. Furthermore, CUP is associated with poor outcome which may be at least partially explained by use of suboptimal therapeutic interventions since there is general agreement that CUP tumors retain the biologic properties of the putative primary malignancy [1], [2]. Immunohistochemical (IHC) testing has long been the gold standard method to diagnose the site of tumor origin, especially in cases of poorly-differentiated or undifferentiated tumors. Meta-analysis of studies assessing the accuracy of IHC in challenging cases reported an accuracy of 60–70% in the characterization of metastatic tumors [3], [4], [5], [6]. Since therapeutic regimens may depend upon diagnosis, there is a need for improved diagnosis of CUP. To address these challenges, assays aiming at tissue-of-origin (TOO) identification based on assessment of differential gene expression have been developed and tested clinically. However, integration of such assays into clinical practice is hampered by relatively poor performance characteristics, e.g., low accuracy such as < 90% combined with high call rate such as 100% or higher accuracy such as <~90% combined with low call rate such as < 90%, and limited sample availability. See **Table 135**. Nevertheless, initial clinical studies demonstrate possible benefit of matching treatments to tumor types predicted by the assay [8]. With increasing

availability of comprehensive molecular profiling assays, particularly next-generation DNA sequencing, genomic features have been incorporated in CUP treatment strategies [9]. Although this approach has not been a panacea for unambiguous identification of the TOO, it has revealed targetable molecular alterations in some patients [9].

Table 135. Landscape of tissue of origin approaches

Assay	Cancer Categories	N Independent Test Set	Accuracy (%)	Cases Called (%)
MI GPSai	21	13,661	94.7	93
PCAWG 2020 [32]	14	1436	88	100
MSK IMPACT 2019 [10]	22	11,644	74.1	100
Cancer Genetics Tissue of Origin 2012 [11]	9	27	94.1	89
Biotheranostics CancerTYPE ID 2011 [7]	30	187	83	100
Park SY 2007 [5]	7	60	75	78
Dennis JL 2005 [12]	7	130	88	100
Brown RW 1997 [6]	5	128	66	86
Gamble AR 1993 [13]	14	100	70	100

As described above and further detailed in this Example, we used a machine-learning approach to build TOO classifiers based on data from a large next-generation DNA sequencing panel in conjunction with data from whole transcriptome sequencing, which are both used broadly for routine molecular tumor profiling. *See, e.g.,* **Example 1**. This panomic computational classification system identified TOO at an accuracy significantly exceeding that of other currently available technologies. See **Table 135**. Moreover, this assay simultaneously determines the presence of genetic abnormalities that guide treatment selection, thus generating substantial clinical utility in a single test.

Methods

Next-Generation Sequencing (NGS) - DNA

Genomic DNA was isolated from formalin-fixed paraffin-embedded (FFPE) tumor samples which were microdissected to enrich tumor purity. FFPE specimens underwent pathology review to measure percent tumor content and tumor size; a minimum of 20% of tumor content in the area for microdissection was set as a threshold to enable enrichment and extraction of tumor-specific DNA. Matched normal tissue was not routinely sequenced. A custom-designed SureSelect XT assay was used to enrich 592 or whole exome whole-gene targets (Agilent Technologies, Santa Clara, CA). See **Example 1** for further details. Enriched DNA was subjected to NGS using the NextSeq platform (Illumina, Inc., San Diego, CA). All variants were detected with > 99% confidence based on allele frequency and probe panel coverage, with an average sequencing depth of coverage of > 500 and an analytic sensitivity of 5%. Genetic variants identified were interpreted by board-certified molecular geneticists and categorized as 'pathogenic,' 'presumed pathogenic,' 'variant of unknown significance,' 'presumed benign,' or 'benign,' according to the American College of Medical Genetics and Genomics (ACMG) standards. When assessing mutation frequencies of individual genes, 'pathogenic,' 'presumed pathogenic,' and 'variants of unknown significance' were counted as mutations while 'benign' and 'presumed benign' variants were excluded. Copy number alteration (CNA; also commonly referred to as copy number variation (CNV) herein) was simultaneously determined by NGS by comparing the depth of sequencing of genomic loci to a diploid control as well as the known performance of the genomic loci. Calculated gains of 6 copies or greater were considered amplified.

Next-Generation Sequencing (NGS) - RNA

FFPE specimens were microdissected as described above prior to enrichment and extraction of tumor-specific RNA. Qiagen RNA FFPE tissue extraction kit was used for extraction (Qiagen LLC, Germantown, MD), and the RNA quality and quantity were determined using the Agilent TapeStation. Biotinylated RNA baits were hybridized to the synthesized and purified cDNA targets and the bait-target complexes were amplified in a post capture PCR reaction. The Illumina NovaSeq 6500 was used to sequence the whole transcriptome from patients to an average of 60 M reads. Raw data was demultiplexed by Illumina Dragen BioIT accelerator, trimmed, counted, PCR-duplicates removed and aligned to human reference genome hg19 by STAR aligner [14]. For transcription counting, transcripts per million molecules was generated using the Salmon expression pipeline [15].

RNA expression

RNA expression, as defined by transcripts per million (TPM) from the Salmon RNA expression pipeline [15] using our whole transcriptome sequencing assay (WTS; see **Example 1**), was validated using IHC results from over 5000 human breast adenocarcinoma cases. Protein amounts were measured by FDA-approved antibodies using standard quantitative IHC assays. IHC scores come directly from histopathology review by board-certified pathologists for ER/ESR1 (human estrogen receptor), PR/PGR (human progesterone receptor), AR (human androgen receptor), and HER2/neu/ERBB2 (human Herceptin, receptor tyrosine kinase CD340). 50 IHC 'positive' and 50 IHC

'negative' cases were used to decide the TPM thresholds corresponding to IHC positive and IHC negative for these 4 genes. The thresholds were evaluated on 5197 independent cases and all four markers had a sensitivity > 86% with specificities ranging from 85% to 99%. Validation results are shown in **Table 136** and **FIGs. 6A-D**, which show ROC curves for calculating IHC result from WTS expression for the indicated biomarkers.

Table 136 - Results of independent validation of IHC result derivation from WTS expression data

Category	N	Sensitivity	Specificity	PPV	NPV	Accuracy
ER (**FIG. 6A**)	5098	93.5%	90.7%	94.6%	88.8%	92.5%
PR (**FIG. 6B**)	5024	86.3%	85.1%	79.6%	90.3%	85.6%
HER2 (**FIG. 6C**)	5197	91.0%	99.7%	97.8%	98.6%	98.5%
AR (**FIG. 6D**)	5142	88.5%	88.5%	94.4%	77.9%	88.5%

Additionally, we compared data between our WTS expression assay to the Illumina DASL Expression Microarray and publicly available Affymetrix U133A expression arrays from the expO project (Gene Expression Omnibus accession GSE2109) in a cross-platform comparison method [33]. We selected 10 cases from each dataset from a diagnosed Stage IV uterine carcinoma and 10 cases diagnosed with Stage IV colon adenocarcinoma. We identified 14,473 genes which are common across these three platforms. Although these cases are from different people, without being bound by theory, we hypothesized that the gene expression profiles from uterine tumors and colon tumors are sufficiently different from each other and sufficiently common within a tumor type that common patterns of over- and under-expression would be detectable. To visualize this, we took the log2 ratio of the 14,473 genes between uterine (numerator) and colon (denominator) cancer and plotted the ratios. **FIGs. 6E-G** show the ratios plotted against each other with R^2 listed in **FIGs. 6E** (WTS (X axis) and Illumina (Y axis)), **9F** (Illumina (X axis) and Affymetrix (Y axis)) and **9G** (WTS (X axis) and Affymetrix (Y axis)). Note that the expression data was averaged across 10 patients. The Pearson's correlation coefficient for each is 0.68, 0.75 and 0.73 respectively.

Results

Patients

To identify patients for this Example, we used a database of over 200,000 samples analyzed from 2008 to 2020 as described in **Example 1**. We identified 77,044 cases that had next-generation DNA and RNA sequencing results with an available pathology diagnosis including CUP. CUP cases were defined as those assigned a primary tumor site of "Unknown primary site" and for which the "Cancer of Unknown Primary" lineage was selected by the submitting site. The submitted pathological diagnosis was used as the training label. Subsequent independent validation of the classifier was accomplished by including 13,661 cases with a known primary and 1,107 CUP cases that were analyzed prospectively as part of routine tumor profiling. See **FIG. 6H**, which shows a

CONSORT diagram **600** (www.consort-statement.org/consort-statement/flow-diagram). The DNA and RNA components of MI GPSai were trained **603** using a combined 57,489 patients (**601** + **602**), which were then locked **604** and validated on 4,602 non-CUP **605** and 185 CUP patients **606** to determine optimal performance settings. Following this evaluation, MI GPSai rendered a prediction on routinely profiled cases resulting in the final prospective validation set **608** and CUP cases **609**.

Artificial intelligence training

Molecular profiles from 57,489 patients were used for initial training of the global tumor classification algorithm designated MI GPSai. This panomic dataset was comprised of 34,352 cases with genomic data (**FIG. 6H 601**) and 23,137 with both genomic and transcriptomic data (**FIG. 6H 602**). MI GPSai was generated using an artificial intelligence platform that leverages the "Deliberation Analytics" (DEAN) framework as described herein. DEAN uses biomarker data as feature inputs into an ensemble of over 300 well-established machine learning algorithms, including random forest, support vector machine, logistic regression, K-nearest neighbor, artificial neural network, naïve Bayes, quadratic discriminant analysis, and Gaussian processes models. Multiple feature selection methods were employed to build models along with 5-fold cross validation during training to assess performance. High-performing models deliberate against one another to determine a final result. For DNA, a set of 115 distinct primary tumor site and histology classes were defined and used to generate subpopulations of patients. For training the GPS, all 115 disease types were trained against each other using the training set to generate 6,555 model signatures, where each signature is built to differentiate between a pair of disease types. The signatures were generated using Gradient Boosted Forests. The models were validated using the test cases where each test case was processed individually through all 6,555 signatures, thereby providing a pairwise analysis between every disease type for every case. The results are analyzed in a 115 × 115 matrix where each column and each row is a single disease type and the cell at the intersection is the probability that a case is one disease type or the other. The probabilities for each disease type are summed for each column which results in 115 disease types with their probability sums. These disease types are ranked by their probability sums. See **Example 2** and **Tables 2-116** and related discussion for details. For RNA, gradient boosted forests were trained using a selection of RNA transcripts to separately determine a cancer type, organ group and histology. See **FIGs. 4A-B**, and **Tables 117-120** and related discussion for additional details.

The scheme set forth in **FIG. 4B** was used to obtain a final prediction. The 115 × 115 matrix described above is used as an intermediate model to assess DNA **416** and the gradient boosted forests were applied to the transcripts in **Table 117** to build intermediate models to assess cancer type **412**, organ group **413** and histology **414**. A gradient boosted forest was applied to the outputs of the intermediate models to dynamically combine the results **415**. Using this approach, a total of 6,559 models were generated and used to determine a final probability (termed a MI GPS Score) for each case belonging to each of the final desired cancer categories. These MI GPS Scores were then

clustered into multidimensional signatures which were empirically evaluated in our molecular profiling database to determine the predicted prevalence in each cancer category. The prevalence is the final output of the MI GPSai machine learning platform **417**. The desired cancer categories comprised 21 broad cancer categories selected in order to achieve the highest predictive power for a clinically relevant category that would assist with therapy selection in challenging cases. These 21 cancer categories include breast adenocarcinoma; central nervous system cancer; cervical adenocarcinoma; cholangiocarcinoma; colon adenocarcinoma; gastroesophageal adenocarcinoma; gastrointestinal stromal tumor (GIST); hepatocellular carcinoma; lung adenocarcinoma; melanoma; meningioma; ovarian granulosa cell tumor; ovarian, fallopian tube adenocarcinoma; pancreas adenocarcinoma; prostate adenocarcinoma; renal cell carcinoma; squamous cell carcinoma; thyroid cancer; urothelial carcinoma; uterine endometrial adenocarcinoma; and uterine sarcoma.

The top DNA and RNA features that contribute the largest amount of information to the predictions made for each of the 21 cancer categories are shown in **FIGs. 6I-6AC**. In each figure, the 10 leftmost biomarkers are the top contributors based on DNA analysis whereas the 10 rightmost biomarkers are the top contributors based on RNA analysis. In some cases, e.g., GATA3 in breast carcinoma in **FIG. 6I**, the same gene was identified as a top contributor by both DNA and RNA. Without being bound by theory, much of the DNA results are copy number alterations (see, e.g, **Tables 2-116**), and copy number may have a direct impact on transcript levels.

Without being bound by theory, several observations can be made regarding the biomarkers in **FIGs. 6I-6AC**. For example, various canonical driver mutations are found among the top contributing biomarkers. Examples include IDH1 and EGFR for gliomas, cKIT/PDGFRA in gastrointestinal stromal tumors (GIST), BRAF/NRAS in melanoma, KRAS/CDKN2A in pancreatic cancer, GATA3 and CDH1 in breast cancer, VHL in renal cell carcinoma, BRAF in thyroid, PTEN in endometrial cancer, and FOXL2 in ovarian granulosa cell tumors [16], [17], [18], [19], [20], [21]. Expression of genes relatively specific to tissue lineage are also among the top contributors, e.g., CDX2 in gastroesophageal cancer, KIT in GIST, MITF in melanoma and NKX3–1 in prostate cancer [22], [23], [24], [25]. Without being bound by theory, markers in the figures were most useful for differentiating TOO are found in these lists, canonical cancer markers such as BRCA1 are not in the top 10 for the machine learning as they may be found in a number of cancer categories. Additional biomarkers that have not been explicitly associated with the particular cancer types are also included in the algorithm, revealing previously uncovered linkages with biomarkers and pathways. Additional details of the machine learning configurations and inputs are described here [26].

Validation of algorithmic disease classification in independent cohorts

Following the lock of the algorithm (**FIG. 6H 604**), predictions made by the MI GPSai platform were first validated in an independent set of 4,602 patients with known cancer category (**FIG. 6H 605**) and 185 patients with CUP (**FIG. 6H 606**). MI GPSai provided a top prediction for each case along with a score related to the confidence in the call. When evaluating the MI GPSai top

prediction on every case in the cohort irrespective of the score, the top prediction was concordant with the pathologist-assigned disease type in 90.3% of cases. An assessment of the scores in this dataset led us to select a threshold of 0.835 as a minimum score to report a result as it was the intersection of accuracy of the top prediction and the call rate (percentage of cases resulted), resulting in 93.3% accuracy on 93.3% of cases with a defined primary and 75.6% of CUP cases. See **FIG. 6AD**, which shows selection of this threshold in the independent validation set. The x-axis represents all cases with that MI GPSai Score and greater. In the non-CUP cases (N=4,602), the predictor demonstrates a 93.3% sensitivity on 93.3% of cases at the selected threshold of 0.835, annotated as the upper asterisk. In the CUP cases (N=185), 75.6% of cases exceeded the selected threshold, annotated as the lower asterisk. At this threshold, the assay was robust within both primary and metastatic tumors as well as various ranges of tumor purity. See, e.g., **Table 137**.

Table 137. Summary of performance in the independent validation cohort at the selected threshold

Category	n	Call Rate (%)	Sensitivity (%)
Global	4602	93.3	93.3
Primary Specimen	2544	94	94.1
Metastatic Specimen	1969	92.2	92.5
Percent Tumor >= 20, <= 50	2885	92.7	93.4
Percent Tumor > 50, <= 80	1657	94.1	93.1
Percent Tumor > 80	54	100	100

Prospective validation

Subsequently, the assay was used in clinical testing to prospectively evaluate the tumor of each patient with molecular profiling performed (**FIG. 6H 607**). Pathologists were notified of the MI GPSai score and empirical prevalence tables if the assay returned a MI GPSai Score of >= 0.835 for any cancer category. The tumors of 13,661 non-CUP patients were evaluated by the algorithm as a prospective validation cohort. See **Table 138**, wherein sensitivity is abbreviated as "Sens." Globally, this cohort exhibited a similar call rate compared to the initial independent validation cohort (93.0% vs 93.3%) and exhibited a higher sensitivity (94.7% vs 93.3%). The sensitivity of the assay remained above 93% in both primary and metastatic tumors regardless of tumor purity (**Table 138**).

Table 138. Summary of algorithm performance in the prospective validation cohort.

Category	n	Above Threshold	Call Rate (%)	Sens. in Top 1 (%)	Sens. in Top 2 (%)	Sens. in Top 3 (%)	Sens. in Top 4 (%)	Sens. in Top 5 (%)	Rule Outs / Case

Global	13,661	12,699	93	94.7	97.2	97.9	98.1	98.2	17.6
Primary Specimen	7521	7087	94.2	96.1	98.2	98.7	98.8	98.9	17.8
Metastatic Specimen	5942	5426	91.3	93	96	97	97.2	97.4	17.4
Percent Tumor < 20	4	3	75	100	100	100	100	100	18.7
Percent Tumor >= 20, <= 50	8227	7636	92.8	94.5	97	97.8	97.9	98	17.4
Percent Tumor > 50, <= 80	5189	4835	93.2	95	97.7	98.2	98.4	98.5	17.9

This prospective dataset also allowed us to evaluate the diagnostic rule-out power (i.e., negative predictive value) of the assay. For all patients, the empirical prevalence tables yielded an average of 17.6 cancer categories that had not been observed per patient (i.e., could be ruled out) for their respective MI GPSai scores. The correct cancer category had a non-zero empirical probability in 98.9% of all cases, and the 1.1% of observations in which the true cancer category was incorrectly ruled out represents less than 0.1% of the total disease types ruled out. Thus, the rule out accuracy exceeds 99.9%.

Each of the 21 cancer categories was represented in the prospective validation dataset both with respect to true tumor type and highest prediction. See **Table 139**. Sixteen of the 21 cancer categories had an observed positive predictive value (PPV) of >= 90% and three had a PPV of >= 99%. The minimum rule-out accuracy was 98.0%. Five cancer categories (e.g. central nervous system cancers, GIST, melanoma, meningioma, and prostate) each exhibited > 99% sensitivity while twelve (e.g., breast, colon, gastroesophageal, hepatocellular, lung, two subtypes of ovarian, pancreatic, renal, squamous cell, uterine adenocarcinoma, and uterine sarcoma) achieved > 90% sensitivity.

Table 139. Summary of algorithm performance in the prospective validation cohort by cancer category

Category	n	Call Rate (%)	Sensitivity (%)	PPV (%)	Rule Out Accuracy (%)
Breast Adenocarcinoma	1533	98	98.4	99	100

Central Nervous System Cancer	445	99.8	99.8	100	100
Cervical Adenocarcinoma	60	51.7	38.7	66.7	98
Cholangiocarcinoma	363	73.8	69.4	83	99.7
Colon Adenocarcinoma	2119	97	98.5	98.2	100
Gastroesophageal Adenocarcinoma	613	84.5	90.9	89.5	99.9
GIST	23	95.7	100	95.7	100
Hepatocellular Carcinoma	66	84.9	92.9	96.3	99.7
Lung Adenocarcinoma	2287	95	96.4	93.6	100
Melanoma	373	96.5	99.7	99.7	100
Meningioma	21	90.5	100	95	100
Ovarian Granulosa Cell Tumor	25	88	95.5	95.5	100
Ovarian, Fallopian Tube Adenocarcinoma	1493	91.6	92.5	94.3	99.9
Pancreas Adenocarcinoma	815	87.6	91.9	87.7	100
Prostate Adenocarcinoma	556	97.1	99.1	98.7	100
Renal Cell Carcinoma	176	92.6	95.7	96.9	99.8
Squamous Cell Carcinoma	1193	93	93.5	93.4	99.9
Thyroid Cancer	74	85.1	85.7	91.5	99.2
Urothelial Carcinoma	354	90.7	85.4	96.1	99.9
Uterine Endometrial Adenocarcinoma	989	89.4	91.4	89.7	100
Uterine Sarcoma	83	83.1	98.6	94.4	100

FIG. AE and **FIG. AF** show confusion matrices with respect to prediction and truth for the cancer categories, respectively. **FIG. AE** shows a prediction matrix in the prospective validation set. Each row shows the percentage of the actual disease types observed when a MI GPSai achieves a score > 0.835. The diagonal represents the PPV for the given disease type. Blank cells have values between 0 and 1. **FIG. AE** shows a confusion matrix in the prospective validation set. Each column shows observed predictions for each disease type when a MI GPSai achieves a score > 0.835. The diagonal represents the sensitivity for the given disease type. Blank cells have values between 0 and 1.

Analysis of CUP

Of the 1292 CUP cases analyzed by MI GPSai, 71.7% achieved a score exceeding the reportable threshold. See **FIG. 6AG**, which shows the distribution of MI GPSai predictions in CUP cases. The top panel in the figure shows the score distributions, where 71.7% of cases return a reportable result, and the bottom panel represents the predictions that were made. Validation of a CUP assay at the individual patient level is fundamentally uncertain as the "truth" is unknown. As such, comparing the populations generated by MI GPSai for each cancer category in terms of mutation frequencies against the mutation frequencies in populations of known primaries yields insight into the similarities of these populations. The genes with mutation frequencies with a 95% confidence interval which does not overlap with that of any other cancer category along with their frequencies in the populations created by MI GPSai can be seen in **Table 140**. In the table, "*" represents the observed value among the known cancer category of the combined training and testing datasets, and "**" represents the observed value among CUP cases predicted to be of the cancer category in each row. Many of the pathogenic mutation frequencies were similar in the labeled and CUP predicted populations, but not all. In particular, VHL pathogenic mutations were not seen in the 18 CUP cases classified as Renal Cell Carcinoma. This could potentially be due to lower proportions of clear cell carcinoma in CUP [27].

Table 140 - Percentages of pathogenic variants detected among biomarkers per cancer category

	Of This Cancer Category		**Not Of This Cancer Category**	
Biomarker	**Train + Test***	**CUP****	**Train + Test**	**CUP****
Breast Adenocarcinoma				
CDH1	10.7% (9.7-11.7)	11.1% (3.4-18.6)	0.8% (0.7-0.9)	0.8% (0.2-1.4)
ESR1	9.2% (8.2-10.1)	0.0% (0.0-0.0)	0.2% (0.2-0.3)	0.1% (0.0-0.4)
GATA3	9.5% (8.6-10.5)	1.8% (0.0-5.1)	0.1% (0.1-0.1)	0.0% (0.0-0.0)
MAP3K1	5.2% (4.5-5.9)	2.6% (0.0-6.8)	0.8% (0.7-0.9)	0.3% (0.0-0.7)
Cholangiocarcinoma				
IDH1	8.6% (7.0-10.4)	19.5% (13.2-25.7)	0.4% (0.3-0.4)	0.4% (0.0-0.9)
Colon Adenocarcinoma				

AMER1	6.5% (5.9-7.1)	4.7% (1.2-9.3)	0.4% (0.3-0.4)	0.6% (0.1-1.2)
APC	76.3% (75.3-77.3)	34.1% (24.4-44.2)	2.4% (2.2-2.6)	2.5% (1.5-3.6)
Lung Adenocarcinoma				
EGFR	14.7% (13.8-15.6)	1.5% (0.4-3.2)	0.3% (0.2-0.3)	0.5% (0.0-1.1)
KEAP1	9.3% (8.7-10.0)	20.2% (15.8-25.1)	0.9% (0.8-1.0)	1.2% (0.3-2.2)
SMARCA4	5.8% (5.3-6.4)	19.9% (15.1-24.4)	1.3% (1.2-1.5)	2.4% (1.3-3.6)
STK11	14.4% (13.5-15.2)	26.9% (21.5-31.9)	0.9% (0.8-1.0)	1.3% (0.5-2.2)
Ovarian, Fallopian Tube Adenocarcinoma				
BRCA1	8.8% (7.9-9.7)	4.8% (0.0-11.6)	1.3% (1.2-1.4)	1.4% (0.7-2.2)
TP53	81.9% (80.6-83.1)	90.5% (81.4-97.7)	61.9% (61.4-62.5)	51.8% (48.2-55.2)
Pancreas Adenocarcinoma				
CDKN2A	24.2% (22.3-26.3)	18.1% (10.0-27.2)	4.8% (4.5-5.0)	7.8% (6.1-9.8)
KRAS	88.9% (87.5-90.3)	94.2% (88.6-98.6)	19.0% (18.6-19.4)	18.1% (15.4-20.8)
SMAD4	18.1% (16.4-19.8)	25.6% (15.7-37.1)	4.0% (3.8-4.2)	3.5% (2.3-4.9)
Renal Cell Carcinoma				
KDM5C	17.7% (13.1-22.4)	0.0% (0.0-0.0)	1.2% (1.1-1.4)	1.5% (0.6-2.6)
PBRM1	35.1% (31.1-39.3)	21.4% (5.6-39.0)	1.3% (1.2-1.4)	3.8% (2.5-5.2)
SETD2	25.5% (21.5-29.1)	33.1% (11.1-55.6)	1.4% (1.3-1.5)	1.7% (0.8-2.6)
VHL	59.7% (55.4-64.1)	0.0% (0.0-0.0)	0.0% (0.0-0.1)	0.1% (0.0-0.3)
Squamous Cell Carcinoma				
NFE2L2	7.6% (6.7-8.4)	6.9% (2.5-11.9)	0.6% (0.5-0.7)	0.4% (0.0-0.9)
NOTCH1	7.2% (6.3-8.0)	6.8% (2.5-11.9)	0.8% (0.7-0.9)	1.3% (0.6-2.2)
Urothelial Carcinoma				
CREBBP	6.9% (5.4-8.4)	12.5% (0.0-29.4)	1.5% (1.4-1.7)	2.3% (1.4-3.4)
EP300	5.8% (4.4-7.2)	6.6% (0.0-17.6)	1.2% (1.1-1.3)	1.5% (0.8-2.3)
ERBB2 (Her2/Neu)	7.8% (6.2-9.3)	6.4% (0.0-17.6)	1.5% (1.3-1.6)	2.4% (1.5-3.5)
FGFR3	14.6% (12.5-16.8)	6.5% (0.0-17.6)	0.2% (0.2-0.3)	0.6% (0.1-1.1)
KDM6A	21.9% (19.5-24.5)	13.2% (0.0-35.3)	1.3% (1.2-1.5)	2.4% (1.4-3.4)
KMT2D	26.9% (24.3-29.8)	14.5% (0.0-29.6)	5.3% (5.0-5.5)	6.5% (4.9-8.3)
TSC1	9.2% (7.6-10.9)	0.0% (0.0-0.0)	0.7% (0.6-0.8)	0.9% (0.3-1.6)

Uterine Endometrial Adenocarcinoma				
ARID1A	82.4% (80.2-84.6)	100.0% (100.0-100.0)	27.8% (26.9-28.8)	25.1% (20.1-30.2)
ASXL1	22.6% (19.3-26.1)	20.0% (5.3-36.8)	6.9% (6.4-7.4)	5.9% (2.9-9.2)
BCOR	8.5% (7.5-9.6)	17.0% (0.0-36.8)	0.9% (0.8-1.0)	1.2% (0.6-1.9)
FBXW7	13.7% (12.5-15.0)	21.4% (5.3-42.1)	3.7% (3.5-3.9)	2.5% (1.5-3.6)
FGFR2	5.9% (5.1-6.8)	11.0% (0.0-26.3)	0.4% (0.3-0.4)	1.4% (0.7-2.3)
JAK1	10.4% (9.3-11.5)	22.5% (5.3-42.1)	0.7% (0.7-0.8)	0.4% (0.0-0.8)
MSH6	5.2% (4.5-6.0)	10.8% (0.0-26.3)	1.1% (1.0-1.2)	1.5% (0.8-2.3)
MSI	20.1% (18.7-21.7)	28.2% (10.5-47.4)	2.2% (2.0-2.4)	2.6% (1.7-3.7)
PIK3CA	39.3% (37.5-41.1)	52.8% (31.6-73.7)	12.2% (11.9-12.6)	6.0% (4.5-7.5)
PIK3R1	21.7% (20.1-23.2)	22.4% (5.3-42.1)	1.5% (1.4-1.6)	0.9% (0.3-1.6)
PPP2R1A	11.7% (10.6-12.9)	11.2% (0.0-26.3)	0.4% (0.3-0.5)	0.2% (0.0-0.6)
PTCH1	6.7% (5.5-8.1)	18.2% (5.3-36.8)	1.3% (1.1-1.5)	2.2% (1.1-3.4)
PTEN	42.9% (41.0-44.8)	49.9% (26.3-73.7)	4.5% (4.2-4.7)	3.7% (2.6-5.0)
RNF43	7.8% (6.8-8.8)	15.7% (0.0-31.6)	1.9% (1.8-2.1)	1.1% (0.5-1.8)

Clinical utility and case examples

In a non-limiting real world example, we received an inguinal lymph node biopsy on an 82-year-old man which was sent for molecular profiling (see **Example 1**). At the time of biopsy, the serum PSA was not elevated, and workup had not identified the primary tumor. Evaluation by the referring pathologist included negative IHC stains with CK7, CK20, PSA, PSAP, CDX2, p40, GATA3, SOX10, and CD45. A cytokeratin stain was positive (AE1/3) and case was diagnosed as carcinoma of unknown primary. Notably, this carcinoma was evaluated appropriately for prostatic lineage with PSA and PSAP IHC, and given the concurrent low serum PSA, prostatic adenocarcinoma was considered ruled out.

MI GPSai predicted with high probability that the sample was prostate adenocarcinoma (MI GPSai score 0.9998) and review of the gene expression data showed high expression of androgen receptor (AR). IHC of AR protein was performed and AR was found highly expressed, which supported the MI GPSai call. The patient had a follow-up biopsy of the prostate which confirmed prostatic adenocarcinoma. After discussion with the ordering physician, the diagnosis was changed from CUP to metastatic prostatic adenocarcinoma. Importantly, the patient's molecular profiling also identified pathogenic variants in BRCA2 and PTEN, highlighting the utility of diagnosis and biomarker analysis from the same platform.

In addition to assigning lineage and identifying biomarker data with CUP cases, MI GPSai can assist with pathologic diagnosis fidelity. We prospectively monitored discrepancies between MI GPSai and the pathologist-assigned diagnoses in 1292 cases. In cases where the pathologist-assigned diagnosis was different than the top MI GPSai prediction and the MI GPSai score for the top prediction exceeded 0.999, an automated email was sent to the pathologist in charge of the case alerting them to this discrepancy. The pathology group was previously educated on the design and performance of MI GPSai and instructed to consider the discrepant cases with their medical judgement. The pathologists were able to review patient clinical history, imaging results if available, order immunohistochemistry, and discuss the case with the referring oncologist and/or pathologist.

There were 46 cases with a MI GPSai score greater than 0.999 where pathologists were alerted. After review with additional immunohistochemistry and consultation with the referring physician, the diagnosis was changed in 19 cases (41.3%). In 11 cases (23.9%), where the submitted diagnosis was not changed despite MI GPSai predictions, the predicted diagnosis was pancreatic adenocarcinoma, a cancer with limited specific IHC markers for confirmation. All cases did not result in a diagnosis revision for various reasons ranging from a lack of diagnostic IHCs to verify the prediction (such as cholangiocarcinoma vs pancreatic carcinoma) to a lack of response from the oncologist.

In one non-limiting real world example, the patient's treatment course was altered based on MI GPSai. See **FIGs. 6AH-AL**. We received a cervical lymph node from a 61-year-old man for molecular profiling. The referring pathologist assigned a diagnosis of poorly-differentiated squamous cell carcinoma (**FIG. 6AH**). The patient had systemic metastasis and had not responded well to squamous cell carcinoma directed therapy. The MI GPSai predicted diagnosis was urothelial carcinoma (MI GPSai score 0.9999). Our whole transcriptome expression data was used to select for lineage specific gene expression to guide immunohistochemical antibody selection, the current gold-standard for lineage assignment. The mean RNA expression of Uroplakin II and GATA3 of the urothelial carcinoma cases in our database is relatively high based on WTS data across numerous cancers, both relatively specific for urothelial carcinoma and not typically expressed in squamous cell carcinoma. See **FIGs. 6AI** and **9AJ**, respectively. Thus the patient sample was probed with antibodies to these proteins. This additional IHC was positive for Uroplakin II and GATA3. See **FIGs. 6AK** and **9AL**, respectively. Importantly, the choice of the PD-L1 clone and scoring system was affected by the lineage of cancer being tested. In this case, the referring pathologist and oncologist asked to change the diagnosis to urothelial carcinoma and run the SP142 PD-L1 antibody according to the label indications for atezolizumab. This PD-L1 score was positive and the patient therapy was changed. These non-limiting real world patient examples show that MI GPSai has significant clinical utility with both CUP and diagnostic fidelity.

Discussion

Cancer of unknown primary remains a major clinical challenge and outcomes are poor. Molecular predictors of tumor origin can assist in addressing this problem by providing critical information in CUP cases that can inform treatment decisions and potentially improve outcomes. Herein we provide an artificial intelligence-derived panomic molecular classifier that uses DNA and RNA information to make tumor type predictions across a broad spectrum of diagnostic classes with high accuracy.

Prior molecular assays for the identification of cancers of unknown primary have focused on RNA profiles which have degraded performance in situations where the tumor is from a site of metastasis or if the tumor percentage is low [7]. Our method is robust to these limitations. Without being bound by theory, this is at least in part because we isolate nucleic acid from microdissected material, thus enriching for tumor cells, and because we use combined analysis of DNA and RNA, which further reduces susceptibility to the effects of normal cell contamination. As demonstrated in the case examples above, availability of mutational and gene expression analysis data further enhances the clinical utility of our approach from a diagnostic and therapeutic perspective.

The accuracy of MI GPSai surpasses recently reported uses of DNA NGS panels for tissue of origin identification or guidance of utilization of targeted- and immunotherapies [10], [28]. Moreover, overall accuracy of these approaches may be limited. For example, predictions made by a Random Forest Classifier using results from a 468-gene NGS panel as input, resulted in an overall accuracy of 74.1% [10]. Analysis of circulating tumor DNA data from a commercial 70-gene NGS panel revealed potentially targetable mutations. However, an attempt to identify the underlying TOO was not made [28], possibly due to the limited number of genes analyzed. In contrast, analysis of DNA methylation across the genome might add additional information to above-mentioned assays, as it has been shown to predict a primary tumor in 87% of CUP cases [29].

In addition to its role in understanding CUP, MI GPSai provides a quality control tool that can be integrated into a pathology laboratory workflow. As part of our prospective evaluation of MI GPSai, pathologists were alerted to discrepancies between submitted diagnosis and MI GPSai prediction, resulting in change in diagnosis in 41.3% of these cases. Considering that the rate of inaccurate diagnosis ranges between 3% and 9% [30], inclusion of MI GPSai in clinical routine could improve diagnostic fidelity overall.

In summary, MI GPSai displayed robust performance in the diagnostic workup of CUP cases that was consistent across 13,661 cases including both metastatic and low percentage tumors. At the same time, MI GPSai can also play an important role in quality control of anatomical pathology laboratories. Since the MI GPSai analysis uses the results of DNA and RNA profiles obtained as part of routine clinical tumor profiling, both diagnostic and therapeutic information can be returned that optimize patients' treatment strategy from a single test. This workflow improves the current standard of multiple tests that require more tissue and increased turnaround time, which can delay treatment.

Our approach aims to utilize the context-specific information gained by lineage assignment when considering biomarker-directed therapy.

References (*bracketed numbers [#] correspond to those in the text of this Example*)

[1] C. Massard, et al. Carcinomas of an unknown primary origin–diagnosis and treatment. Nat. Rev. Clin. Oncol., 8 (12) (2011), pp. 701-710

[2] G.R. Varadhachary, M.N. Raber. Cancer of unknown primary site. N. Engl. J. Med., 371 (8) (2014), pp. 757-765

[3] B.R. DeYoung, M.R. Wick. Immunohistologic evaluation of metastatic carcinomas of unknown origin: an algorithmic approach. Semin. Diagn. Pathol., 17 (3) (2000), pp. 184-193

[4] G.G. Anderson, L.M. Weiss. Determining tissue of origin for metastatic cancers: meta-analysis and literature review of immunohistochemistry performance. Appl. Immunohistochem. Mol. Morphol., 18 (1) (2010), pp. 3-8

[5] S.Y. Park, et al. Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. Arch. Pathol. Lab. Med., 131 (10) (2007), pp. 1561-1567

[6] R.W. Brown, et al. Immunohistochemical identification of tumor markers in metastatic adenocarcinoma. A diagnostic adjunct in the determination of primary site. Am. J. Clin. Pathol., 107 (1) (1997), pp. 12-19

[7] M.G. Erlander, et al. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. J. Mol. Diagn., 13 (5) (2011), pp. 493-503

[8] J.D. Hainsworth, et al. Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. J. Clin. Oncol., 31 (2) (2013), pp. 217-223

[9] J.S. Ross, et al. Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. JAMA Oncol., 1 (1) (2015), pp. 40-49

[10] A. Penson, et al. Development of genome-derived tumor type prediction to inform clinical cancer care. JAMA Oncol., 6 (1) (2019), pp. 84-91

[11] G.A. Stancel, et al. Identification of tissue of origin in body fluid specimens using a gene expression microarray assay. Cancer Cytopathol., 120 (1) (2012), pp. 62-70

[12] J.L. Dennis, et al. Markers of adenocarcinoma characteristic of the site of origin: development of a diagnostic algorithm. Clin. Cancer Res., 11 (10) (2005), pp. 3766-3772

[13] A.R. Gamble, et al. Use of tumour marker immunoreactivity to identify primary site of metastatic cancer. BMJ, 306 (6873) (1993), pp. 295-298

[14] A. Dobin, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29 (1) (2013), pp. 15-21

[15] R. Patro, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods, 14 (4) (2017), pp. 417-419

[16] C.W. Brennan, et al. The somatic genomic landscape of glioblastoma. Cell, 155 (2) (2013), pp. 462-477

[17] S.P. Shah, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. N. Engl. J. Med., 360 (26) (2009), pp. 2719-2729

[18] ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature, 578 (7793) (2020), pp. 82-93

[19] F. Sanchez-Vega, et al. Oncogenic signaling pathways in the cancer genome atlas. Cell, 173 (2) (2018), pp. 321-337.e10

[20] M.C. Heinrich, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. J. Clin. Oncol., 21 (23) (2003), pp. 4342-4349

[21] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature, 490 (7418) (2012), pp. 61-70

[22] P. Tan, et al. Genetics and molecular pathogenesis of gastric adenocarcinoma. Gastroenterology, 149 (5) (2015), pp. 1153-1162

[23] M. Miettinen, et al. Immunohistochemical spectrum of GISTs at different sites and their differential diagnosis with a reference to CD117 (KIT). Mod. Pathol., 13 (10) (2000), pp. 1134-1142

[24] L.A. Garraway, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature, 436 (7047) (2005), pp. 117-122

[25] M.C. Markowski, et al. Inflammatory cytokines induce phosphorylation and ubiquitination of prostate suppressor protein NKX3.1. Cancer Res., 68 (17) (2008), pp. 6896-6901

[26] Abraham J., et al. Genomic Profiling Similarity. WO2020146554.

[27] F.A. Greco, J.D. Hainsworth. Renal cell carcinoma presenting as carcinoma of unknown primary site: recognition of a treatable patient subset. Clin. Genitourin. Cancer, 16 (4) (2018), pp. e893-e898

[28] S. Kato, et al. Utility of genomic analysis in circulating tumor DNA from patients with carcinoma of unknown primary. Cancer Res., 77 (16) (2017), pp. 4238-4246

[29] S. Moran, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol., 17 (10) (2016), pp. 1386-1395

[30] M. Peck, et al. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. J. Clin. Pathol., 71 (11) (2018), pp. 995-1000

[31] K. Bera, et al. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat. Rev. Clin. Oncol., 16 (11) (2019), pp. 703-715

[32] W. Jiao, G. Atwal, P. Polak, et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nat. Commun., 11 (2020), p. 728

[33] P. Stafford, M. Brun. Three methods for optimization of cross-laboratory and cross-platform microarray expression data. Nucl. Acids Res., 35 (10) (2007), p. e72

[34] Haskell CM, et al. Metastasis of unknown origin. Curr Probl Cancer. 1988 Jan-Feb;12(1):5-58. Review. PubMed PMID: 3067982.

[35] Haigis KM, et al. Tissue-specificity in cancer: The rule, not the exception. Science. 2019 Mar 15;363(6432):1150-1151. doi: 10.1126/science.aaw3472. PubMed PMID: 30872507.

Example 4: Molecular Profiling Report and Use for Patient with Metastatic Adenocarcinoma

FIGs. 7A-P present a molecular profiling report which is de-identified but from molecular profiling of a real life patient according to the systems and methods provided herein.

FIG. 7A illustrates page 1 of the report indicating the specimen as reported in the test requisition from the ordering physician was taken from the liver and was presented with primary tumor site as ascending colon. The diagnosis was metastatic adenocarcinoma. In the "Results with Therapy Associations" section, **FIG. 7A** further displays a summary of therapies associated with potential benefit and therapies associated with potential lack of benefit based on the relevant biomarkers for the therapeutic associations. Here, the report notes that mutations were not detected in KRAS, NRAS and BRAF, thereby indicated potential benefit of cetuximab or panitumumab. Conversely, lack of expression of HER2 protein indicates potential lack of benefit from anti-HER2 therapies (lapatinib, pertuzumab, trastuzamab). The section "Cancer Type Relevant Biomarkers" highlights certain of the molecular profiling results for particularly relevant biomarkers. The "Genomic Signatures" section indicates the results of microsatellite instability (MSI) and tumor mutational burden (TMB). Note both characteristics were also highlighted in the section just above. This patient was found to be MSI stable and TMB low.

FIG. 7B is page 2 of the report and lists a summary of biomarker results from the indicated assays. Of note, APC and TP53 were found to have known pathogenic mutations via sequencing of tumor genomic DNA. The section "Other Findings" notes a number of genes with indeterminate sequencing results due to low coverage.

FIG. 7C is page 3 of the report and continues the list of "Other Findings" with genes where genomic DNA sequencing (by NGS) did not find point mutations, indels, or copy number amplification.

FIG. 7D is page 4 of the report and further continues the list of "Other Findings" with genes where RNA sequencing (by NGS) did not find alterations (e.g., no fusion genes detected).

FIG. 7E is page 5 of the report and shows the results of the Genomic Profiling Similarity (GPS) analysis as provided herein performed on the specimen. Recall the specimen comprises a metastatic lesion taken from the liver and was reported to be an adenocarcinoma of the ascending colon by the ordering physician (see **FIG. 7A**). As shown in the figure, the report provides a probability that the specimen is from each of the listed organ groups (i.e., Bladder; Brain; Breast; Colon; Female Genital Tract & Peritoneum; Gastroesophageal; Head, Face or Neck, NOS; Kidney;

Liver, Gall Bladder, Ducts; Lung; Melanoma/Skin; Pancreas; Prostate; Other). The Similarity for each Organ type shown is in the vertical bars. In this case, GPS assigned a score of 97 to Organ type "Colon," and the starred shape indicates a probability of correct match > 98%. See "Legend" box. The Organ group Gastroesophageal had a similarity of 1, and the circular shape indicates that the probability is inconclusive. All other organs had a similarity of less than 1 or 0, indicating that those Organ groups were excluded with a > 99% probability.

FIG. 7F is page 6 of the report and provides a listing of "Notes of Significance," here an available clinical trial based on the profiling results, and additional specimen information.

FIG. 7G is page 7 of the report and provides a "Clinical Trial Connector," which identifies potential clinical trials for the patient based on the molecular profiling results. A trial connected to the APC gene mutation (see **FIG. 7B**) is noted.

FIG. 7H presents a disclaimer. For example, that decisions on patient care and treatment must be based on the independent medical judgment of the treating physician, taking into consideration all available information concerning the patient's condition. This page ends the main body of the report and an Appendix follows.

FIGs. 7I-M provide more details about results obtained using Next-Generation Sequencing (NGS). **FIG. 7I** is page 1 of the appendix and provides information about the Tumor Mutational Burden (TMB) and Microsatellite Instability (MSI) analyses and results. The report notes that high mutational load is a potential indicator of immunotherapy response (Le et al., PD-1 Blockade in Tumors with Mismatch-Repair Deficiency, N Engl J Med 2015; 372:2509-2520; Rizvi et al., Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. Science. 2015 Apr 3; 348(6230): 124–128; Rosenberg et al., Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single arm, phase 2 trial. Lancet. 2016 May 7; 387(10031): 1909–1920; Snyder et al., Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. N Engl J Med. 2014 Dec 4; 371(23): 2189–2199; all of which references are incorporated by reference herein in their entirety). **FIG. 7J** is page 2 of the appendix and lists details concerning the genes found to harbor alterations, namely APC and TP53. See also **FIG. 7B**. **FIG. 7K** is page 3 of the appendix and notes genes that were tested by NGS with either indeterminate results due to low coverage for some or all exons, or no detected mutations. **FIG. 7L** is page 4 of the appendix and continues the listing of genes that were tested by NGS with no detected mutations and adds more information about how Next Generation Sequencing was performed. **FIG. 7M** is page 5 of the appendix and provides information about copy number alterations (CNA; copy number variation; CNV), e.g., gene amplification, detected by NGS analysis and corresponding methodology. **FIG. 7N** is page 6 of the appendix and provides information about gene fusion and transcript variant detection by RNA Sequencing analysis and corresponding methodology. In this specimen, no fusions or variant transcripts were detected. **FIG. 7O** is page 7 of the appendix and provides more information about the IHC analysis performed on the

patient specimen, e.g., the staining threshold and results for each marker. **FIG. 7P** and **FIG. 7Q** are pages 8 and 9 of the appendix, respectively, and provide a listing of references used to provide evidence of the biomarker – agent association rules used to construct the therapy recommendations.

Example 5: Molecular Profiling Report – Metastatic Ovarian Carcinoma

FIGs. 8A-P present another molecular profiling report which is de-identified but from molecular profiling of a real life patient according to the systems and methods provided herein.

FIG. 8A illustrates page 1 of the report indicating the specimen as reported in the test requisition from the ordering physician was taken from the ascending colon and was presented with primary tumor site as ovary. The diagnosis was carcinoma, NOS. In the "Results with Therapy Associations" section, **FIG. 8A** further displays a summary of therapies associated with potential benefit and therapies associated with potential lack of benefit based on the relevant biomarkers for the therapeutic associations. Here, the report notes that the sample was identified as PD-L1 positive by IHC, thereby indicated potential benefit of pembrolizamab. Conversely, lack of expression of HER2 protein indicates potential lack of benefit from anti-HER2 therapies pertuzumab or trastuzamab. The section "Cancer Type Relevant Biomarkers" highlights certain of the molecular profiling results for particularly relevant biomarkers, including results from various analytes: genomic DNA (microsatellite instability (MSI), mismatch repair status, tumor mutational burden (TMB), and ATM and BRCA1/2 status); whole transcriptome sequencing (NTRK1/2/3 fusion); and IHC (ER/PR protein status). The sample was found to be MSI stable, MMR proficient, TMB low, no NTRK fusions detected, no mutation detected in ATM or BRCA1/2, and ER/PR negative. The section "Other Findings" notes that a pathogenic variant was found in the TP53 gene by NGS of genomic DNA.

FIG. 8B is page 2 of the report and lists additional summary of biomarker results from the indicated assays. "Genomic Signatures" provides additional insight into the MSI and TMB results. "Genes Tested with Pathogenic or Likely Pathogenic Alterations" provides further detail about the TP53 pathogenic mutation detected via sequencing of tumor genomic DNA. The section "Immunohistochemistry Results" provides further detail about the protein expression results, e.g., criteria used to determine the result, and details results of the MMR genes (MLH1, MSH2, MSH6, PMS2). "Genes Tested with Indeterminate Results by Tumor DNA Sequencing" notes certain genes of interest with indeterminate results due to low sequencing coverage of some or all exons.

FIG. 8C is page 3 of the report and shows the results of the MI GPSai (GPS) analysis as provided herein performed on the specimen. See, e.g., **Example 3**. Recall the specimen comprises a metastatic lesion taken from the ascending colon and was reported to be an ovarian carcinoma by the ordering physician (see **FIG. 8A**). As shown in **FIG. 8C**, the report provides a probability that the specimen is from each of the listed cancer categories (i.e., breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma,

lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma). The predicted Prevalence for each cancer category is shown is in the horizontal bars. In this case, GPS assigned a prevalence of 96% to cancer category "Ovarian, Fallopian Tube Adenocarcinoma." The cancer category "Uterine Endometrial Adenocarcinoma" had a prevalence of 3%, and "Cervical Adenocarcinoma" had a prevalence of <1%. All other categories had a prevalence of ~0%. Thus, the GPS result was consistent with the original diagnosis.

FIG. 8D is page 4 of the report and provides a listing of "Notes of Significance," here an available clinical trial based on the profiling results, and additional specimen information.

FIG. 8E is page 5 of the report and provides a "Clinical Trial Connector," which identifies potential clinical trials for the patient based on the molecular profiling results. A trial connected to the PD-L1 IHC result (see **FIG. 8A**) is noted.

FIG. 8F is page 6 of the report and presents a disclaimer. For example, that decisions on patient care and treatment must be based on the independent medical judgment of the treating physician, taking into consideration all available information concerning the patient's condition. This page ends the main body of the report and an Appendix follows.

FIGs. 8G-I are pages 7-9 of the report (and 1-3 of the Appendix) and provide more details about results obtained using Next-Generation Sequencing (NGS) of genomic tumor DNA. **FIG. 8G** is page 1 of the appendix and provides information about the Tumor Mutational Burden (TMB) and Microsatellite Instability (MSI) analyses and results, and provides details concerning mutations in genes found to harbor alterations, here TP53. **FIG. 8H** is page 2 of the appendix and notes genes that were tested by NGS with either indeterminate results due to low coverage for some or all exons and provides details about the NGS assay. **FIG. 8I** is page 3 of the appendix and provides information about copy number alterations (CNA; copy number variation; CNV), e.g., gene amplification, detected by NGS analysis and corresponding methodology. **FIG. 8J** is page 4 of the appendix and provides information about gene fusion and transcript variant detection by RNA Sequencing analysis and corresponding methodology. In this specimen, no fusions or variant transcripts were detected. **FIGs. 8K-L** are pages 5-6 of the appendix, respectively, and provides more information about the IHC analysis performed on the patient specimen, e.g., the staining threshold and results for each marker. **FIG. 8M** is page 7 of the appendix, and provide a listing of references used to provide evidence of the biomarker – agent association rules used to construct the therapy recommendations.

Example 6: Selecting Treatment for a Cancer

An oncologist is treating a cancer patient with a metastatic tumor of unknown primary and desires to perform molecular profiling on the tumor sample to assist in selecting a treatment regimen for the patient. A biological sample is collected from a tumor located in the retroperitoneum. The

oncologist's pathology report states that the specimen is adenocarcinoma, NOS with unknown primary origin, i.e., CUP. The oncologist requisitions a molecular profiling panel to be performed on the tumor sample. The sample is sent to our laboratory for molecular profiling according to **Example 1** herein.

We perform molecular profiling comprising NGS of genomic DNA, NGS of RNA transcripts, and IHC analysis on the tumor specimen. A molecular profile is generated for the sample. The machine learning models described in **Examples 2-3** are used to predict the primary site of the tumor. The classification leans strongly towards "ovarian, fallopian, retroperitoneal adenocarcinoma." Mutations in APC and TP53 are identified. No mutations in KRAS, BRAF, and NRAS are found. HER2 is not overexpressed. The molecular profiling results are included in the report such as in the Examples above. The report suggests treatment with cetuximab or panitumumab but not anti-HER2 therapy. The report is provided to the oncologist. The oncologist uses the information provided in the report to assist in determining a treatment regimen for the patient.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope as described herein, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A data processing apparatus for generating input data structure for use in training a machine learning model to predict at least one attribute of a biological sample, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the data processing apparatus including one or more processors and one or more storage devices storing instructions that when executed by the one or more processors cause the one or more processors to perform operations, the operations comprising:

obtaining, by the data processing apparatus one or more biomarker data structures and one or more sample data structures;

extracting, by the data processing apparatus, first data representing one or more biomarkers associated with the sample from the one or more biomarker data structures, second data representing the sample data from the one or more sample data structures, and third data representing a predicted at least one attribute;

generating, by the data processing apparatus, a data structure, for input to a machine learning model, based on the first data representing the one or more biomarkers and the second data representing the predicted at least one attribute and sample;

providing, by the data processing apparatus, the generated data structure as an input to the machine learning model;

obtaining, by the data processing apparatus, an output generated by the machine learning model based on the machine learning model's processing of the generated data structure;

determining, by the data processing apparatus, a difference between the third data representing a predicted predicted at least one attribute for the sample and the output generated by the machine learning model; and

adjusting, by the data processing apparatus, one or more parameters of the machine learning model based on the difference between the third data representing a predicted predicted at least one attribute for the sample and the output generated by the machine learning model.

2. The data processing apparatus of claim **1**, wherein the set of one or more biomarkers include one or more biomarkers listed in any one of **Tables 121-129**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof.

3. The data processing apparatus of claim **1**, wherein the set of one or more biomarkers include each of the biomarkers in claim **2**.

4. The data processing apparatus of claim **1**, wherein the set of one or more biomarkers includes at least one of the biomarkers in claim **2**, optionally wherein the set of one or more

biomarkers comprises each of the biomarkers in **Table 118**, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**.

5. A data processing apparatus for generating input data structure for use in training a machine learning model to predict at least one attribute of a biological sample, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the data processing apparatus including one or more processors and one or more storage devices storing instructions that when executed by the one or more processors cause the one or more processors to perform operations, the operations comprising:

obtaining, by the data processing apparatus, a first data structure that structures data representing a set of one or more biomarkers associated with a biological sample from a first distributed data source, wherein the first data structure includes a key value that identifies the sample;

storing, by the data processing apparatus, the first data structure in one or more memory devices;

obtaining, by the data processing apparatus, a second data structure that structures data representing data for the at least one attribute for the sample having the one or more biomarkers from a second distributed data source, wherein the data for the at least one attribute includes data identifying a sample, at least one attribute, and an indication of the predicted at least one attribute, wherein second data structure also includes a key value that identifies the sample;

storing, by the data processing apparatus, the second data structure in the one or more memory devices;

generating, by the data processing apparatus and using the first data structure and the second data structure stored in the memory devices, a labeled training data structure that includes (i) data representing the set of one or more biomarkers and the sample, and (ii) a label that provides an indication of a predicted at least one attribute, wherein generating, by the data processing apparatus and using the first data structure and the second data structure includes correlating, by the data processing apparatus, the first data structure that structures the data representing the set of one or more biomarkers associated with the sample with the second data structure representing predicted at least one attribute data for the sample having the one or more biomarkers based on the key value that identifies the subject; and

training, by the data processing apparatus, a machine learning model using the generated label training data structure, wherein training the machine learning model using the generated labeled training data structure includes providing, by the data processing apparatus and to the machine learning model, the generated label training data structure as an input to the machine learning model.

6. The data processing apparatus of claim **5**, wherein operations further comprising:

obtaining, by the data processing apparatus and from the machine learning model, an output generated by the machine learning model based on the machine learning model's processing of the generated labeled training data structure; and

determining, by the data processing apparatus, a difference between the output generated by the machine learning model and the label that provides an indication of the predicted at least one attribute.

7. The data processing apparatus of claim **6**, the operations further comprising:

adjusting, by the data processing apparatus, one or more parameters of the machine learning model based on the determined difference between the output generated by the machine learning model and the label that provides an indication of the predicted at least one attribute.

8. The data processing apparatus of claim **5**, wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Tables 121-127**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof.

9. The data processing apparatus of claim **5**, wherein the set of one or more biomarkers include each of the biomarkers in **Table 118**, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**.

10. The data processing apparatus of claim **5**, wherein the set of one or more biomarkers includes at least one of the biomarkers in claim **8**.

11. A method comprising steps that correspond to each of the operations of claims **1-10**.

12. A system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described with reference to any one of claims **1-10**.

13. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described with reference to any one of claims **1-10**.

14. A method for determining at least one attribute of a biological sample, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the method comprising:

for each particular machine learning model of a plurality of machine learning models that have each been trained to perform an prediction operation between received input data representing a sample and the at least one attribute:

providing, to the particular machine learning model, input data representing a sample of a subject, wherein the sample was obtained from tissue or an organ of the subject; and

obtaining output data, generated by the particular machine learning model based on the particular machine learning model's processing the provided input data, that represents a probability or likelihood that the sample represented by the provided input data corresponds to the at least one attribute;

providing, to a voting unit, the output data obtained for each of the plurality of machine learning models, wherein the provided output data includes data representing initial sample attributes determined by each of the plurality of machine learning models; and

determining, by the voting unit and based on the provided output data, the predicted at least one attribute.

15. The method of claim **14**, wherein the predicted at least one attribute is determined by applying a majority rule to the provided output data, by using the provided output data as input into a dynamic voting model, or a combination thereof.

16. The method of claim **14** or **15**, wherein determining, by the voting unit and based on the provided output data, the predicted at least one attribute comprises:

determining, by the voting unit, a number of occurrences of each initial attribute class of the multiple candidate attribute classes; and

selecting, by the voting unit, the initial attribute class of the multiple candidate attribute classes having the highest number of occurrences.

17. The method of any one of claims **14-16**, wherein each machine learning model of the plurality of machine learning models comprises a random forest classification algorithm, support vector machine, logistic regression, k-nearest neighbor model, artificial neural network, naive Bayes model, quadratic discriminant analysis, Gaussian processes model, or any combination thereof.

18. The method of any one of claims **14-16**, wherein each machine learning model of the plurality of machine learning models comprises a random forest classification algorithm.

19. The method of any one of claims **14-18**, wherein the plurality of machine learning models includes multiple representations of a same type of classification algorithm.

20. The method of any one of claims **14-18**, wherein the input data represents a description of (i) sample attributes and (ii) origins.

21. The method of claim **20**, wherein the multiple candidate attribute classes include at least one class for prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas, rectum, colon, breast, intrahepatic bile duct, cecum, gastroesophageal junction, frontal lobe,

kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin.

22. The method of claim **20**, wherein the multiple candidate attribute classes include at least at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or all 21 of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma.

23. The method of any one of claims **20-22**, wherein the sample attributes includes one or more biomarkers for the sample, wherein optionally the one or more biomarkers comprises one or more biomarkers listed in any one of **Tables 121-127**, **Tables 117-120**, INSM1, any table selected from **Tables 2-116**, and any combination thereof, optionally wherein the set of one or more biomarkers comprises one or more biomarkers listed in any one of **Table 117**, **Table 118**, **Table 119**, **Table 120**, INSM1, or any combination thereof.

24. The method of claim **23**, wherein the one or more biomarkers comprises each of the biomarkers in **Table 118**, **Table 119**, **Table 120**, and INSM1, and wherein optionally the set of one or more biomarkers further comprises the markers in any table selected from **Tables 2-116**.

25. The method of claim **23**, wherein the one or more biomarkers includes a panel of genes that is less than all known genes of the sample.

26. The method of claim **23**, wherein the one or more biomarkers includes a panel of genes that comprises all known genes for the sample.

27. The method of any one of claims **20-26**, wherein the input data further includes data representing a description of the sample and/or subject.

28. A system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described with reference to any one of claims **14-27**.

29. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described with reference to any one of claims **14-27**.

30. A method for classifying a biological sample, the method comprising:

obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample;

obtaining, by one or more computers, second data representing another initial classification for the biological sample that were previously determined based on DNA sequences of the biological sample;

providing, by one or more computers, at least a portion of the first data and the second data as an input to a dynamic voting engine that has been trained to predict a target biological sample classification based on processing of multiple initial biological sample classifications;

processing, by one or more computers, the provided input data through the dynamic voting engine;

obtaining, by one or more computers, output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the provided input data; and

determining, by one or more computers, a target biological sample classification for the biological sample based on the obtained output data.

31. The method of claim **30**,

wherein obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample comprises:

obtaining data representing a cancer type classification for the biological sample based the RNA sequences of the biological sample;

obtaining data representing an organ from which the biological sample originated based on the RNA sequences of the biological sample; and

obtaining data representing a histology for the biological sample based on the RNA sequences of the biological sample,

and

wherein providing at least a portion of the first data and the second data as an input to the dynamic voting engine comprises:

providing the obtained data representing the cancer type classification, the obtained data representing the organ from which the biological sample originated, the obtained data representing the histology, and the second data as an input to the dynamic voting engine.

32. The method of claim **30**, wherein the dynamic voting engine comprises one or more machine learning models.

33. The method of claim **30**, wherein training the dynamic voting engine comprises:

obtaining a labeled training data item that includes (I) one or more initial classifications that

include data indicating a cancer classification type, data indicating an initial organ of origin, data indicating a histology, or data indicating output of a DNA analysis engine and (II) a target biological sample classification;

generating training input data for input to the dynamic voting engine based on the obtained training data item;

processing the generated training input data through the dynamic voting engine;

obtaining output data generated by the dynamic voting engine based on the dynamic voting engine's processing of the generated training input data; and

adjusting one or more parameters of the dynamic voting engine based on the level of similarity between the output data and the label of the obtained training data item.

34. The method of claim **30**, wherein previously determining an initial classification for the biological sample based on DNA sequences of the biological sample comprises:

receiving, by one or more computers, a biological signature representing the biological sample that was obtained from a cancerous neoplasm in a first portion of a body, wherein the model includes a cancerous biological signature for each of multiple different types of cancerous biological samples, wherein each of the cancerous biological signatures include at least a first cancerous biological signature representing a molecular profile of a cancerous biological sample from the first portion of one or more other bodies and a second cancerous biological signature representing a molecular profile of a cancerous biological sample from a second portion of one or more other bodies;

performing, by one or more computers and using a pairwise-analysis model, pairwise analysis of the biological signature using the first cancerous biological signature and the second cancerous biological signature;

generating, by one or more computers and based on the performed pairwise analysis, a likelihood that the cancerous neoplasm in the first portion of the body was caused by cancer in a second portion of the body; and

storing, by one or more computers, the generated likelihood in a memory device.

35. A system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform each of the operations described with reference to any one of claims **30-34**.

36. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform the operations described with reference to any one of claims **30-34**.

37. A method comprising:

(a) obtaining a biological sample from a subject having a cancer;

(b) performing at least one assay on the sample to assess one or more biomarkers, thereby obtaining a biosignature for the sample;

(c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises at least one pre-determined biosignature indicative of at least one attribute, and wherein the at least one attribute of the cancer is selected from the group comprising primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof;

(d) processing, by one or more computers, the provided biosignature through the model; and

(e) outputting from the model a prediction of the at least one attribute of the cancer.

38. The method of claim **37**, wherein the biological sample comprises formalin-fixed paraffin-embedded (FFPE) tissue, fixed tissue, a core needle biopsy, a fine needle aspirate, unstained slides, fresh frozen (FF) tissue, formalin samples, tissue comprised in a solution that preserves nucleic acid or protein molecules, a fresh sample, a malignant fluid, a bodily fluid, a tumor sample, a tissue sample, or any combination thereof.

39. The method of claim **37** or **38**, wherein the biological sample comprises cells from a solid tumor, a bodily fluid, or a combination thereof.

40. The method of any one of claims **38-39**, wherein the bodily fluid comprises a malignant fluid, a pleural fluid, a peritoneal fluid, or any combination thereof.

41. The method of any one of claims **38-40**, wherein the bodily fluid comprises peripheral blood, sera, plasma, ascites, urine, cerebrospinal fluid (CSF), sputum, saliva, bone marrow, synovial fluid, aqueous humor, amniotic fluid, cerumen, breast milk, broncheoalveolar lavage fluid, semen, prostatic fluid, Cowper's fluid, pre-ejaculatory fluid, female ejaculate, sweat, fecal matter, tears, cyst fluid, pleural fluid, peritoneal fluid, pericardial fluid, lymph, chyme, chyle, bile, interstitial fluid, menses, pus, sebum, vomit, vaginal secretions, mucosal secretion, stool water, pancreatic juice, lavage fluids from sinus cavities, bronchopulmonary aspirates, blastocyst cavity fluid, or umbilical cord blood.

42. The method of any one of claims **37-41**, wherein performing the at least one assay in step (b) comprises determining a presence, level, or state of a protein or nucleic acid for each of the one or more biomarkers, wherein optionally the nucleic acid comprises deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or a combination thereof.

43. The method of claim **42**, wherein:

i. the presence, level or state of at least one of the proteins is determined using a technique selected from immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof, wherein optionally the presence, level or state of all of the proteins is determined using the technique; and/or

ii. the presence, level or state of at least one of the nucleic acids is determined using a technique selected from polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, whole genome sequencing, whole transcriptome sequencing, or any combination thereof, wherein optionally the presence, level or state of all of the nucleic acids is determined using the technique.

44. The method of claim **43**, wherein the state of the nucleic acid comprises a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof.

45. The method of claim **44**, wherein the state of the nucleic acid consists of or comprises a copy number.

46. The method of any one of claims **37-45**, wherein the at least one assay comprises next-generation sequencing, wherein optionally the next-generation sequencing is used to assess: i) at least one of the genes, genomic information / signatures, and fusion transcripts in any of **Tables 121-130**, or any combination thereof; ii) at least one of the genes and/or transcripts in any table selected from **Tables 117-120**, INSM1, and any combination thereof; iii) the whole exome; iv) the whole transcriptome; v) at least one gene in any table selected from **Tables 2-116**, and any combination thereof; or vi) any combination thereof.

47. The method of any one of claims **37-46**, wherein the predicting the at least one attribute of the cancer comprises determining a probability that the attribute is each member of a plurality of such attributes and selecting the attribute with the highest probability.

48. The method of any one of claims **37-47**, wherein:

i. the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, or all 38 of prostate, bladder, endocervix, peritoneum, stomach, esophagus, ovary, parietal lobe, cervix, endometrium, liver, sigmoid colon, upper-outer quadrant of breast, uterus, pancreas, head of pancreas, rectum, colon, breast, intrahepatic bile duct, cecum,

gastroesophageal junction, frontal lobe, kidney, tail of pancreas, ascending colon, descending colon, gallbladder, appendix, rectosigmoid colon, fallopian tube, brain, lung, temporal lobe, lower third of esophagus, upper-inner quadrant of breast, transverse colon, and skin;

ii. the primary tumor origin or plurality of primary tumor origins consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or all 21 of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma;

iii. the cancer/disease type consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or all 28 of adrenal cortical carcinoma; bile duct, cholangiocarcinoma; breast carcinoma; central nervous system (CNS); cervix carcinoma; colon carcinoma; endometrium carcinoma; gastrointestinal stromal tumor (GIST); gastroesophageal carcinoma; kidney renal cell carcinoma; liver hepatocellular carcinoma; lung carcinoma; melanoma; meningioma; Merkel; neuroendocrine; ovary granulosa cell tumor; ovary, fallopian, peritoneum; pancreas carcinoma; pleural mesothelioma; prostate adenocarcinoma; retroperitoneum; salivary and parotid; small intestine adenocarcinoma; squamous cell carcinoma; thyroid carcinoma; urothelial carcinoma; uterus;

iv. the organ group consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or all 17 of adrenal gland; bladder; brain; breast; colon; eye; female genital tract and peritoneum (FGTP); gastroesophageal; head, face or neck, NOS; kidney; liver, gallbladder, ducts; lung; pancreas; prostate; skin; small intestine; thyroid; and/or

v. the histology consists of, comprises, or comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or all 29 of adenocarcinoma, adenoid cystic carcinoma, adenosquamous carcinoma, adrenal cortical carcinoma, astrocytoma, carcinoma, carcinosarcoma, cholangiocarcinoma, clear cell carcinoma, ductal carcinoma in situ (DCIS), glioblastoma (GBM), GIST, glioma, granulosa cell tumor, infiltrating lobular carcinoma, leiomyosarcoma, liposarcoma, melanoma, meningioma, Merkel cell carcinoma, mesothelioma,

neuroendocrine, non-small cell carcinoma, oligodendroglioma, sarcoma, sarcomatoid carcinoma, serous, small cell carcinoma, squamous.

49. The method of any one of claims **37-48**, wherein the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, optionally a cancer/disease type, comprises selections of biomarkers according to **Table 118**, wherein optionally:

i. a pre-determined biosignature indicative of adrenal cortical carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from INHA, MIB1, SYP, CDH1, NKX3-1, CALB2, KRT19, MUC1, S100A5, CD34, TMPRSS2, KRT8, NCAM2, ARG1, TG, NCAM1, SERPINA1, PSAP, TPM3, and ACVRL1;

ii. a pre-determined biosignature indicative of bile duct, cholangiocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from HNF1B, VIL1, SERPINA1, ESR1, ANO1, SOX2, MUC4, S100A2, KRT5, KRT7, CNN1, AR, ENO2, S100A9, NKX2-2, SATB2, PSAP, S100A6, CALB2, and TMPRSS2;

iii. a pre-determined biosignature indicative of breast carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, ANKRD30A, KRT15, KRT7, S100A2, PAX8, MUC4, KRT18, HNF1B, S100A1, PIP, SOX2, MDM2, MUC5AC, PMEL, TFF1, KRT16, KRT6B, S100A6, and SERPINB5;

iv. a pre-determined biosignature indicative of central nervous system (CNS) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT18, KRT8, SOX2, ANO1, NCAM1, PDPN, NKX2-2, KRT19, S100A14, S100A11, S100A1, MSH2, CEACAM1, GPC3, ERBB2, TG, KRT7, CGB3, and S100A2;

v. a pre-determined biosignature indicative of cervix carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ESR1, CDKN2A, CCND1, LIN28A, PGR, SMARCB1, CEACAM4, S100B, FUT4, PSAP, MUC2, MDM2, NCAM1, SATB2, TNFRSF8, CD79A, S100A13, VHL, CD3G, and TPSAB1;

vi. a pre-determined biosignature indicative of colon carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, KRT7, MUC2, KRT20, MUC1, SATB2, VIL1, CEACAM5, CDH17, S100A6, CEACAM20, KRT6B, TFF3, FUT4, BCL2, KRT6A, KRT18, CEACAM18, TFF1, and MLH1;

vii. a pre-determined biosignature indicative of endometrium carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, PGR, ESR1, VHL, CALD1, LIN28B, NAPSA, KRT5, S100A6, DES, FLI1, DSC3, S100P, CEACAM16, PDPN, ARG1, TLE1, WT1, BCL6, and MLH1;

viii. a pre-determined biosignature indicative of gastrointestinal stromal tumor (GIST) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ANO1, SDC1, KRT19, MUC1, KRT8, ACVRL1, KIT, CDH1, S100A2, KRT7, ERBB2, S100A16, ENO2, S100A9, TPSAB1, KRT17, PAX8, PGR, ESR1, and VHL;

ix. a pre-determined biosignature indicative of gastroesophageal carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FUT4, CDX2, SERPINB5, MUC5AC, AR, TFF1, NCAM2, TFF3, ISL1, ANO1, VIL1, PAX8, SOX2, CEACAM6, S100A13, ENO2, NAPSA, TPSAB1, S100B, and CD34;

x. a pre-determined biosignature indicative of kidney renal cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, CDH1, CDKN2A, S100P, S100A14, HAVCR1, HNF1B, KL, KRT7, MUC1, POU5F1, VHL, PAX2, AMACR, BCL6, S100A13, CA9, MDM2, SALL4, and SYP;

xi. a pre-determined biosignature indicative of liver hepatocellular carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, CEACAM16, KRT19, AFP, MUC4, CEACAM5, MSH2, BCL6, DSC3, KRT15, S100A6, CEACAM20, GPC3, MUC1, CD34, VIL1, ERBB2, POU5F1, KRT18, and KRT16;

xii. a pre-determined biosignature indicative of lung carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,

16, 17, 18, 19, or 20 features selected from NAPSA, SOX2, CEACAM7, KRT7, S100A10, CEACAM6, S100A1, PAX8, AR, VHL, S100A13, CD99L2, KRT5, MUC1, CEACAM1, SFTPA1, TMPRSS2, TFF1, KRT15, and MUC4;

xiii. a pre-determined biosignature indicative of melanoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT8, PMEL, KRT19, MUC1, MLANA, S100A14, S100A13, MITF, S100A1, VIM, CDKN2A, ACVRL1, MS4A1, POU5F1, TPM1, UPK3A, S100P, GATA3, and CEACAM1;

xiv. a pre-determined biosignature indicative of meningioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SDC1, KRT8, ANO1, VIM, S100A14, S100A2, CEACAM1, MSH2, PGR, KRT10, TP63, CD5, INHA, CDH1, CCND1, MDM2, KRT16, SPN, SMARCB1, and S100A9;

xv. a pre-determined biosignature indicative of Merkel cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, ERBB2, S100A12, S100A14, MYOG, SDC1, KRT7, S100PBP, MME, TMPRSS2, CEACAM5, CPS1, CR1, MUC4, CEACAM4, CA9, ENO2, FLI1, LIN28B, and MLANA;

xvi. a pre-determined biosignature indicative of neuroendocrine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, ISL1, ENO2, POU5F1, TFF3, SYP, TPM4, S100A1, S100Z, MUC4, MPO, DSC3, CEACAM4, S100A7, ERBB2, CDX2, S100A11, KRT10, CEACAM5, and CEACAM3;

xvii. a pre-determined biosignature indicative of ovary granulosa cell tumor consists of, comprises, or comprises at least, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FOXL2, SDC1, MSH6, MUC1, KRT8, PGR, MME, SERPINA1, FLI1, S100B, CEACAM21, AMACR, KRT1, SFTPA1, TPM1, CALCA, S100A11, NCAM1, ISL1, and ENO2;

xviii. a pre-determined biosignature indicative of ovary, fallopian, peritoneum consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from WT1, PAX8, INHA, TFE3, S100A13, FOXL2, TLE1, MSLN, POU5F1, CEACAM3, ALPP, S100A10, FUT4, NKX3-1, CEACAM5, SOX2, ESR1, ENO2, ACVRL1, and SYP;

xix. a pre-determined biosignature indicative of pancreas carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, GATA3, ANO1, SERPINA1, ISL1, MUC5AC, FUT4, SMAD4, CD5, CALB2, S100A4, SMN1, ESR1, HNF1B, AMACR, MSH2, PDPN, MSLN, TFF1, and KRT6C;

xx. a pre-determined biosignature indicative of pleural mesothelioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from UPK3B, CALB2, WT1, SMARCB1, PDPN, INHA, CEACAM1, MSLN, KRT5, CA9, S100A13, SF1, CDH1, CDKN2A, FLI1, SYP, CEACAM3, CPS1, SATB2, and BCL6;

xxi. a pre-determined biosignature indicative of prostate adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT7, KLK3, NKX3-1, AMACR, S100A5, MUC1, MUC2, UPK3A, KL, CPS1, MSLN, PMEL, CNN1, SERPINA1, KRT2, CGB3, TMPRSS2, CEACAM6, SDC1, and AR;

xxii. a pre-determined biosignature indicative of retroperitoneum consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT18, KRT8, TPM1, S100A14, CD34, TPM4, CDH1, CNN1, SDC1, AR, MDM2, KIT, TLE1, CPS1, CDK4, UPK3A, TMPRSS2, TPM3, and CEACAM1;

xxiii. a pre-determined biosignature indicative of salivary and parotid consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ENO2, PIP, TPM1, KRT14, S100A1, ERBB2, TFF1, ALPP, DSC3, CTNNB1, CALB2, SALL4, ANO1, CEACAM16, HNF1B, KIT, ARG1, CEACAM18, TMPRSS2, and HAVCR1;

xxiv. a pre-determined biosignature indicative of small intestine adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, DES, MUC2, CDH17, CEACAM5, SERPINA1, KRT20, HNF1B, ESR1, ARG1, CD5, TLE1, PMEL, SOX2, SFTPA1, MME, CD99L2, MPO, S100P, and CA9;

xxv. a pre-determined biosignature indicative of squamous cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, SOX2, KRT6A, KRT17, S100A1, CD3G, SFTPA1, AR, KRT5, SDC1, KRT20, DSC3, CNN1, MSH2, ESR1, S100A2, SERPINB5, PDPN, S100A14, and TPM3;

xxvi. a pre-determined biosignature indicative of thyroid carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TG, PAX8, CPS1, S100A2, TPSAB1, CALB2, HNF1B, INHA, ARG1, CNN1, CDK4, VIM, CEACAM5, TLE1, TFF3, KRT8, S100P, FOXL2, MUC1, and GATA3;

xxvii. a pre-determined biosignature indicative of urothelial carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, UPK2, KRT20, MUC1, S100A2, CPS1, TP63, CALB2, MITF, S100P, SERPINA1, DES, CTNNB1, MSLN, SALL4, VHL, KRT7, CD2, PAX8, and UPK3A; and/or

xxviii. a pre-determined biosignature indicative of uterus consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT18, NCAM1, DES, FOXL2, CD79A, S100A14, ESR1, MSLN, MITF, UPK3B, TPM1, ENO2, S100P, MLH1, KRT8, CDH1, TPM4, SATB2, and MDM2.

50. The method of any one of claims **37-48**, wherein the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, optionally an organ type, comprises selections of biomarkers according to **Table 119**; wherein optionally:

i. a pre-determined biosignature indicative of adrenal gland consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from INHA, CDH1, SYP, MIB1, CALB2, KRT8, PSAP, KRT19, NCAM2, NKX3-1, ARG1, SERPINA1, CD34, TPM3, S100A7, ACVRL1, PMEL, CR1, ERG, and PECAM1;

ii. a pre-determined biosignature indicative of bladder consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, KRT20, UPK2, CPS1, SALL4, SERPINA1, DES, CALB2, MUC1, S100A2, MSLN, MITF, PAX8, S100A10, CNN1, UPK3A, CD3G, NAPSA, CD2, and MME;

iii. a pre-determined biosignature indicative of brain consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT8, ANO1, S100B, S100A14, SOX2, PDPN, CEACAM1, S100A2, NCAM1, MSH2, KRT18, NKX2-2, WT1, S100A1, GPC3, TLE1, CD5, S100Z, S100A16, and PGR;

iv. a pre-determined biosignature indicative of breast consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, ANKRD30A, KRT15, KRT7,

S100A2, S100A1, MUC4, HNF1B, KRT18, SOX2, PIP, PAX8, MDM2, KRT16, MUC5AC, S100A6, TP63, TFF1, KRT5, and SERPINA1;

v. a pre-determined biosignature indicative of colon consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, KRT7, MUC2, KRT20, MUC1, CEACAM5, CDH17, TFF3, KRT18, KRT6B, VIL1, SATB2, S100A6, SOX2, S100A14, HAVCR1, FUT4, ERG, HNF1B, and PTPRC;

vi. a pre-determined biosignature indicative of eye consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PMEL, MLANA, MITF, BCL2, S100A13, S100A2, S100A10, S100A1, MIB1, SOX2, ENO2, S100A16, VIM, VHL, PDPN, WT1, S100B, KRT7, KRT10, and PSAP;

vii. a pre-determined biosignature indicative of female genital tract and peritoneum (FGTP) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, ESR1, WT1, PGR, CDKN2A, FOXL2, KRT5, TPM4, SMARCB1, DES, TMPRSS2, CDK4, GATA3, AR, S100A13, MSH2, ANO1, CALB2, MS4A1, and CCND1;

viii. a pre-determined biosignature indicative of gastroesophageal consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDX2, ANO1, FUT4, SERPINB5, SPN, NCAM2, VIL1, CD34, ENO2, TFF3, AR, S100A13, TPM1, CEACAM6, SOX2, PAX8, MUC5AC, CDH1, S100A11, and ISL1;

ix. a pre-determined biosignature indicative of head, face or neck, NOS consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT5, DSC3, TP63, HNF1B, MUC5AC, PAX5, KRT15, PGR, S100A6, TMPRSS2, MME, S100B, ENO2, CEACAM8, SALL4, ANO1, GATA3, LIN28B, CD99L2, and UPK3A;

x. a pre-determined biosignature indicative of kidney consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, CDH1, HNF1B, S100A14, HAVCR1, CDKN2A, S100P, KL, KRT7, S100A13, VHL, PAX2, POU5F1, MUC1, AMACR, ENO2, MDM2, WT1, SYP, and AR;

xi. a pre-determined biosignature indicative of liver, gallbladder, ducts consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, VIL1, HNF1B,

ANO1, ESR1, SOX2, MUC4, S100A2, ENO2, CNN1, POU5F1, KRT5, S100A9, UPK3B, PSAP, KRT7, KL, TMPRSS2, SATB2, and S100A14;

xii. a pre-determined biosignature indicative of lung consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NAPSA, SOX2, SFTPA1, VHL, S100A1, S100A10, AR, TMPRSS2, CD99L2, CEACAM7, CEACAM6, KRT6A, KRT7, NCAM2, TP63, CEACAM1, MUC4, KRT20, CNN1, and ISL1;

xiii. a pre-determined biosignature indicative of pancreas consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PDX1, ANO1, SERPINA1, GATA3, ISL1, MUC5AC, SMAD4, FUT4, CD5, SMN1, NKX2-2, TFF1, AMACR, SOX2, HNF1B, S100Z, MSLN, DES, S100A4, and CALB2;

xiv. a pre-determined biosignature indicative of prostate consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KLK3, KRT7, NKX3-1, AMACR, CPS1, S100A5, UPK3A, KL, MUC1, CGB3, MUC2, TMPRSS2, MSLN, PMEL, S100A10, SERPINA1, KRT20, SFTPA1, BCL6, and TFF1;

xv. a pre-determined biosignature indicative of skin consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT8, PMEL, KRT7, KRT19, GATA3, MDM2, AMACR, TPM1, TLE1, CEACAM19, CEACAM16, MLANA, TMPRSS2, AR, TFF3, BCL6, CR1, NCAM1, and MS4A1;

xvi. a pre-determined biosignature indicative of small intestine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MUC2, CDH17, FLI1, KRT20, CDX2, CD5, KRT7, MPO, CNN1, DSC3, DES, ANO1, S100A1, CALD1, TFF1, SPN, MITF, TMPRSS2, CALB2, and CEACAM16; and/or

xvii. a pre-determined biosignature indicative of thyroid consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from PAX8, TG, CPS1, SERPINB5, INHA, ARG1, CNN1, CEACAM5, TPSAB1, CALB2, HNF1B, VIM, CDK4, S100P, S100A2, LIN28B, TFF3, CGA, TLE1, and TPM3.

51. The method of any one of claims **37-48**, wherein the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, optionally a histology, comprises selections of biomarkers according to **Table 120**; wherein optionally:

i. a pre-determined biosignature indicative of adenocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TMPRSS2, HNF1B, KRT5, MUC1, CEACAM5, MUC5AC, CDH17, TP63, ALPP, GATA3, CEACAM1, TFF3, S100A1, KRT8, PDX1, KRT17, CDH1, KLK3, CPS1, and S100A2;

ii. a pre-determined biosignature indicative of adenoid cystic carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT14, KIT, TPM3, CGA, SMAD4, CTNNB1, DSC3, S100A6, TP63, TPM1, CALD1, MIB1, CD2, CDH1, ANO1, ENO2, CD3G, TPM2, CEACAM1, and BCL2;

iii. a pre-determined biosignature indicative of adenosquamous carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, SFTPA1, OSCAR, KRT19, KRT15, NAPSA, GPC3, MS4A1, S100A12, ERG, CEACAM6, VHL, SOX2, SERPINA1, KRT6A, CDKN2A, CD3G, PIP, NCAM2, and CEACAM7;

iv. a pre-determined biosignature indicative of adrenal cortical carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MIB1, INHA, CDH1, SYP, CALB2, NKX3-1, KRT19, ERBB2, MUC1, ARG1, VIM, CD34, CALD1, S100A9, MSLN, S100A10, CD5, PMEL, SDC1, and TP63;

v. a pre-determined biosignature indicative of astrocytoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, SOX2, NCAM1, MUC1, S100A4, KRT17, KRT8, S100A1, TPM4, CNN1, TPM2, OSCAR, AR, SDC1, SALL4, SMN1, SFTPA1, KIT, CA9, and S100A9;

vi. a pre-determined biosignature indicative of carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, MITF, MUC5AC, PDPN, VIL1, CEACAM5, CDH1, CDH17, IL12B, S100P, KRT20, KRT7, SPN, TMPRSS2, ENO2, NKX2-2, PMEL, IMP3, BCL6, and S100A8;

vii. a pre-determined biosignature indicative of carcinosarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT6B, GPC3, MSLN, MUC1, S100A6, S100A2, MME, CDKN2A, CDH1, FOXL2, KRT7, CALB2, SFTPA1, ERG, PGR, KRT17, NAPSA, CALD1, LIN28B, and KIT;

viii. a pre-determined biosignature indicative of cholangiocarcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SERPINA1, HNF1B, VIL1, TFF1, ENO2, NKX2-2, FUT4, MUC4, MLH1, TMPRSS2, WT1, KL, KRT7, ESR1, MDM2, SFTPA1, SMN1, KRT18, UPK3B, and COQ2;

ix. a pre-determined biosignature indicative of clear cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from POU5F1, HAVCR1, CEACAM6, HNF1B, PAX8, NAPSA, CD34, MYOG, FOXL2, MITF, S100P, S100A9, S100A14, S100Z, WT1, CDH1, TTF1, SYP, MLH1, and KRT16;

x. a pre-determined biosignature indicative of ductal carcinoma in situ (DCIS) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from GATA3, HNF1B, DES, MME, ANKRD30A, SATB2, SOX2, NCAM2, PAX8, CEACAM4, PIP, MUC4, NKX3-1, SERPINA1, KRT20, KIT, NCAM1, KRT14, S100A2, and CDKN2A;

xi. a pre-determined biosignature indicative of glioblastoma (GBM) consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, KRT18, PDPN, NKX2-2, SOX2, NCAM1, KRT8, ERBB2, KRT15, KRT19, GATA3, CDKN2A, BCL6, S100A14, KRT10, UPK3A, SF1, CA9, CCND1, and KRT5;

xii. a pre-determined biosignature indicative of GIST consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ANO1, SDC1, MUC1, KRT19, KRT8, ACVRL1, KIT, ERBB2, CDH1, CEACAM19, FUT4, TFF3, S100A16, S100A13, ISL1, S100A9, TPSAB1, KRT18, IMP3, and KRT3;

xiii. a pre-determined biosignature indicative of glioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT8, S100B, SYP, NCAM2, CD3G, SDC1, SOX2, CEACAM1, POU5F1, MIB1, SATB2, MDM2, NCAM1, KRT7, CGB3, CPS1, PDPN, CALCA, ERBB2, and TNFRSF8;

xiv. a pre-determined biosignature indicative of granulosa cell tumor consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from FOXL2, SDC1, MSH6, KRT18, KRT8, MME, FLI1, S100A9, CALCA, S100B, CCND1, CEACAM21, TLE1, SERPINA1, S100A11, SFTPA1, SYP, NCAM2, CD3G, and SOX2;

xv. a pre-determined biosignature indicative of infiltrating lobular carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from CDH1, GATA3, S100A1, TFF3, CA9, MUC1, NKX3-1, ANKRD30A, SOX2, S100A5, MUC4, KRT7, OSCAR, MME, SERPINA1, CDK4, AR, CEACAM3, BCL6, and KRT5;

xvi. a pre-determined biosignature indicative of leiomyosarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT19, KRT8, KRT18, CNN1, TPM4, FOXL2, TPM2, TPM1, CD79A, CALB2, SATB2, S100A5, DES, S100A14, KRT2, ERBB2, PDPN, ENO2, CD2, and CALD1;

xvii. a pre-determined biosignature indicative of liposarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from KRT18, MDM2, CDK4, CDH1, KRT19, KRT7, PDPN, CD34, TPM4, CR1, ACVRL1, MME, KRT8, AMACR, CEACAM5, S100B, OSCAR, LIN28A, S100A12, and SDC1;

xviii. a pre-determined biosignature indicative of melanoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from S100B, PMEL, KRT19, KRT8, MUC1, S100A14, MLANA, S100A13, TPM1, MITF, VIM, CEACAM19, POU5F1, SATB2, CPS1, CDKN2A, KRT10, AR, ACVRL1, and LIN28A;

xix. a pre-determined biosignature indicative of meningioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from SDC1, KRT8, S100A14, ANO1, CEACAM1, VIM, KRT10, PGR, MSH2, CD5, S100A2, CDH1, TP63, SMARCB1, KRT16, S100A10, S100A4, DSC3, CCND1, and GATA3;

xx. a pre-determined biosignature indicative of Merkel cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, ERBB2, MME, MYOG, CPS1, KRT7, SALL4, S100A12, S100A14, S100PBP, CR1, SMAD4, CEACAM5, MUC4, CA9, KRT10, SYP, CCND1, MSLN, and MLANA;

xxi. a pre-determined biosignature indicative of mesothelioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from UPK3B, CALB2, PDPN, SMARCB1, MSLN, KRT5, CEACAM3, WT1, INHA, CEACAM1, CA9,

TLE1, SATB2, CDH1, MUC2, CDKN2A, CEACAM18, MSH2, DSC3, and PTPRC;

xxii. a pre-determined biosignature indicative of neuroendocrine consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ISL1, NCAM1, S100A11, ENO2, S100A1, SYP, MUC1, TFF3, S100Z, PAX8, ERBB2, ESR1, S100A10, CEACAM5, SDC1, MUC4, MPO, S100A4, S100A7, and TP63;

xxiii. a pre-determined biosignature indicative of non-small cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from ESR1, TMPRSS2, AR, S100A1, SFTPA1, MSLN, SOX2, ENO2, TP63, SMAD4, PTPRC, ISL1, CEACAM7, CEACAM20, S100Z, INHA, NCAM1, MUC2, TFF3, and PAX8;

xxiv. a pre-determined biosignature indicative of oligodendroglioma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, KRT18, CD2, S100A11, SYP, CDH1, S100A4, S100A14, CEACAM1, S100PBP, SDC1, SALL4, UPK2, COQ2, TPM2, CD99L2, TTF1, CD79A, INHA, and VIM;

xxv. a pre-determined biosignature indicative of sarcoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from NCAM1, KRT19, S100A14, NKX2-2, KRT2, KRT7, SATB2, MYOG, CALD1, CEACAM19, CA9, KRT15, CDKN2A, S100P, WT1, TMPRSS2, S100A7, SERPINB5, DSC3, and ENO2;

xxvi. a pre-determined biosignature indicative of sarcomatoid carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from MME, VIM, S100A14, CD99L2, S100A11, NKX3-1, SATB2, CPS1, MSLN, SFTPA1, POU5F1, CDH1, OSCAR, S100A5, IMP3, CEACAM1, PMS2, NCAM2, KRT15, and S100A12;

xxvii. a pre-determined biosignature indicative of serous consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from WT1, PAX8, KRT7, CDKN2A, MSLN, ACVRL1, SATB2, CDK4, DSC3, AR, S100A16, ANO1, S100A5, SDC1, IMP3, SERPINA1, KRT4, ESR1, FOXL2, and KRT15;

xxviii. a pre-determined biosignature indicative of small cell carcinoma consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,

16, 17, 18, 19, or 20 features selected from NCAM1, ISL1, PAX5, KIT, MUC4, S100A10, MUC1, CTNNB1, MITF, NKX2-2, S100A11, SMN1, MSLN, S100A6, BCL2, SYP, KL, CGB3, TPSAB1, TFF3; and/or

xxix. a pre-determined biosignature indicative of squamous consists of, comprises, or comprises at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 features selected from TP63, KRT5, KRT17, SOX2, AR, CD3G, KRT6A, S100A1, DSC3, SERPINB5, HNF1B, SDC1, S100A6, TPSAB1, KRT20, HAVCR1, TTF1, MSH2, PMS2, and CNN1.

52. The method of any one of claims **37-51**, wherein the at least one pre-determined biosignature indicative of the at least one attribute of the cancer comprises selections of biomarkers according claim **49**, claim **50**, and/or claim **51**.

53. The method of any one of claims **49-52**, wherein performing the at least one assay to assess the one or more biomarkers in step (b) comprises assessing the markers in the at least one pre-determined biosignature using DNA analysis and/or expression analysis, wherein:

i. the DNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof;

ii. the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof; and/or

iii. the expression analysis consists of or comprises analysis of RNA, where optionally:

  i. the RNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, amount, level, expression level, presence, or any combination thereof; and/or

  ii. the RNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole transcriptome sequencing, or any combination thereof;

iv. the expression analysis consists of or comprises analysis of protein, where optionally:

i. the protein analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, fusion, amplification, amount, level, expression level, presence, or any combination thereof; and/or

ii. the protein analysis is performed using immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof; and/or

v. any combination thereof.

54. The method of claim **53**, wherein performing the assay to assess the one or more biomarkers in step (b) comprises assessing the markers in the at least one pre-determined biosignature using: a combination of the DNA analysis and the RNA analysis; a combination of the DNA analysis and the protein analysis; a combination of the RNA analysis and the protein analysis; or a combination of the DNA analysis, the RNA analysis, and the protein analysis.

55. The method of claim **53** or **54**, wherein performing the assay to assess the one or more biomarkers in step (b) comprises RNA analysis of messenger RNA transcripts.

56. The method of any one of claims **37-55**, wherein the at least one pre-determined biosignature indicative of the at least one attribute of the cancer, optionally a primary tumor origin, comprises selections of biomarkers according to at least one of **FIGs. 6I-AC**; wherein optionally:

i. a pre-determined biosignature indicative of breast adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, CDH1, PAX8, KRAS, ELK4, CCND1, MECOM, PBX1, CREBBP, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, NY-BR-1, KRT15, CK7, S100A2, RCCMa, MUC4, CK18, HNF1B and S100A1;

ii. a pre-determined biosignature indicative of central nervous system cancer comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from IDH1, SOX2, OLIG2, MYC, CREB3L2, SPECC1, EGFR, FGFR2, SETBP1, and ZNF217, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from S100B, CK18, CK8, SOX2, DOG1, CD56, PDPN, NKX2-2, CK19, and S100A14;

iii. a pre-determined biosignature indicative of cervical adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, MECOM, RPN1, U2AF1, GNAS, RAC1, KRAS, FL11, EXT1, and CDK6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8,

9, or 10 features selected from ER, p16, CYCLIND1, LIN28A, PR, SMARCB1, CEACAM4, S100B, CD15, and PSAP;

iv. a pre-determined biosignature indicative of cholangiocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, ARID1A, MAF, KRAS, CACNA1D, SPEN, SETBP1, CDK12, LHFPL6, and MDS2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from HNF1B, VILLIN, ANTITRYPSIN, ER, DOG1, SOX2, MUC4, S100A2, KRT5, and CK7;

v. a pre-determined biosignature indicative of colon adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from APC, CDX2, KRAS, SETBP1, FLT3, LHFPL6, CDKN2A, FLT1, ASXL1, and CDKN2B, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CDX2, CK7, MUC2, CK20, MUC1, SATB2, VILLIN, CEACAM5, CDK17, and S100A6;

vi. a pre-determined biosignature indicative of gastroesophageal adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CDX2, ERG, TP53, KRAS, U2AF1, ZNF217, CREB3L2, IRF4, TCF7L2, and LHFPL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CD15, CDX2, MASPIN, MUC5AC, AR, TFF1, NCAM2, TFF3, ISL1, and DOG1;

vii. a pre-determined biosignature indicative of gastrointestinal stromal tumor (GIST) comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from c-KIT (KIT), TP53, MAX, PDGFRA, TSHR, MSI2, SPEN, JAK1, SETBP1, and CDH11, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from DOG1, CD138, CK19, MUC1, CK8, ACVRL1, KIT, E-CADHERIN, S100A2, and CK7;

viii. a pre-determined biosignature indicative of hepatocellular carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from HLF, CACNA1D, HMGN2P46, KRAS, FANCF, PRCC, ERG, FLT1, FGFR1, and ACSL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from ANTITRYPSIN, CEACAM16, CK19, AFP, MUC4, CEACAM5, MSH2, BCL6, DSC3, and KRT15;

ix. a pre-determined biosignature indicative of lung adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from NKX-2, KRAS, TP53, TPM4, CDX2, TERT, FOXA1, SETBP1, CDKN2A, and LHFPL6, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or

10 features selected from Napsin A, SOX2, CEACAM7, CK7, S100A10, CEACAM6, S100A1, RCCMa, AR and VHL;

x. a pre-determined biosignature indicative of melanoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from IRF4, SOX10, TP53, BRAF, FGFR2, TRIM27, EP300, CDKN2A, LRP1B, and NRAS, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from S100B, CK8, HMB-45, CD19, MUC1, MLANA, S100A14, S100A13, MITF, and S100A1;

xi. a pre-determined biosignature indicative of meningioma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CHEK2, TP53, MYCL, THRAP3, MPL, EBF1, EWSR1, PMS2, FLI1, and NTRK2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from CD138, CK8, DOG1, VIM, S100A14, S100A2, CEACAM1, MSH2, PR, and KRT10;

xii. a pre-determined biosignature indicative of ovarian granulosa cell tumor comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXL2, TP53, EWSR1, CBFB, SPECC1, BCL3, MYH9, TSHR, GID4, and SOX2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXL2, CD138, MSH6, MUC1, CK8, PR, MME, ANTITRYPSIN, FLI1, and S100B;

xiii. a pre-determined biosignature indicative of ovarian & fallopian tube adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, MECOM, KRAS, TPM4, RAC1, ASXL1, EP300, CDX2, RPN1, and WT1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from WT1, RCCMa, INHIBIN-alpha, TFE3, S100A13, FOLX2, TLE1, MSLN, POU5F1, and CEACAM3;

xiv. a pre-determined biosignature indicative of pancreas adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from KRAS, CDKN2A, CDKN2B, FANCF, IRF4, TP53, ASXL1, SETBP1, APC, and FOXO1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from PDX1, GATA3, DOG1, ANTITRYPSIN, ISL1, MUC5AC, CD15, SMAD4, CD5, and CALB2;

xv. a pre-determined biosignature indicative of prostate adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from FOXA1, PTEN, KLK2, FOXO1, GATA2, FANCA, LHFPL6, KRAS, ETV6, and ERCC3, and/or expression analysis of at least, 1, 2, 3, 4,

5, 6, 7, 8, 9, or 10 features selected from CK7, PSA, NKX3-1, AMACR, S100A5, MUC1, MUC2, UPK3A, KL and HEPPAR-1;

xvi. a pre-determined biosignature indicative of renal cell carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from VHL, TP53, EBF1, MAF, RAF1, CTNNA1, XPC, MUC1, KRAS, and BTG1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RCCMa, E-CADHERIN, p16, S100P, S100A14, HAVCR1, HNF1B, KL, CK7, and MUC1;

xvii. a pre-determined biosignature indicative of squamous cell carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from TP53, SOX2, KLHL6, CDKN2A, LPP, CACNA1D, TFRC, KRAS, RPN1, and CDX2, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from P63, SOX2, CK6, KRT17, S100A1, CD3G, SFTPA1, AR, KRT5, and CD138;

xviii. a pre-determined biosignature indicative of thyroid cancer comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from BRAF, NKX2-1, TP53, MYC, KDSR, TRRAP, CDX2, KRAS, FHIT, and SETBP1, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from THYROGLOBULIN, RCCMa, HEPPAR-1, S100A2, TPSAB1, CALB2, HNF1B, INHIBIN-alpha, ARG1, and CNN1;

xix. a pre-determined biosignature indicative of urothelial carcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, ASXL1, CDKN2B, TP53, CTNNA1, CDKN2A, KRAS, IL7R, CREBBP, and VHL, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from GATA3, UPII, CK20, MUC1, S100A2, HEPPAR-1, P63, CALB2, MITF, and S100P;

xx. a pre-determined biosignature indicative of uterine endometrial adenocarcinoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from PTEN, PAX8, PIK3CA, CCNE1, TP53, MECOM, ESR1, CDX2, CDKN2A, and KRAS, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RCCMa, PR, ER, VHL, CALD1, LIN28B, Napsin A, KRT5, S100A6, and DES; and/or

xxi. a pre-determined biosignature indicative of uterine sarcoma comprises DNA analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 features selected from RB1, SPECC1, FANCC, TP53, CACNA1D, JAK1, ETV1, PRRX1, PTCH1, and HOXD13, and/or expression analysis of at least, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10

features selected from CK19, CK18, CD56, DES, FOXL2, CD79A, S100A14, ER, MSLN, and MITF.

57. The method of claim **56**, wherein:

i. the DNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA), or any combination thereof;

ii. the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof;

iii. the expression analysis consists of or comprises analysis of RNA, where optionally:

i. the RNA analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, amount, level, expression level, presence, or any combination thereof; and/or

ii. the RNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole transcriptome sequencing, or any combination thereof;

iv. the expression analysis consists of or comprises analysis of protein, where optionally:

i. the protein analysis consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, fusion, amplification, amount, level, expression level, presence, or any combination thereof; and/or

ii. the protein analysis is performed using immunohistochemistry (IHC), flow cytometry, an immunoassay, an antibody or functional fragment thereof, an aptamer, mass spectrometry, or any combination thereof; and/or

v. any combination thereof.

58. The method of any one of claims **37-57**, wherein the at least one pre-determined biosignature comprises or further comprises selections of biomarkers according to any one of **Tables 2-116** assessed using DNA analysis, and the DNA analysis:

i. consists of or comprises determining a sequence, mutation, polymorphism, deletion, insertion, substitution, translocation, fusion, break, duplication, amplification, repeat, copy number, copy number variation (CNV; copy number alteration; CNA) or any combination thereof; and/or

ii. the DNA analysis is performed using polymerase chain reaction (PCR), in situ hybridization, amplification, hybridization, microarray, nucleic acid sequencing, dye termination sequencing, pyrosequencing, next generation sequencing (NGS; high-throughput sequencing), whole exome sequencing, or any combination thereof.

59. The method of claim **58**, wherein the at least one pre-determined biosignature comprising selections of biomarkers according to any one of **Tables 2-116** comprises:

i. a pre-determined biosignature indicative of adrenal cortical carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 2**;

ii. a pre-determined biosignature indicative of anus squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 3**;

iii. a pre-determined biosignature indicative of appendix adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 4**;

iv. a pre-determined biosignature indicative of appendix mucinous adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 5**;

v. a pre-determined biosignature indicative of bile duct NOS cholangiocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 6**;

vi. a pre-determined biosignature indicative of brain astrocytoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 7**;

vii. a pre-determined biosignature indicative of brain astrocytoma anaplastic origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 8**;

viii. a pre-determined biosignature indicative of breast adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 9**;

ix. a pre-determined biosignature indicative of breast carcinoma NOS consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 10**;

x. a pre-determined biosignature indicative of breast infiltrating duct adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 11**;

xi. a pre-determined biosignature indicative of breast infiltrating lobular adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 12**;

xii. a pre-determined biosignature indicative of breast metaplastic carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 13**;

xiii. a pre-determined biosignature indicative of cervix adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 14**;

xiv. a pre-determined biosignature indicative of cervix carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 15**;

xv. a pre-determined biosignature indicative of cervix squamous carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 16**;

xvi. a pre-determined biosignature indicative of colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 17**;

xvii. a pre-determined biosignature indicative of colon carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 18**;

xviii. a pre-determined biosignature indicative of colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 19**;

xix. a pre-determined biosignature indicative of conjunctiva malignant melanoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 20**;

xx. a pre-determined biosignature indicative of duodenum and ampulla adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 21**;

xxi. a pre-determined biosignature indicative of endometrial endometrioid adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5,

6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 22**;

xxii. a pre-determined biosignature indicative of endometrial adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 23**;

xxiii. a pre-determined biosignature indicative of endometrial carcinosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 24**;

xxiv. a pre-determined biosignature indicative of endometrial serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 25**;

xxv. a pre-determined biosignature indicative of endometrium carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 26**;

xxvi. a pre-determined biosignature indicative of endometrium carcinoma undifferentiated origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 27**;

xxvii. a pre-determined biosignature indicative of endometrium clear cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 28**;

xxviii. a pre-determined biosignature indicative of esophagus adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 29**;

xxix. a pre-determined biosignature indicative of esophagus carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 30**;

xxx. a pre-determined biosignature indicative of esophagus squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 31**;

xxxi. a pre-determined biosignature indicative of extrahepatic cholangio common bile gallbladder adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 32**;

xxxii. a pre-determined biosignature indicative of fallopian tube adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 33**;

xxxiii. a pre-determined biosignature indicative of fallopian tube carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 34**;

xxxiv. a pre-determined biosignature indicative of fallopian tube carcinosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 35**;

xxxv. a pre-determined biosignature indicative of fallopian tube serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 36**;

xxxvi. a pre-determined biosignature indicative of gastric adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 37**;

xxxvii. a pre-determined biosignature indicative of gastroesophageal junction adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 38**;

xxxviii. a pre-determined biosignature indicative of glioblastoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 39**;

xxxix. a pre-determined biosignature indicative of glioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 40**;

xl. a pre-determined biosignature indicative of gliosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 41**;

xli. a pre-determined biosignature indicative of head, face or neck NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 42**;

xlii. a pre-determined biosignature indicative of intrahepatic bile duct cholangiocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 43**;

xliii. a pre-determined biosignature indicative of kidney carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 44**;

xliv. a pre-determined biosignature indicative of kidney clear cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 45**;

xlv. a pre-determined biosignature indicative of kidney papillary renal cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 46**;

xlvi. a pre-determined biosignature indicative of kidney renal cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 47**;

xlvii. a pre-determined biosignature indicative of larynx NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 48**;

xlviii. a pre-determined biosignature indicative of left colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 49**;

xlix. a pre-determined biosignature indicative of left colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 50**;

l. a pre-determined biosignature indicative of liver hepatocellular carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 51**;

li. a pre-determined biosignature indicative of lung adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 52**;

lii. a pre-determined biosignature indicative of lung adenosquamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 53**;

liii. a pre-determined biosignature indicative of lung carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 54**;

liv. a pre-determined biosignature indicative of lung mucinous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 55**;

lv. a pre-determined biosignature indicative of lung neuroendocrine carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 56**;

lvi. a pre-determined biosignature indicative of lung non-small cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 57**;

lvii. a pre-determined biosignature indicative of lung sarcomatoid carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 58**;

lviii. a pre-determined biosignature indicative of lung small cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 59**;

lix. a pre-determined biosignature indicative of lung squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 60**;

lx. a pre-determined biosignature indicative of meninges meningioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 61**;

lxi. a pre-determined biosignature indicative of nasopharynx NOS squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 62**;

lxii. a pre-determined biosignature indicative of oligodendroglioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 63**;

lxiii. a pre-determined biosignature indicative of oligodendroglioma aplastic origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 64**;

lxiv. a pre-determined biosignature indicative of ovary adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 65**;

lxv. a pre-determined biosignature indicative of ovary carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 66**;

lxvi. a pre-determined biosignature indicative of ovary carcinosarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 67**;

lxvii. a pre-determined biosignature indicative of ovary clear cell carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 68**;

lxviii. a pre-determined biosignature indicative of ovary endometrioid adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 69**;

lxix. a pre-determined biosignature indicative of ovary granulosa cell tumor NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 70**;

lxx. a pre-determined biosignature indicative of ovary high-grade serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 71**;

lxxi. a pre-determined biosignature indicative of ovary low-grade serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 72**;

lxxii. a pre-determined biosignature indicative of ovary mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 73**;

lxxiii. a pre-determined biosignature indicative of ovary serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 74**;

lxxiv. a pre-determined biosignature indicative of pancreas adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 75**;

lxxv. a pre-determined biosignature indicative of pancreas carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 76**;

lxxvi. a pre-determined biosignature indicative of pancreas mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 77**;

lxxvii. a pre-determined biosignature indicative of pancreas neuroendocrine carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 78**;

lxxviii. a pre-determined biosignature indicative of parotid gland carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 79**;

lxxix. a pre-determined biosignature indicative of peritoneum adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 80**;

lxxx. a pre-determined biosignature indicative of peritoneum carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 81**;

lxxxi. a pre-determined biosignature indicative of peritoneum serous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 82**;

lxxxii. a pre-determined biosignature indicative of pleural mesothelioma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 83**;

lxxxiii. a pre-determined biosignature indicative of prostate adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 84**;

lxxxiv. a pre-determined biosignature indicative of rectosigmoid adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 85**;

lxxxv. a pre-determined biosignature indicative of rectum adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 86**;

lxxxvi. a pre-determined biosignature indicative of rectum mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 87**;

lxxxvii. a pre-determined biosignature indicative of retroperitoneum dedifferentiated liposarcoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 88**;

lxxxviii. a pre-determined biosignature indicative of retroperitoneum leiomyosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 89**;

lxxxix. a pre-determined biosignature indicative of right colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 90**;

xc. a pre-determined biosignature indicative of right colon mucinous adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 91**;

xci. a pre-determined biosignature indicative of salivary gland adenoidcystic carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 92**;

xcii. a pre-determined biosignature indicative of skin Merkel cell carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 93**;

xciii. a pre-determined biosignature indicative of skin nodular melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 94**;

xciv. a pre-determined biosignature indicative of skin squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 95**;

xcv. a pre-determined biosignature indicative of skin melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 96**;

xcvi. a pre-determined biosignature indicative of small intestine gastrointestinal stromal tumor (GIST) NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 97**;

xcvii. a pre-determined biosignature indicative of small intestine adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 98**;

xcviii. a pre-determined biosignature indicative of stomach gastrointestinal stromal tumor (GIST) NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 99**;

xcix. a pre-determined biosignature indicative of stomach signet ring cell adenocarcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 100**;

c. a pre-determined biosignature indicative of thyroid carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 101**;

ci. a pre-determined biosignature indicative of thyroid carcinoma anaplastic NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 102**;

cii. a pre-determined biosignature indicative of papillary carcinoma of thyroid origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 103**;

ciii. a pre-determined biosignature indicative of tonsil oropharynx tongue squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 104**;

civ. a pre-determined biosignature indicative of transverse colon adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,

35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 105**;

cv. a pre-determined biosignature indicative of urothelial bladder adenocarcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 106**;

cvi. a pre-determined biosignature indicative of urothelial bladder carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 107**;

cvii. a pre-determined biosignature indicative of urothelial bladder squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 108**;

cviii. a pre-determined biosignature indicative of urothelial carcinoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 109**;

cix. a pre-determined biosignature indicative of uterine endometrial stromal sarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 110**;

cx. a pre-determined biosignature indicative of uterus leiomyosarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 111**;

cxi. a pre-determined biosignature indicative of uterus sarcoma NOS origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 112**;

cxii. a pre-determined biosignature indicative of uveal melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 113**;

cxiii. a pre-determined biosignature indicative of vaginal squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 114**;

cxiv. a pre-determined biosignature indicative of vulvar squamous carcinoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 115**; and/or

cxv. a pre-determined biosignature indicative of skin trunk melanoma origin consisting of, comprising, or comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 features selected from **Table 116.**

60. The method of claim **58** or **59**, wherein the selections of biomarkers according to any one of **Tables 2-116** comprises:

i. the top 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the feature biomarkers with the highest Importance value in the corresponding table/s;

ii. the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 feature biomarkers with the highest Importance value in the corresponding table/s;

iii. at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 40%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,

18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 feature biomarkers with the highest Importance value in the corresponding table/s; and/or

iv. at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the top 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, or 100 feature biomarkers with the highest Importance value in the corresponding table.

61. The method of any one of claims **37-60**, wherein:

i. step (b) comprises determining a gene copy number for at least one member of the biosignature, and step (d) comprises processing the gene copy number;

ii. step (b) comprises determining a sequence for at least one member of the biosignature, and step (d) comprises processing the sequence;

iii. step (b) comprises determining a sequence for a plurality of members of the biosignature, and step (d) comprises comparing the sequence to a reference sequence (e.g., wild type) to identify microsatellite repeats, and identifying members of the biosignature that have microsatellite instability (MSI);

iv. step (b) comprises determining a sequence for a plurality of members of the biosignature, and step (d) comprises comparing the sequence to a reference sequence (e.g., wild type) to identify a tumor mutational burden (TMB); and/or

v. step (b) comprises determining an mRNA transcript level for at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or at least 50 genes in any one of **Tables 117-120**, and/or INSM1, and step (d) comprises processing the transcript levels.

62. The method of claim **61**, wherein a gene copy number, CNV or CNA of a gene in the biosignature is determined by measuring the copy number of at least one proximate region to the gene, wherein optionally the proximate region comprises at least one location in the same sub-band, band, or arm of the chromosome wherein the gene is located.

63. The method of any one of claims **49-62**, wherein the one or more biomarkers in the biosignature are assessed as described in their corresponding table.

64. The method of any one of claims **37-63**, wherein the model comprises a plurality of intermediate models, wherein the plurality of intermediate models comprises at least one pairwise comparison module and/or at least one multi-class classification model.

65. The method of any one of claims **37-64**, wherein the model calculates a statistical measure that the biosignature corresponds to at least one of the at least one pre-determined biosignatures.

66. The method of claim **65**, wherein the processing in step (d) comprises:

i. a pairwise comparison between candidate pre-determined biosignatures, and a probability is calculated that the biosignature corresponds to either one of the pairs of the at least one pre-determined biosignatures; and/or

ii. using at least one multi-class classification model to assess the biosignature.

67. The method of claim **66**, wherein the pairwise comparison between the two candidate primary tumor origins in claim **66**.i) and/or the multi-class classification model in claim **66**.ii) is determined using a machine learning classification algorithm, wherein optionally the machine learning classification algorithm comprises a boosted tree.

68. The method of claim **66** or **67**, wherein the pairwise comparison between the two candidate primary tumor origins in claim **66**.i) is applied to at least one pre-determined biosignature according to any one of claims **58-60**; and/or the multi-class classification model in claim **66**.ii) is applied to at least one pre-determined biosignature according to any one of claims **49-57**.

69. The method of any one of claims **64-68**, further comprising determining intermediate model predictions, wherein the intermediate model predictions comprise:

i. a cancer type determined by the joint pairwise comparisons between at least one pair of pre-determined biosignatures according to any one of claims **58-59**;

ii. a cancer/disease type determined by an intermediate multi-class model applied to at least one pre-determined biosignature according to claim **49**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28 of the pre-determined biosignatures according to claim **49**;

iii. an organ group type determined by an intermediate multi-class model applied to at least one pre-determined biosignature according to claim **50**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, or 27 of the pre-determined biosignatures according to claim **50**; and/or

iv. a histology determined by an intermediate multi-class model applied to at least one pre-determined biosignature according to claim **51**, wherein optionally the intermediate multi-class model is applied to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, or 29 of the pre-determined biosignatures according to claim **51**.

70. The method of claim 69, wherein the processing in step (d) comprises inputting the outputs of each of 69 i)-iv) into a final predictor model that provides the prediction in step (e), wherein optionally the final predictor model comprises a machine learning algorithm, wherein optionally the machine learning algorithm comprises a boosted tree.

71. The method of claim 70, wherein the predicted at least one attribute of the cancer comprises at least one of adrenal cortical carcinoma; anus squamous carcinoma; appendix adenocarcinoma, NOS; appendix mucinous adenocarcinoma; bile duct, NOS, cholangiocarcinoma; brain astrocytoma, anaplastic; brain astrocytoma, NOS; breast adenocarcinoma, NOS; breast carcinoma, NOS; breast infiltrating duct adenocarcinoma; breast infiltrating lobular carcinoma, NOS; breast metaplastic carcinoma, NOS; cervix adenocarcinoma, NOS; cervix carcinoma, NOS; cervix squamous carcinoma; colon adenocarcinoma, NOS; colon carcinoma, NOS; colon mucinous adenocarcinoma; conjunctiva malignant melanoma, NOS; duodenum and ampulla adenocarcinoma, NOS; endometrial adenocarcinoma, NOS; endometrial carcinosarcoma; endometrial endometrioid adenocarcinoma; endometrial serous carcinoma; endometrium carcinoma, NOS; endometrium carcinoma, undifferentiated; endometrium clear cell carcinoma; esophagus adenocarcinoma, NOS; esophagus carcinoma, NOS; esophagus squamous carcinoma; extrahepatic cholangio, common bile, gallbladder adenocarcinoma, NOS; fallopian tube adenocarcinoma, NOS; fallopian tube carcinoma, NOS; fallopian tube carcinosarcoma, NOS; fallopian tube serous carcinoma; gastric adenocarcinoma; gastroesophageal junction adenocarcinoma, NOS; glioblastoma; glioma, NOS; gliosarcoma; head, face or neck, NOS squamous carcinoma; intrahepatic bile duct cholangiocarcinoma; kidney carcinoma, NOS; kidney clear cell carcinoma; kidney papillary renal cell carcinoma; kidney renal cell carcinoma, NOS; larynx, NOS squamous carcinoma; left colon adenocarcinoma, NOS; left colon mucinous adenocarcinoma; liver hepatocellular carcinoma, NOS; lung adenocarcinoma, NOS; lung adenosquamous carcinoma; lung carcinoma, NOS; lung mucinous adenocarcinoma; lung neuroendocrine carcinoma, NOS; lung non-small cell carcinoma; lung sarcomatoid carcinoma; lung small cell carcinoma, NOS; lung squamous carcinoma; meninges meningioma, NOS; nasopharynx, NOS squamous carcinoma; oligodendroglioma, anaplastic; oligodendroglioma, NOS; ovary adenocarcinoma, NOS; ovary carcinoma, NOS; ovary carcinosarcoma; ovary clear cell carcinoma; ovary endometrioid adenocarcinoma; ovary granulosa cell tumor, NOS; ovary high-grade serous carcinoma; ovary low-grade serous carcinoma; ovary mucinous adenocarcinoma; ovary serous carcinoma; pancreas adenocarcinoma, NOS; pancreas carcinoma, NOS; pancreas mucinous adenocarcinoma; pancreas neuroendocrine carcinoma, NOS; parotid gland carcinoma, NOS; peritoneum adenocarcinoma, NOS; peritoneum carcinoma, NOS; peritoneum serous carcinoma; pleural mesothelioma, NOS; prostate adenocarcinoma, NOS; rectosigmoid adenocarcinoma, NOS; rectum adenocarcinoma, NOS; rectum mucinous adenocarcinoma; retroperitoneum dedifferentiated liposarcoma; retroperitoneum leiomyosarcoma, NOS; right colon adenocarcinoma, NOS; right colon

mucinous adenocarcinoma; salivary gland adenoid cystic carcinoma; skin melanoma; skin melanoma; skin merkel cell carcinoma; skin nodular melanoma; skin squamous carcinoma; skin trunk melanoma; small intestine adenocarcinoma; small intestine gastrointestinal stromal tumor, NOS; stomach gastrointestinal stromal tumor, NOS; stomach signet ring cell adenocarcinoma; thyroid carcinoma, anaplastic, NOS; thyroid carcinoma, NOS; thyroid papillary carcinoma of thyroid; tonsil, oropharynx, tongue squamous carcinoma; transverse colon adenocarcinoma, NOS; urothelial bladder adenocarcinoma, NOS; urothelial bladder carcinoma, NOS; urothelial bladder squamous carcinoma; urothelial carcinoma, NOS; uterine endometrial stromal sarcoma, NOS; uterus leiomyosarcoma, NOS; uterus sarcoma, NOS; uveal melanoma; vaginal squamous carcinoma; vulvar squamous carcinoma; and any combination thereof.

72. The method of claim **70**, wherein the predicted at least one attribute of the cancer comprises at least one of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, and uterine sarcoma.

73. The method of claim **70**, wherein the predicted at least one attribute of the cancer comprises at least one of bladder; skin; lung; head, face or neck (NOS); esophagus; female genital tract (FGT); brain; colon; prostate; liver, gall bladder, ducts; breast; eye; stomach; kidney; and pancreas.

74. The method of claim **70**, wherein the predicted at least one attribute of the cancer cancer is according to at least one attribute listed in claim **48**.

75. The method of any one of claims **37-74**, wherein the sample comprises a cancer of unknown primary (CUP).

76. A method of predicting at least one attribute of a cancer, the method comprising:

(a) obtaining a biological sample from a subject having a cancer, wherein the biological sample is according to any one of claims **38-41**;

(b) performing at least one assay to assess one or more biomarkers in the biological sample to obtain a biosignature for the sample, wherein performing the at least one assay is according to any one of claims **42-46**;

(c) providing the biosignature into a model that has been trained to predict at least one attribute of the cancer, wherein the model comprises at least one intermediate model, wherein the at least one intermediate model comprises:

(1) an first intermediate model trained to process DNA data using the predetermined biosignatures according to claim **59**;

(2) a second intermediate model trained to process RNA data using the predetermined biosignatures according to claim **49**;

(3) a third intermediate model trained to process RNA data using the predetermined biosignatures according to claim **50**; and/or

(4) a fourth intermediate model trained to process RNA data using the predetermined biosignatures according to claim **51**;

(d) processing, by one or more computers, the provided biosignature through each of the plurality of intermediate models in part (c), providing the output of each of the plurality of intermediate models into a final predictor model, and processing by one or more computers, the output of each of the plurality of intermediate models through the final predictor model; and

(e) outputting from the final predictor model a prediction of the at least one attribute of the cancer; wherein the predicted at least one attribute of the cancer is a tissue-of-origin selected from the group consisting of breast adenocarcinoma, central nervous system cancer, cervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, gastroesophageal adenocarcinoma, gastrointestinal stromal tumor (GIST), hepatocellular carcinoma, lung adenocarcinoma, melanoma, meningioma, ovarian granulosa cell tumor, ovarian & fallopian tube adenocarcinoma, pancreas adenocarcinoma, prostate adenocarcinoma, renal cell carcinoma, squamous cell carcinoma, thyroid cancer, urothelial carcinoma, uterine endometrial adenocarcinoma, uterine sarcoma, and a combination thereof.

77. The method of claim **76**, wherein step (b) comprises performing DNA analysis by sequencing genomic DNA from the biological sample, wherein the DNA analysis is performed for the genes in **Tables 2-116**; and performing RNA analysis by sequencing messenger RNA transcripts from the biological sample, wherein the RNA analysis is performed for the genes in **Table 117** or **Tables 118-120**.

78. The method of claim **76** or **77**, wherein at least one of the at least one intermediate model and final predictor model comprises a machine learning module, wherein optionally the machine learning module comprises one or more of a random forest, support vector machine, logistic regression, K-nearest neighbor, artificial neural network, naïve Bayes, quadratic discriminant analysis, and Gaussian processes models, wherein optionally the machine learning module comprises an XGBoost decision-tree-based ensemble machine learning algorithm.

79. The method of any one of claims **37-78**, wherein the prediction of the at least one attribute of the cancer is used to:

i. confirm a diagnosis;

ii. change a diagnosis;

iii. perform a quality check; and/or

iv. indicate additional molecular testing to be performed.

80. The method of any one of claims **37-79**, wherein the predicted at least one attribute comprises an ordered list, wherein optionally the list is ordered using a statistical measure.

81. The method of any one of claims **37-80**, further comprising determining whether the prediction of the at least one attribute meets a threshold level, wherein optionally the threshold level is related to a probability of the prediction and/or a confidence in the prediction.

82. The method of any one of claims **37-81**, further comprising generating a molecular profile that identifies the presence, level, or state of the biomarkers in the biosignature, e.g., whether each biomarker has a copy number alteration and/or mutation; and/or a TMB level, MSI, LOH, or MMR status; and/or expression level, wherein the expression level comprises that of at least one transcript and/or protein level.

83. The method of any one of claims **37-82**, further comprising selecting at least one treatment for the patient based at least in part upon the classified at least one attribute of the cancer, wherein optionally the treatment comprises administration of immunotherapy, chemotherapy, or a combination thereof.

84. A method comprising preparing a report, wherein the report comprises a summary or overview of the molecular profile generated according to claim **82**, wherein the report identifies the classified at least one attribute of the cancer, wherein optionally the report further identifies the at least one treatment selected according to claim **83**.

85. The method of claim **84**, wherein the report is computer generated, is a printed report and/or a computer file, and/or is accessible via a web portal.

86. A system comprising one or more computers and one or more storage media storing instructions that, when executed by the one or more computers, cause the one or more computers to perform operations described with reference to any one of claims **37-85**.

87. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform operations described with reference to claims **37-85**.

88. A system for identifying an attribute of a cancer, the system comprising:

(a) at least one host server;

(b) at least one user interface for accessing the at least one host server to access and input data;

(c) at least one processor for processing the inputted data;

(d) at least one memory coupled to the processor for storing the processed data and instructions for carrying out operations with respect to any one of claims **37-85**; and

(e) at least one display for displaying the identified attribute of the cancer.

89. The system of claim **88**, further comprising at least one memory coupled to the processor for storing the processed data and instructions for selecting and/or generating according to any one of claims **83-85**.

90. The system of claims **88** or **89**, wherein the at least one display comprises a report comprising the classified at least one attribute of the cancer.

91. A system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising:

one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising:

obtaining, by the system, a sample biological signature representing the sample that was obtained from the body, wherein the sample comprises cancer cells;

providing, by the system, the sample biological signature as an input to a model, wherein:

the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or

the model is a multi-class model wherein the classes comprise different attributes; and

receiving, by the system, an output generated by the model that represents data indicating a likely attribute of the sample obtained from the body based on the pairwise analysis.

92. A system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising:

one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising:

obtaining, by the system, a sample biological signature representing the sample that was obtained from the body;

providing, by the system, the sample biological signature as an input to a model, wherein:

the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or

the model is a multi-class model wherein the classes comprise different attributes; and

receiving, by the system, an output generated by the model that represents data indicating a probability that an attribute identified by the particular biological signature identifies a likely attribute of the sample.

93. A system for identifying at least one attribute of a sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising:

one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising:

obtaining, by the system, a sample biological signature representing a biological sample that was obtained from the cancer sample in a first portion of the body, wherein the sample biological signature includes data describing a plurality of features of the biological sample, wherein the plurality of features include data describing the first portion of the body;

providing, by the system, the sample biological signature as an input to a model, wherein:

the model is configured to perform analysis between the sample biological signature and each of multiple different biological signatures, wherein each of the multiple different biological signatures corresponds to a different attribute; and/or

the model is a multi-class model wherein the classes comprise different attributes; and

receiving, by the system, an output generated by the model that represents data indicating a likely attribute of the sample obtained from the body.

94. The system of any one of claims **91-93**, wherein the sample obtained from the body is a biological sample according to any one of claims **38-41**.

95. The system of any one of claims **91-94**, wherein the at least one attribute is an attribute listed in claim **48**.

96. The system of any one of claims **91-94**, wherein the sample biological signature includes data representing features obtained based on performance of an assay to assess one or more

biomarkers in the cancer sample, wherein optionally the assay is according to the at least one assay of any one of claims **42-46**.

97. The system of any one of claims **91-96**, the operations further comprising:

determining, based on the output generated by the model, a proposed cancer treatment.

98. The system of any one of claims **91-97**, wherein the at least one attribute is according to any one of claims **71-74**.

99. The system of any one of claims **91-98**, wherein each of the multiple different biological signatures comprise pre-identified biosignatures according to any one of claims **49-59**.

100. The system of any one of claims **91-99**, the operations further comprising:

receiving, by the system, an output generated by the model that represents a likelihood that the sample obtained from the body in a first portion of the body originated from a cancer in a second portion of the body.

101. The system of claim **100**, further comprising

determining, by the system and based on the received output, whether the received output generated by the model satisfies one or more predetermined thresholds; and

based on the determining, by the system, that the received output satisfies the one or more predetermined thresholds, determining, by the system, that the cancerous neoplasm in the first portion of the body originated from a cancer in a second portion of the body or that the cancerous neoplasm in the first portion of the body did not originate from a cancer in a second portion of the body.

102. The system of claim **100**,

wherein the received output generated by the model includes a matrix data structure,

wherein the matrix data structure includes a cell for each feature of the plurality of features evaluated by the pairwise model, wherein each of the cells includes data describing a probability that the corresponding feature indicates that the cancerous neoplasm in the first portion of the body was caused by cancer in the second portion of the first body.

103. A system for identifying at least one attribute of a cancer, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising:

one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising:

receiving, by the system storing a model that is configured to perform analysis of a biological signature, a sample biological signature representing a biological sample that was obtained from a cancerous neoplasm in a first portion of a body, wherein the model includes a

cancerous biological signature for each of multiple different types of cancerous biological samples, wherein the cancerous biological signatures include at least a first cancerous biological signature representing a molecular profile of a cancerous biological sample from the first portion of one or more other bodies;

performing, by the system and using the model, analysis of the sample biological signature using the cancerous biological signatures;

generating, by the system and based on the performed analysis, a likelihood that the cancerous neoplasm in the first portion of the body was caused by cancer in a second portion of the body;

providing, by the system, the generated likelihood to another device for display on the other device.

104. A system for training an analysis model for identifying at least one attribute of a cancer sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primary tumor origin, cancer/disease type, organ group, histology, and any combination thereof, the system comprising:

one or more processors and one or more memory units storing instructions that, when executed by the one or more processors, cause the one or more processors to perform operations, the operations comprising:

generating, by the system, an analysis model, wherein generating the analysis model includes generating a plurality of model signatures, wherein each model signature is configured to differentiate between at least one attribute within each of the at least one attribute;

obtaining, by the system, a set of training data items, wherein each training data item represents DNA or RNA sequencing results and includes data indicating (i) whether or not a variant was detected in the sequencing results and (ii) a number of copies of a gene or transcript in the sequencing results; and

training, by the system, an analysis model using the obtained set of training data items.

105. The system of claim **104**, wherein the plurality of model signatures are generated using random forest models, wherein optionally the random forest models comprise gradient boosted forests.

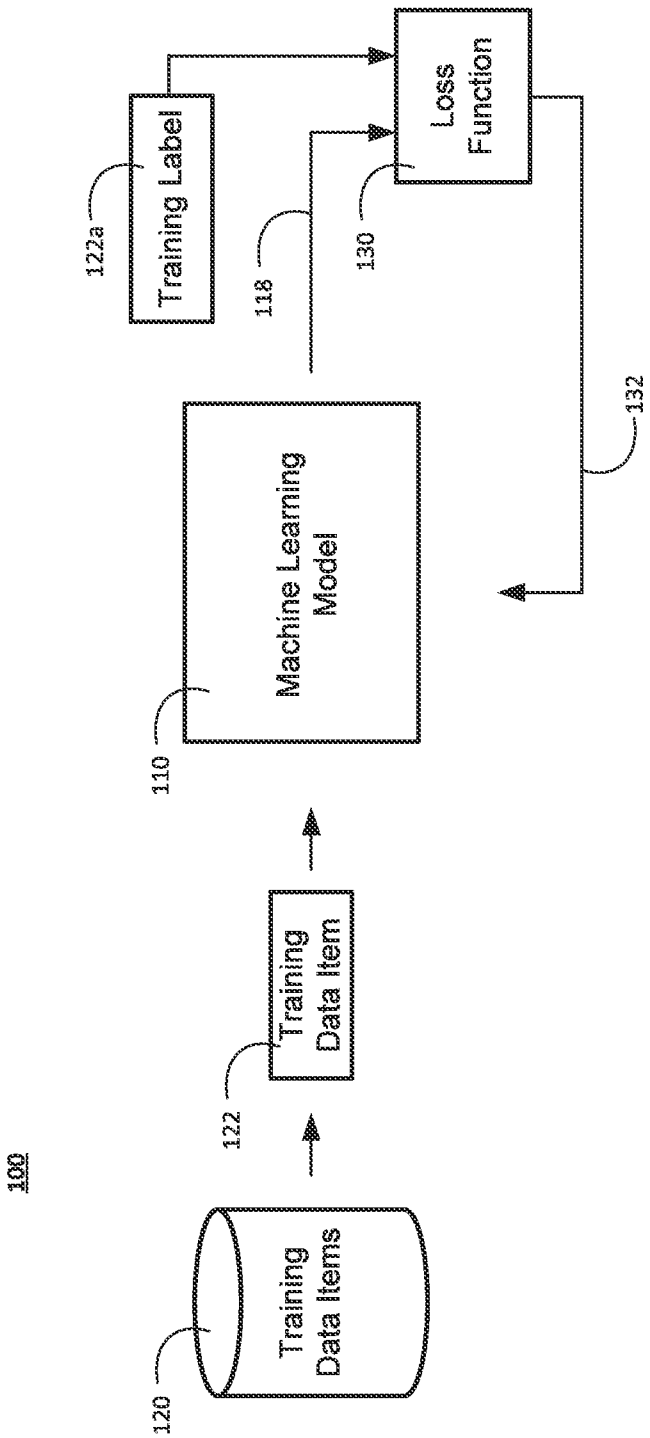

100
Training Data Items
120
Training Data Item
122
Machine Learning Model
110
Training Label
122a
Loss Function
130
118
132

FIG. 1A

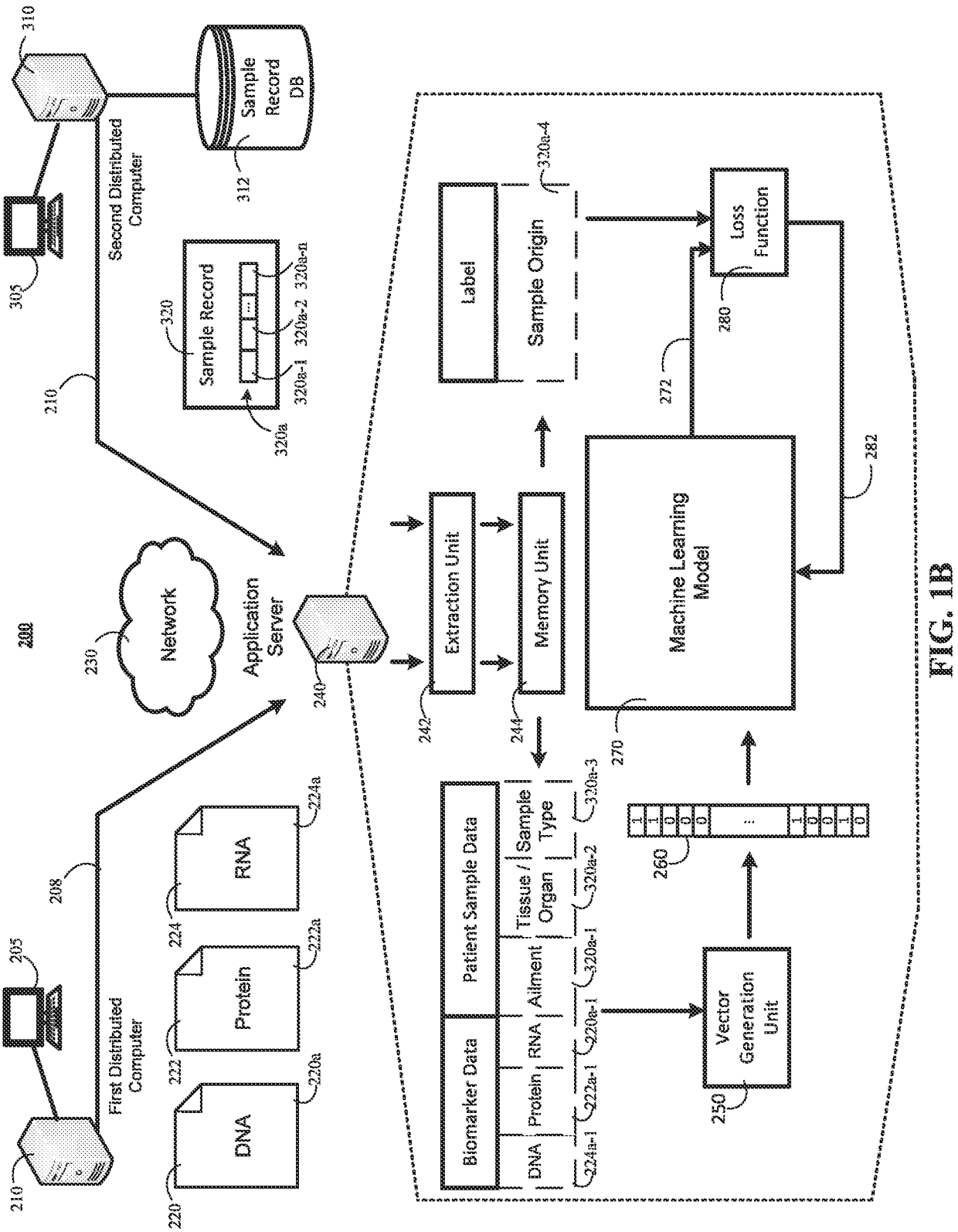

200
First Distributed Computer
Second Distributed Computer
Application Server
Network
Sample Record DB
Sample Record
DNA
Protein
RNA
Biomarker Data
Patient Sample Data
Extraction Unit
Memory Unit
Label
Sample Origin
Machine Learning Model
Loss Function
Vector Generation Unit
Ailment
Tissue / Organ
Sample Type

FIG. 1B

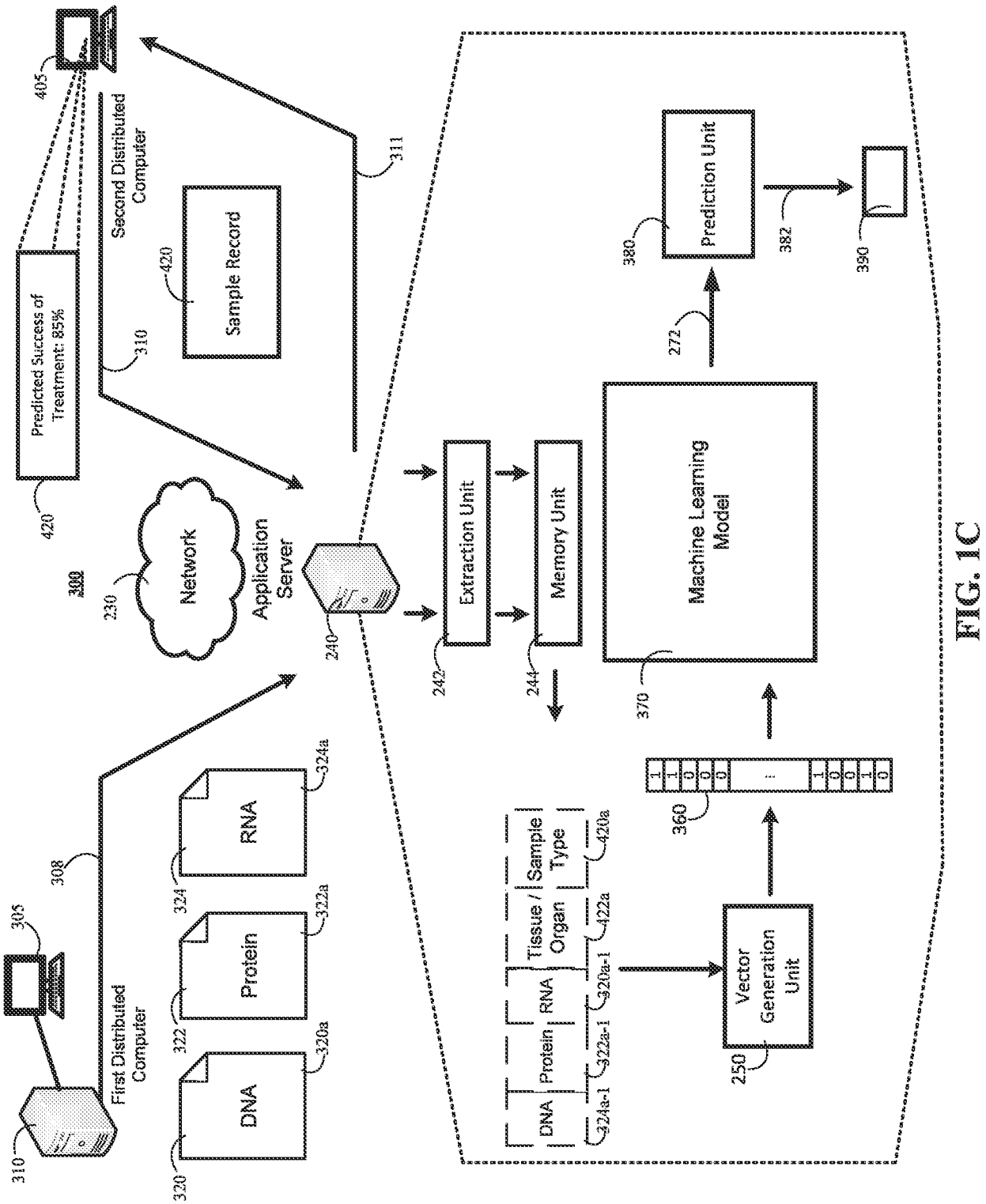

310
First Distributed Computer
305
308
DNA
320
320a
Protein
322
322a
RNA
324
324a
Network
230
Application Server
240
300
Predicted Success of Treatment: 85%
420
310
Second Distributed Computer
Sample Record
420
405
311
Extraction Unit
242
Memory Unit
244
DNA
Protein
RNA
Tissue / Organ
Sample Type
324a-1
322a-1
320a-1
422a
420a
Vector Generation Unit
250
360
Machine Learning Model
370
272
Prediction Unit
380
382
390

FIG. 1C

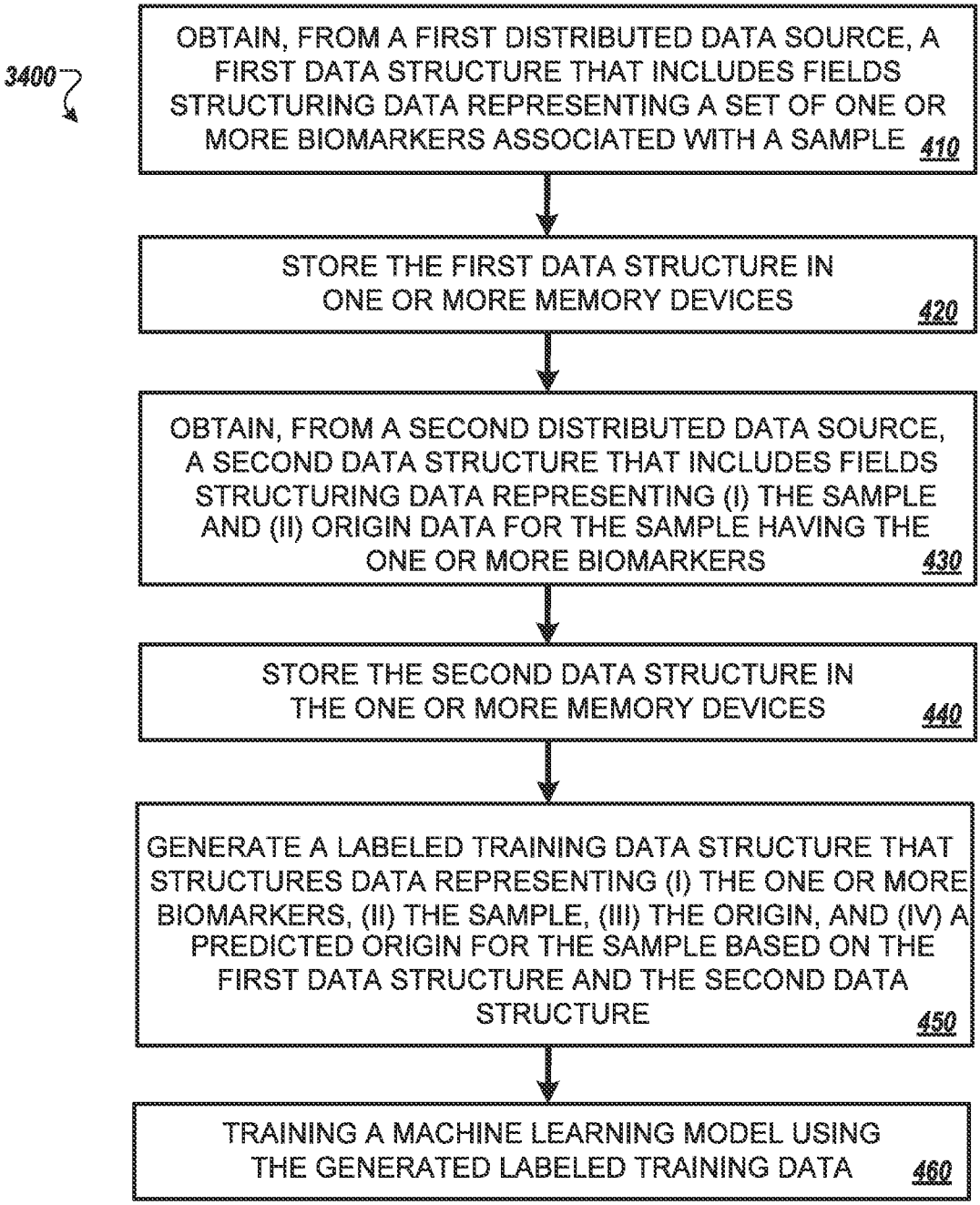

3400
OBTAIN, FROM A FIRST DISTRIBUTED DATA SOURCE, A FIRST DATA STRUCTURE THAT INCLUDES FIELDS STRUCTURING DATA REPRESENTING A SET OF ONE OR MORE BIOMARKERS ASSOCIATED WITH A SAMPLE 410
STORE THE FIRST DATA STRUCTURE IN ONE OR MORE MEMORY DEVICES 420
OBTAIN, FROM A SECOND DISTRIBUTED DATA SOURCE, A SECOND DATA STRUCTURE THAT INCLUDES FIELDS STRUCTURING DATA REPRESENTING (I) THE SAMPLE AND (II) ORIGIN DATA FOR THE SAMPLE HAVING THE ONE OR MORE BIOMARKERS 430
STORE THE SECOND DATA STRUCTURE IN THE ONE OR MORE MEMORY DEVICES 440
GENERATE A LABELED TRAINING DATA STRUCTURE THAT STRUCTURES DATA REPRESENTING (I) THE ONE OR MORE BIOMARKERS, (II) THE SAMPLE, (III) THE ORIGIN, AND (IV) A PREDICTED ORIGIN FOR THE SAMPLE BASED ON THE FIRST DATA STRUCTURE AND THE SECOND DATA STRUCTURE 450
TRAINING A MACHINE LEARNING MODEL USING THE GENERATED LABELED TRAINING DATA 460

FIG. 1D

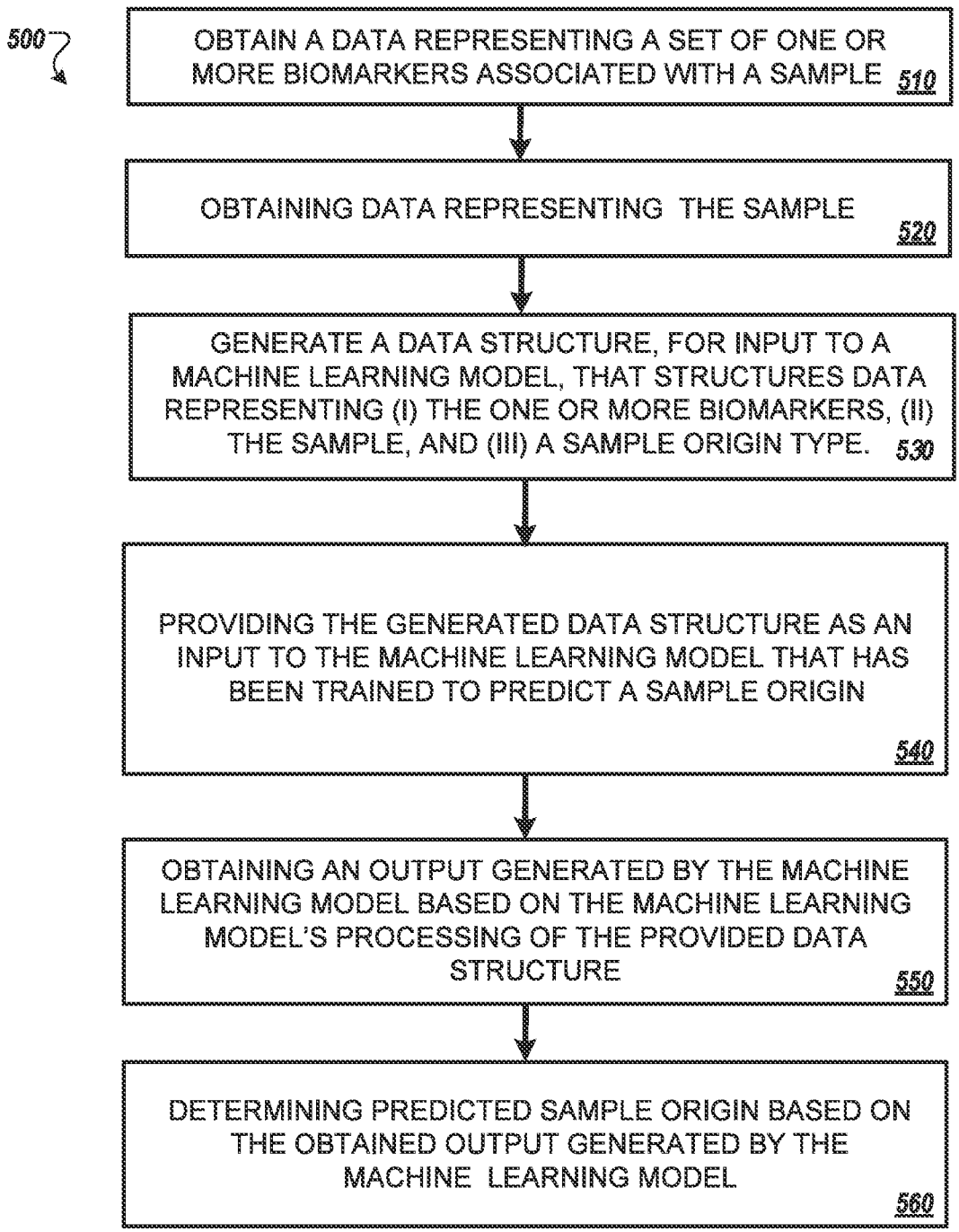

500
OBTAIN A DATA REPRESENTING A SET OF ONE OR MORE BIOMARKERS ASSOCIATED WITH A SAMPLE 510
OBTAINING DATA REPRESENTING THE SAMPLE 520
GENERATE A DATA STRUCTURE, FOR INPUT TO A MACHINE LEARNING MODEL, THAT STRUCTURES DATA REPRESENTING (I) THE ONE OR MORE BIOMARKERS, (II) THE SAMPLE, AND (III) A SAMPLE ORIGIN TYPE. 530
PROVIDING THE GENERATED DATA STRUCTURE AS AN INPUT TO THE MACHINE LEARNING MODEL THAT HAS BEEN TRAINED TO PREDICT A SAMPLE ORIGIN 540
OBTAINING AN OUTPUT GENERATED BY THE MACHINE LEARNING MODEL BASED ON THE MACHINE LEARNING MODEL'S PROCESSING OF THE PROVIDED DATA STRUCTURE 550
DETERMINING PREDICTED SAMPLE ORIGIN BASED ON THE OBTAINED OUTPUT GENERATED BY THE MACHINE LEARNING MODEL 560

FIG. 1E

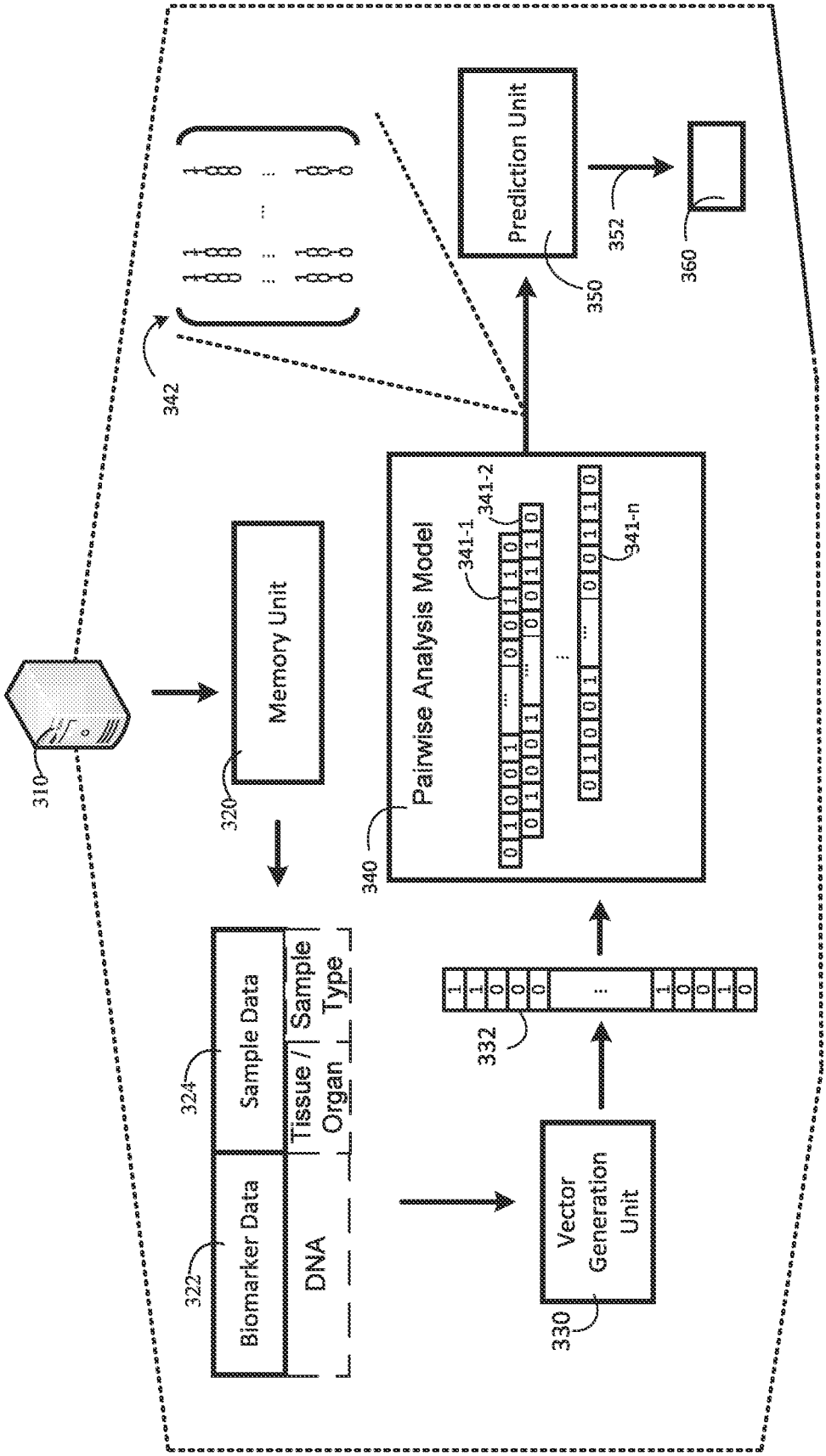

310
320
Memory Unit
340
Pairwise Analysis Model
341-1
341-2
341-n
342
350
Prediction Unit
352
360
322
Biomarker Data
DNA
324
Sample Data
Tissue / Organ
Sample Type
330
Vector Generation Unit
332

FIG. 1F

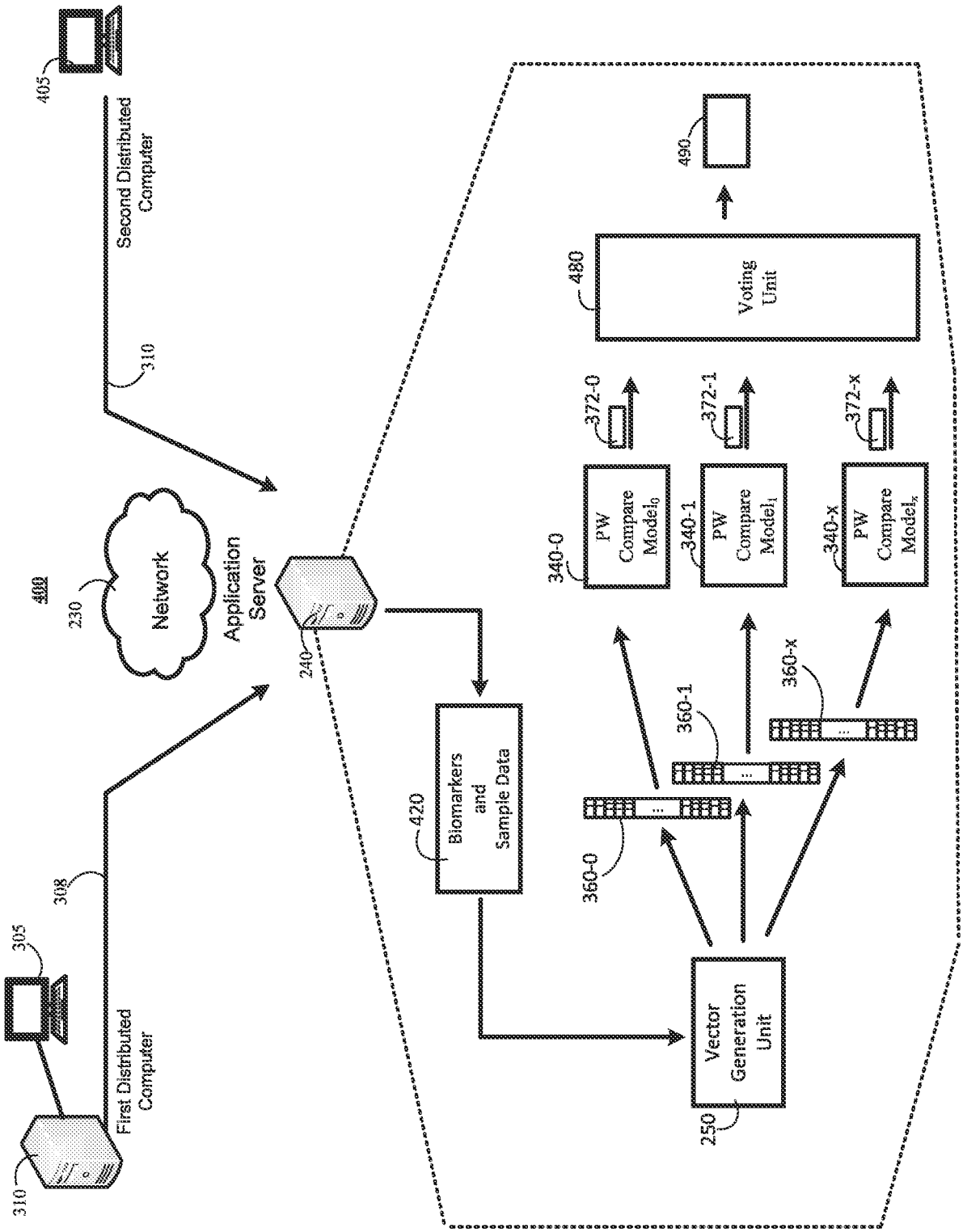

405
Second Distributed Computer
310
400
230
Network
Application Server
240
308
305
First Distributed Computer
310
420
Biomarkers and Sample Data
250
Vector Generation Unit
360-0
360-1
360-x
340-0
PW Compare Model0
340-1
PW Compare Model1
340-x
PW Compare Modelx
372-0
372-1
372-x
480
Voting Unit
490

FIG. 1G

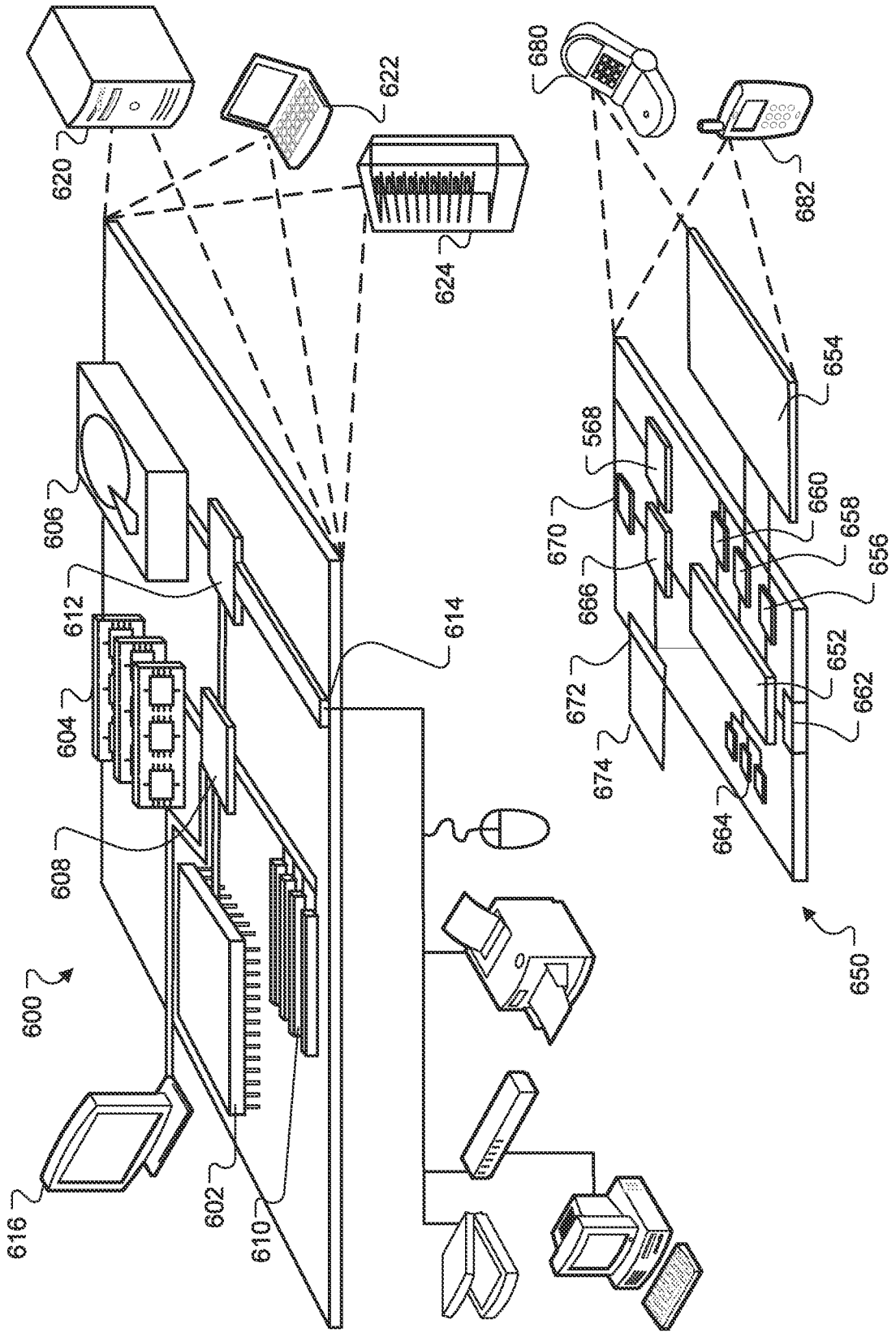

620
622
624
606
612
614
604
608
600
616
602
610
680
682
654
568
660
658
656
670
666
652
672
662
674
664
650

FIG. 1H

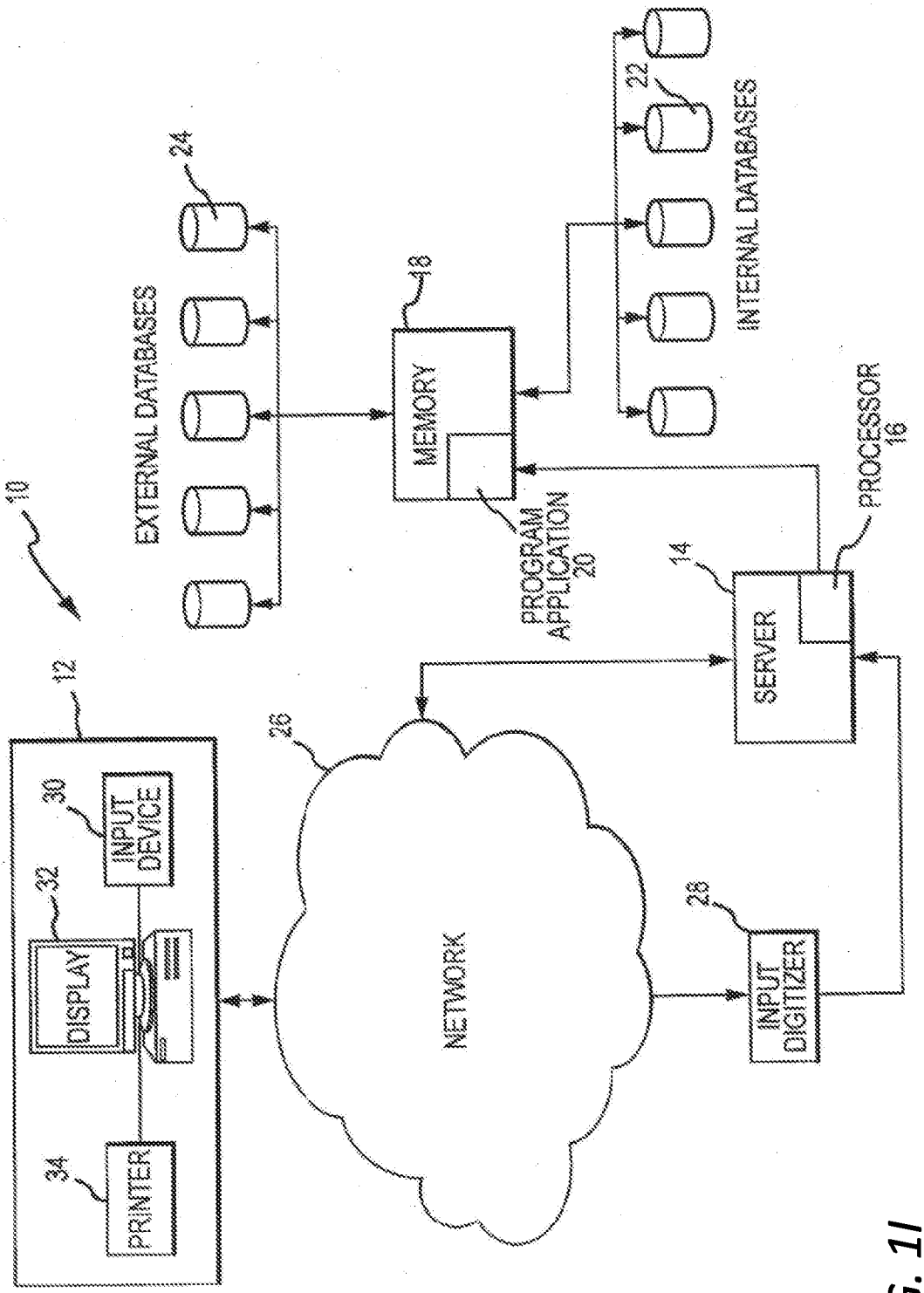

10
12
30
INPUT DEVICE
32
DISPLAY
34
PRINTER
26
NETWORK
28
INPUT DIGITIZER
14
SERVER
PROCESSOR 16
PROGRAM APPLICATION 20
18
MEMORY
EXTERNAL DATABASES
24
INTERNAL DATABASES
22

FIG. 1I

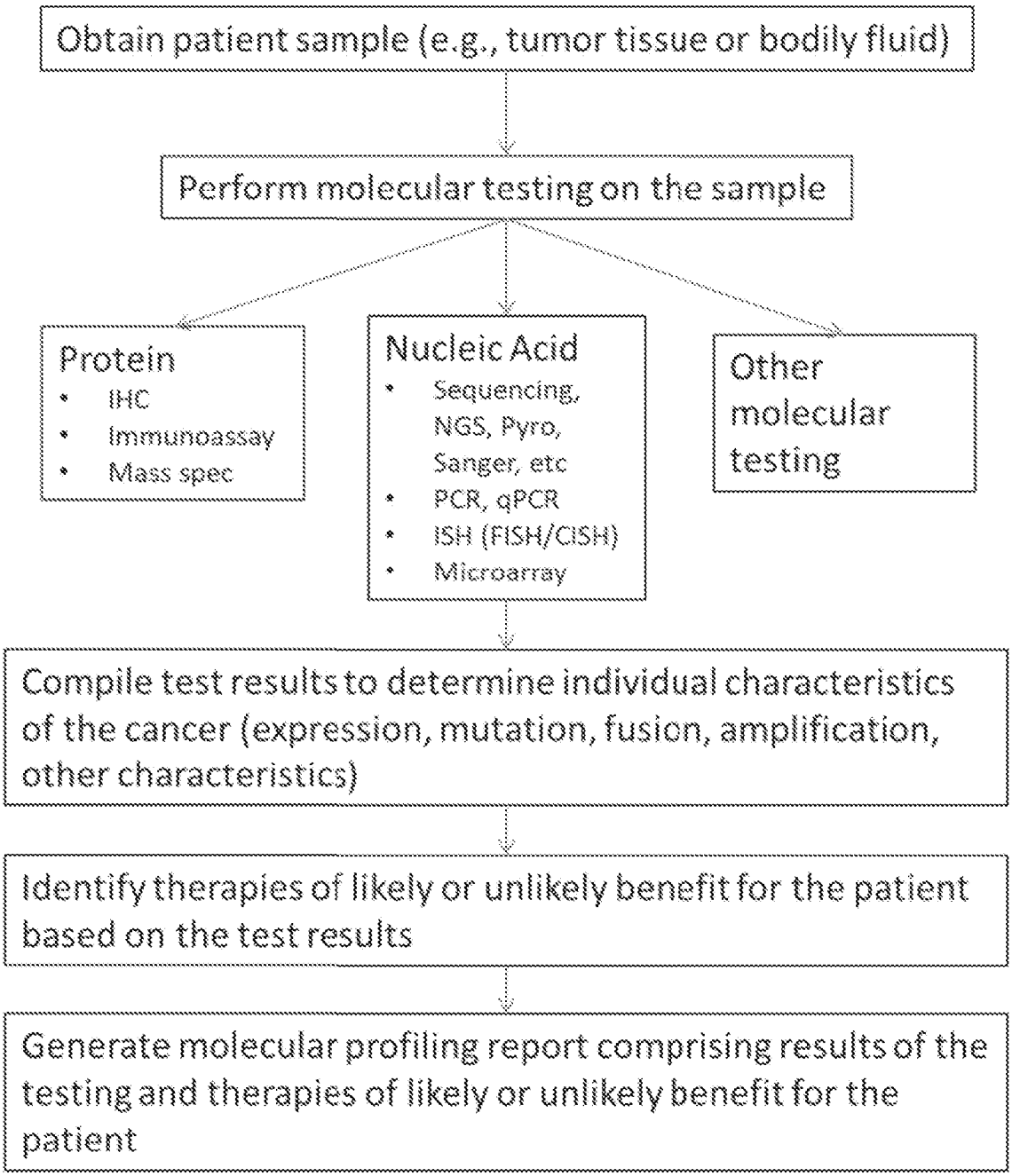

Obtain patient sample (e.g., tumor tissue or bodily fluid)
Perform molecular testing on the sample
Protein
IHC
Immunoassay
Mass spec
Nucleic Acid
Sequencing, NGS, Pyro, Sanger, etc
PCR, qPCR
ISH (FISH/CISH)
Microarray
Other molecular testing
Compile test results to determine individual characteristics of the cancer (expression, mutation, fusion, amplification, other characteristics)
Identify therapies of likely or unlikely benefit for the patient based on the test results
Generate molecular profiling report comprising results of the testing and therapies of likely or unlikely benefit for the patient

FIG. 2A

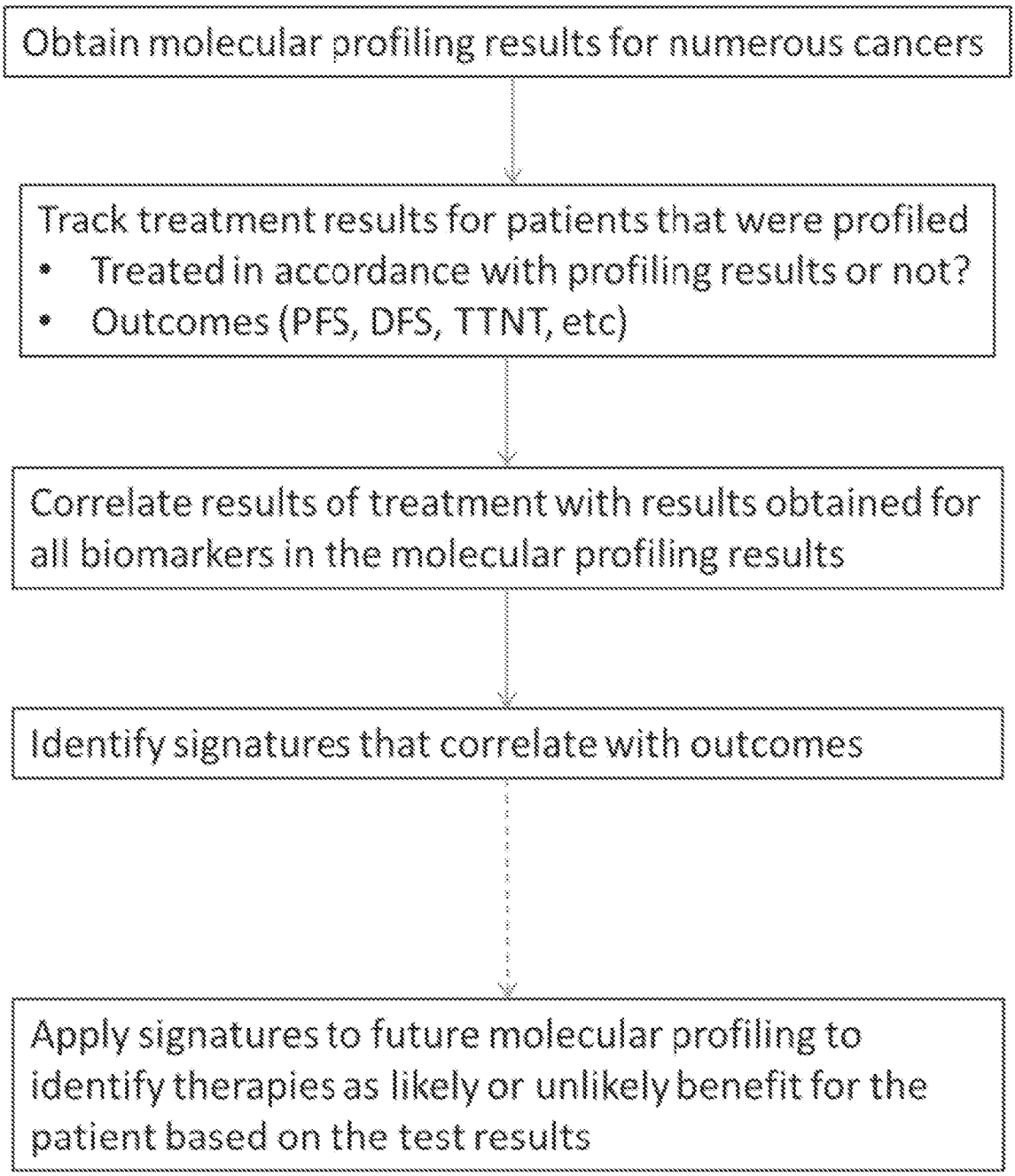

Obtain molecular profiling results for numerous cancers
Track treatment results for patients that were profiled
• Treated in accordance with profiling results or not?
• Outcomes (PFS, DFS, TTNT, etc)
Correlate results of treatment with results obtained for all biomarkers in the molecular profiling results
Identify signatures that correlate with outcomes
Apply signatures to future molecular profiling to identify therapies as likely or unlikely benefit for the patient based on the test results

FIG. 2B

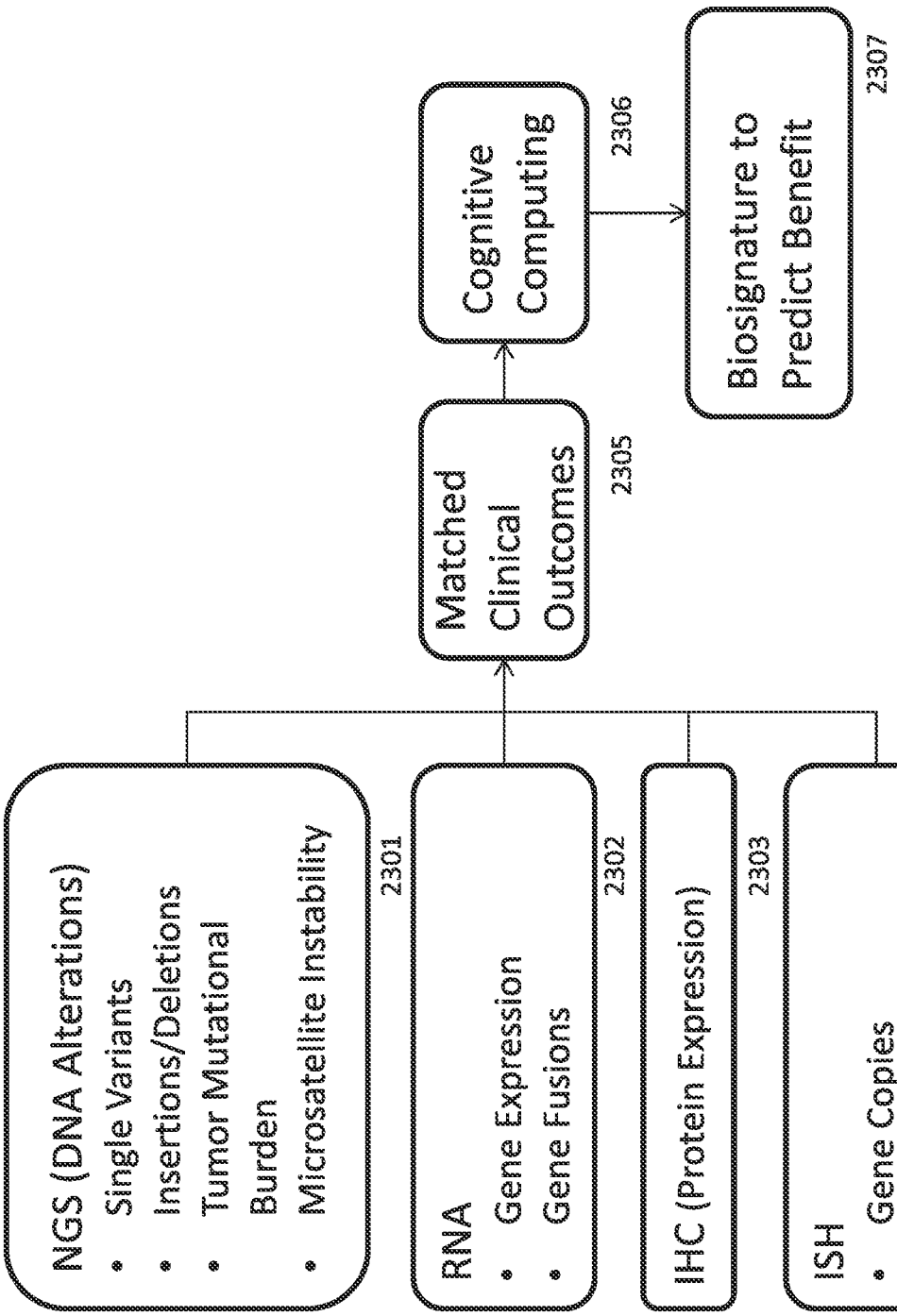

NGS (DNA Alterations)
• Single Variants
• Insertions/Deletions
• Tumor Mutational Burden
• Microsatellite Instability
2301
RNA
• Gene Expression
• Gene Fusions
2302
IHC (Protein Expression)
2303
ISH
• Gene Copies
• Gene Translocations
2304
Matched Clinical Outcomes
2305
Cognitive Computing
2306
Biosignature to Predict Benefit
2307

FIG. 2C

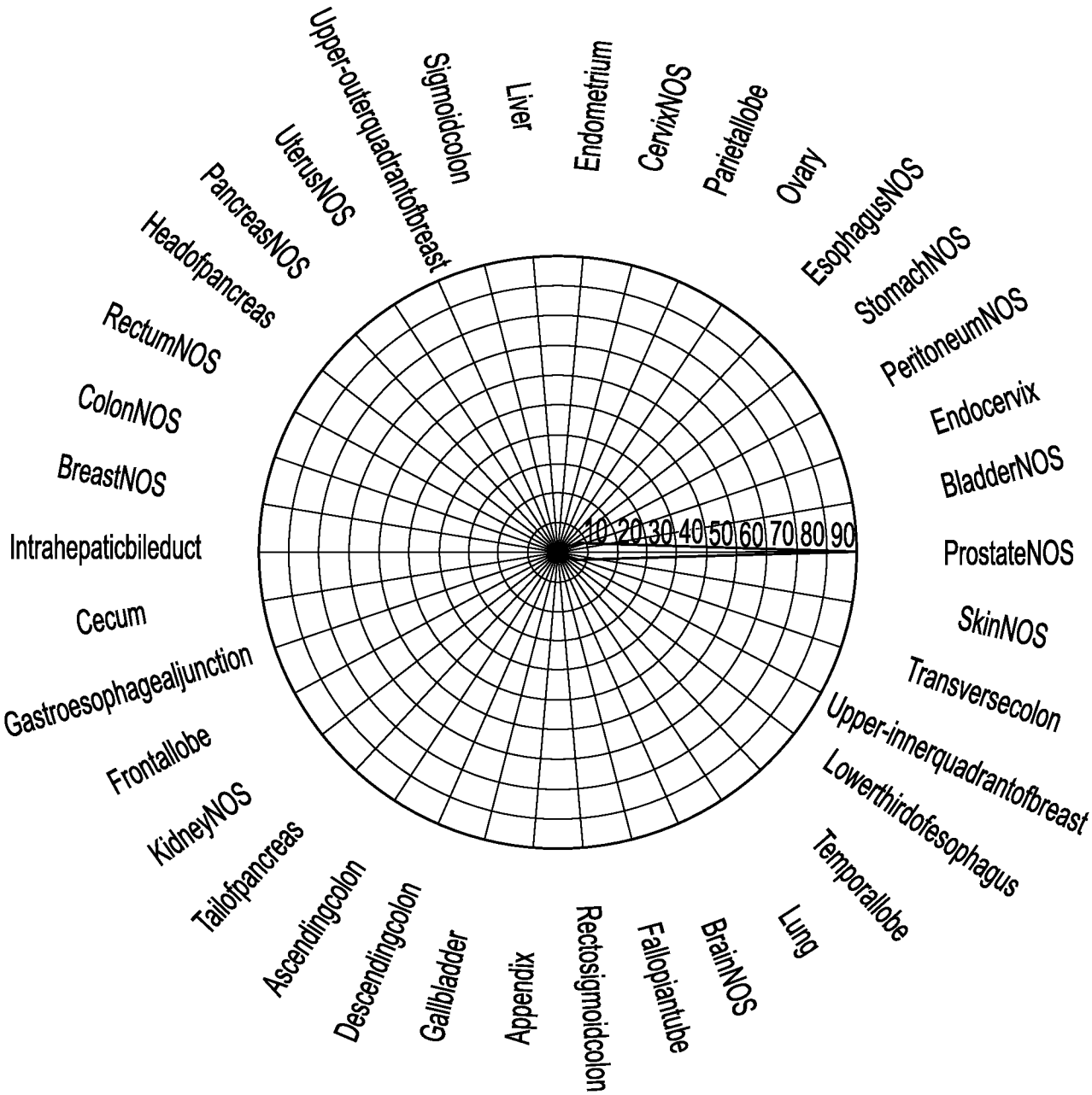

Upper-outerquadrantofbreast
Sigmoidcolon
Liver
Endometrium
CervixNOS
Parietallobe
Ovary
EsophagusNOS
StomachNOS
PeritoneumNOS
Endocervix
BladderNOS
ProstateNOS
SkinNOS
Transversecolon
Upper-innerquadrantofbreast
Lowerthirdofesophagus
Temporallobe
Lung
BrainNOS
Fallopiantube
Rectosigmoidcolon
Appendix
Gallbladder
Descendingcolon
Ascendingcolon
Tailofpancreas
KidneyNOS
Frontallobe
Gastroesophagealjunction
Cecum
Intrahepaticbileduct
BreastNOS
ColonNOS
RectumNOS
Headofpancreas
PancreasNOS
UterusNOS
10 20 30 40 50 60 70 80 90

FIG. 3A

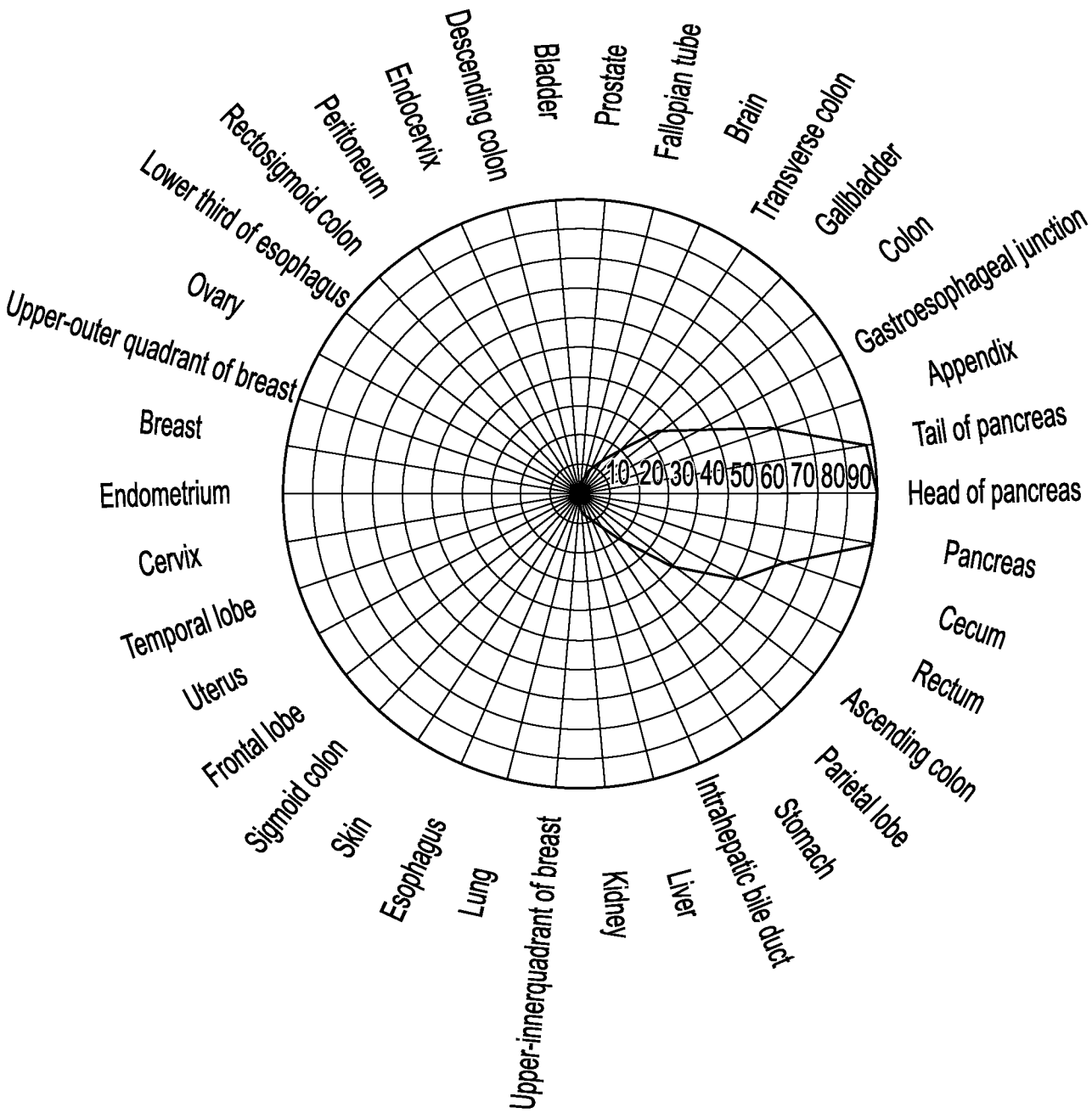

10 20 30 40 50 60 70 80 90
Bladder
Prostate
Fallopian tube
Brain
Transverse colon
Gallbladder
Colon
Gastroesophageal junction
Appendix
Tail of pancreas
Head of pancreas
Pancreas
Cecum
Rectum
Ascending colon
Parietal lobe
Stomach
Intrahepatic bile duct
Liver
Kidney
Upper-innerquadrant of breast
Lung
Esophagus
Skin
Sigmoid colon
Frontal lobe
Uterus
Temporal lobe
Cervix
Endometrium
Breast
Upper-outer quadrant of breast
Ovary
Lower third of esophagus
Rectosigmoid colon
Peritoneum
Endocervix
Descending colon

FIG. 3B

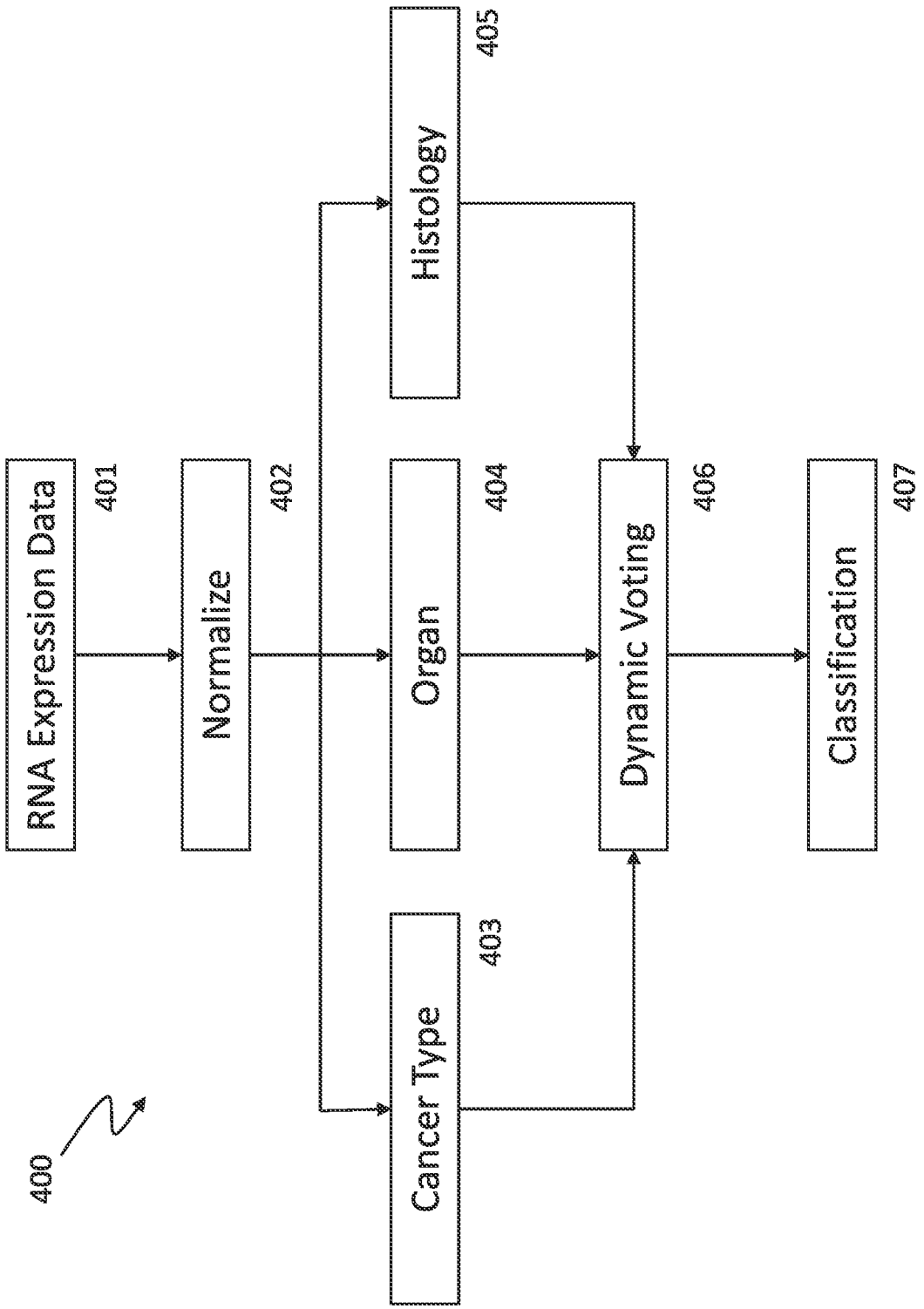

RNA Expression Data
401
Normalize
402
Histology
405
Organ
404
Cancer Type
403
Dynamic Voting
406
Classification
407
400

FIG. 4A

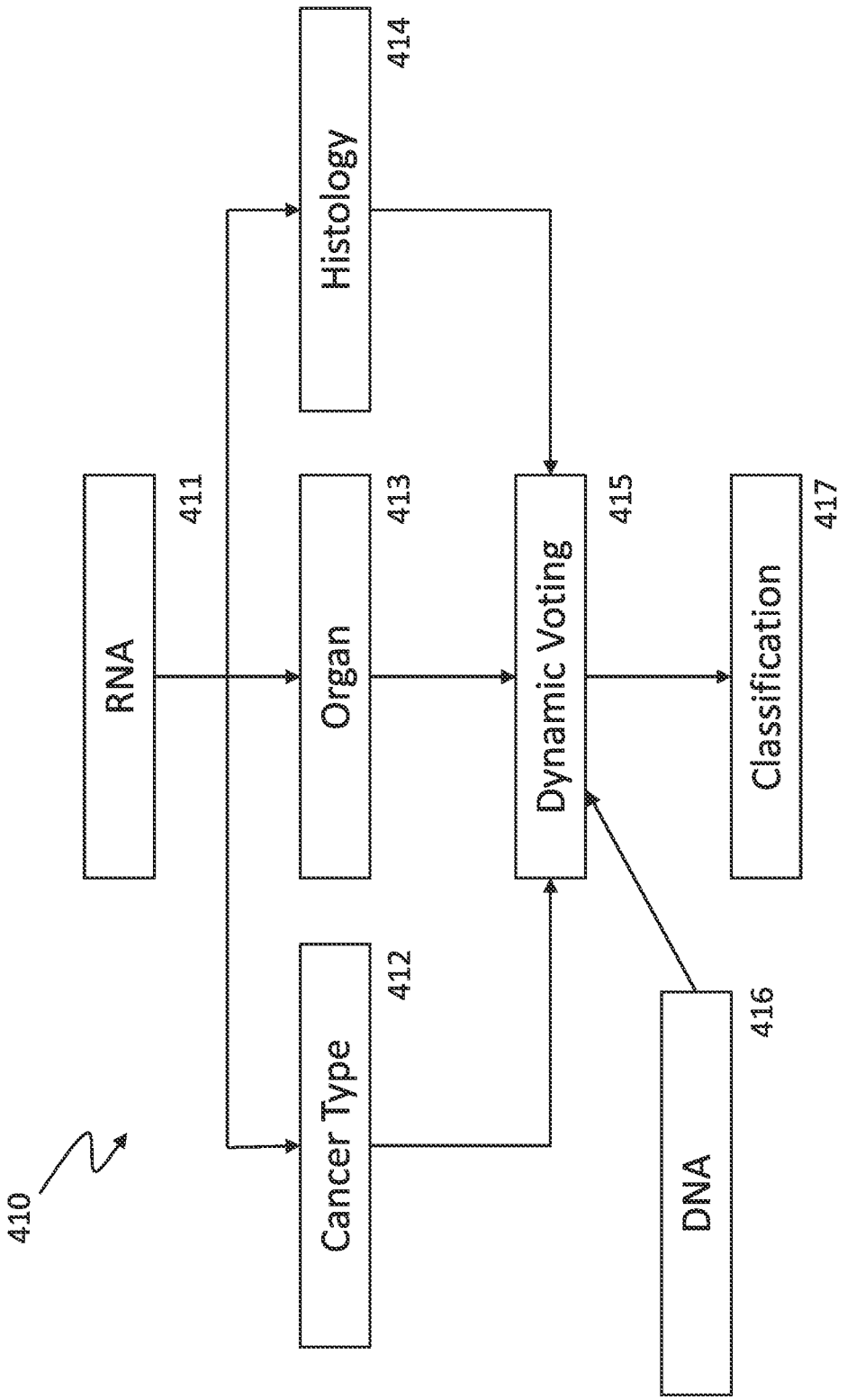

410
RNA
411
Histology
414
Organ
413
Cancer Type
412
Dynamic Voting
415
DNA
416
Classification
417

FIG. 4B

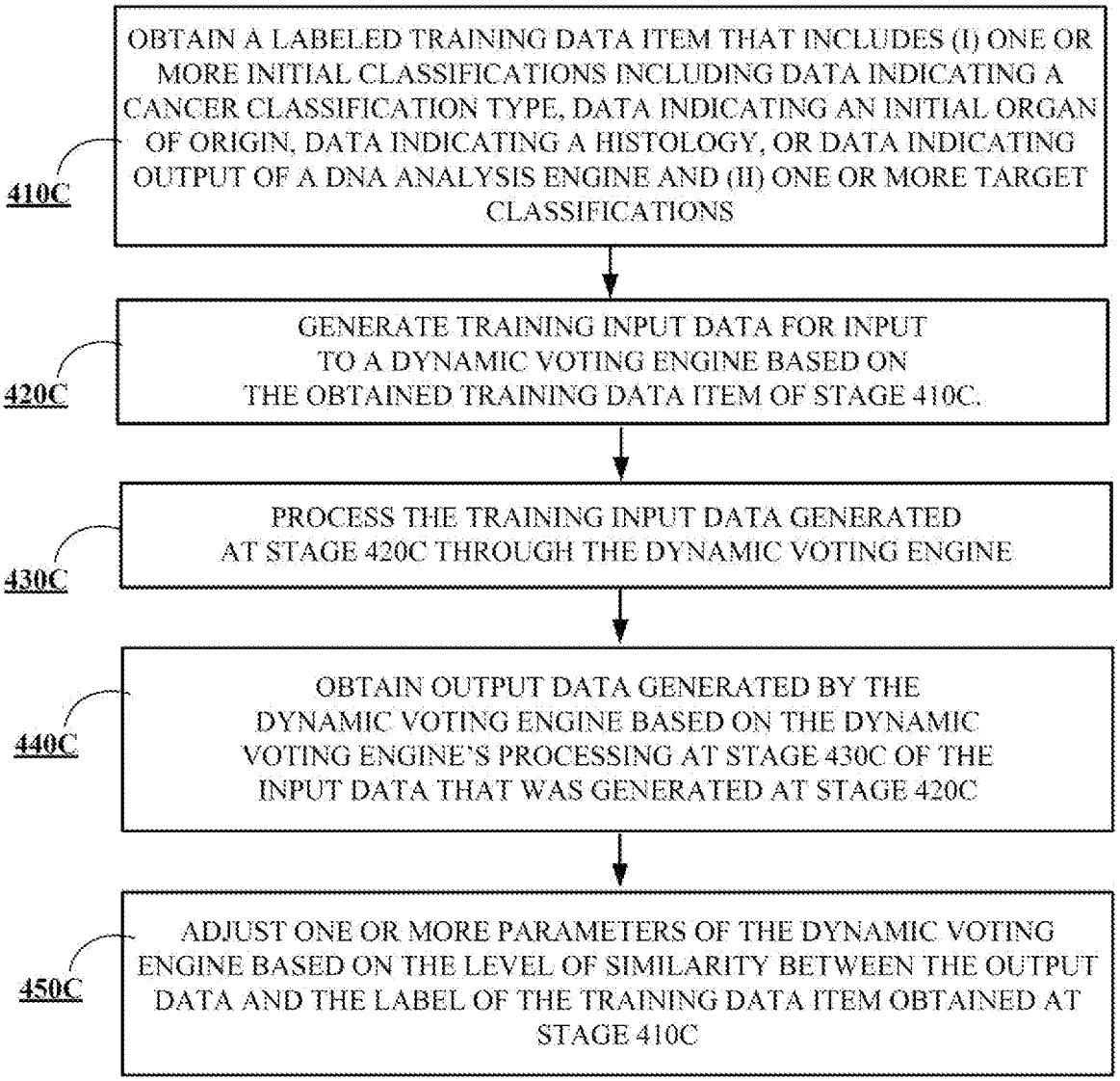

OBTAIN A LABELED TRAINING DATA ITEM THAT INCLUDES (I) ONE OR MORE INITIAL CLASSIFICATIONS INCLUDING DATA INDICATING A CANCER CLASSIFICATION TYPE, DATA INDICATING AN INITIAL ORGAN OF ORIGIN, DATA INDICATING A HISTOLOGY, OR DATA INDICATING OUTPUT OF A DNA ANALYSIS ENGINE AND (II) ONE OR MORE TARGET CLASSIFICATIONS
410C
GENERATE TRAINING INPUT DATA FOR INPUT TO A DYNAMIC VOTING ENGINE BASED ON THE OBTAINED TRAINING DATA ITEM OF STAGE 410C.
420C
PROCESS THE TRAINING INPUT DATA GENERATED AT STAGE 420C THROUGH THE DYNAMIC VOTING ENGINE
430C
OBTAIN OUTPUT DATA GENERATED BY THE DYNAMIC VOTING ENGINE BASED ON THE DYNAMIC VOTING ENGINE'S PROCESSING AT STAGE 430C OF THE INPUT DATA THAT WAS GENERATED AT STAGE 420C
440C
ADJUST ONE OR MORE PARAMETERS OF THE DYNAMIC VOTING ENGINE BASED ON THE LEVEL OF SIMILARITY BETWEEN THE OUTPUT DATA AND THE LABEL OF THE TRAINING DATA ITEM OBTAINED AT STAGE 410C
450C

FIG. 4C

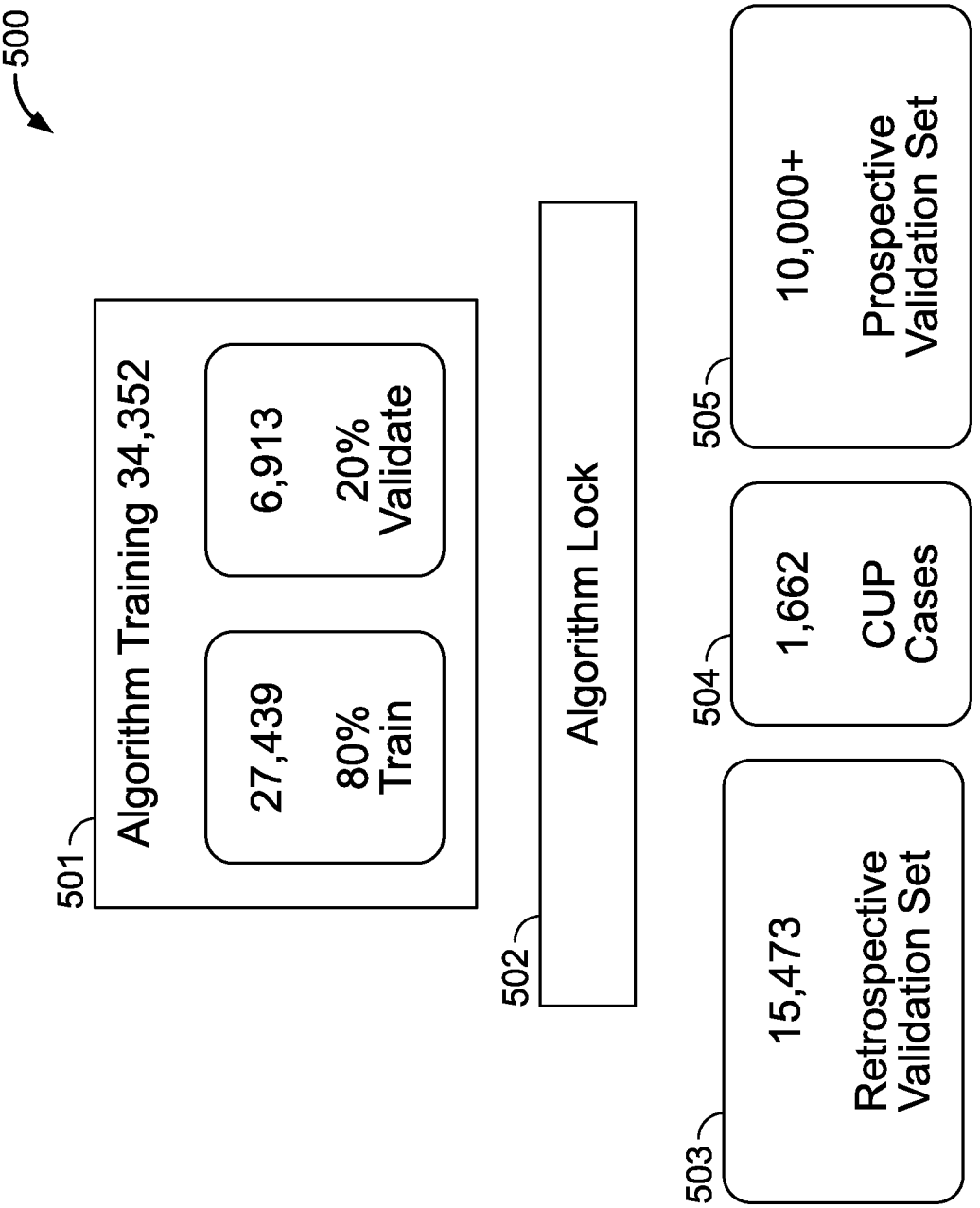

500
501
Algorithm Training 34,352
27,439
80%
Train
6,913
20%
Validate
502
Algorithm Lock
503
15,473
Retrospective
Validation Set
504
1,662
CUP
Cases
505
10,000+
Prospective
Validation Set

FIG. 5A

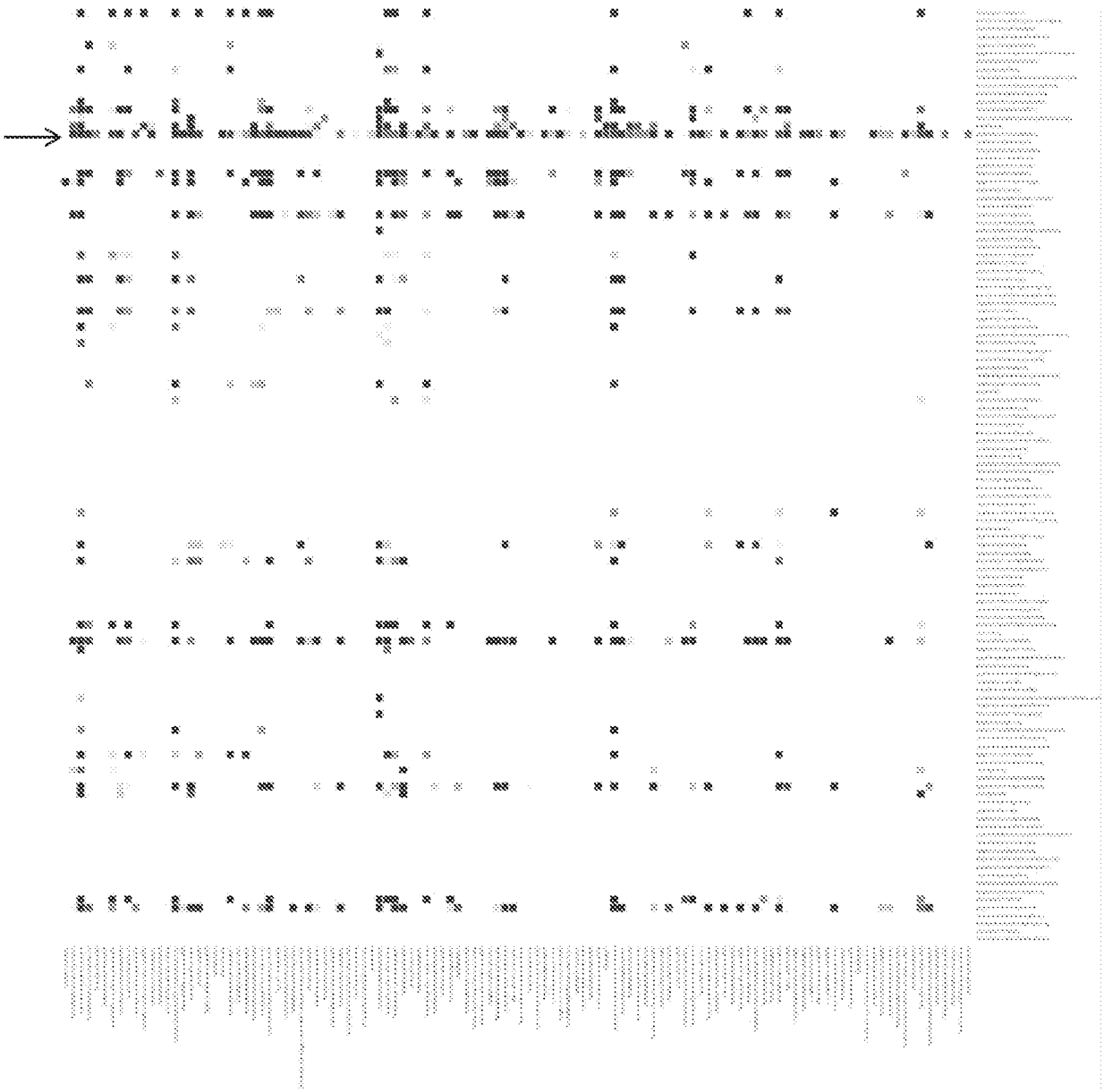


FIG. 5B

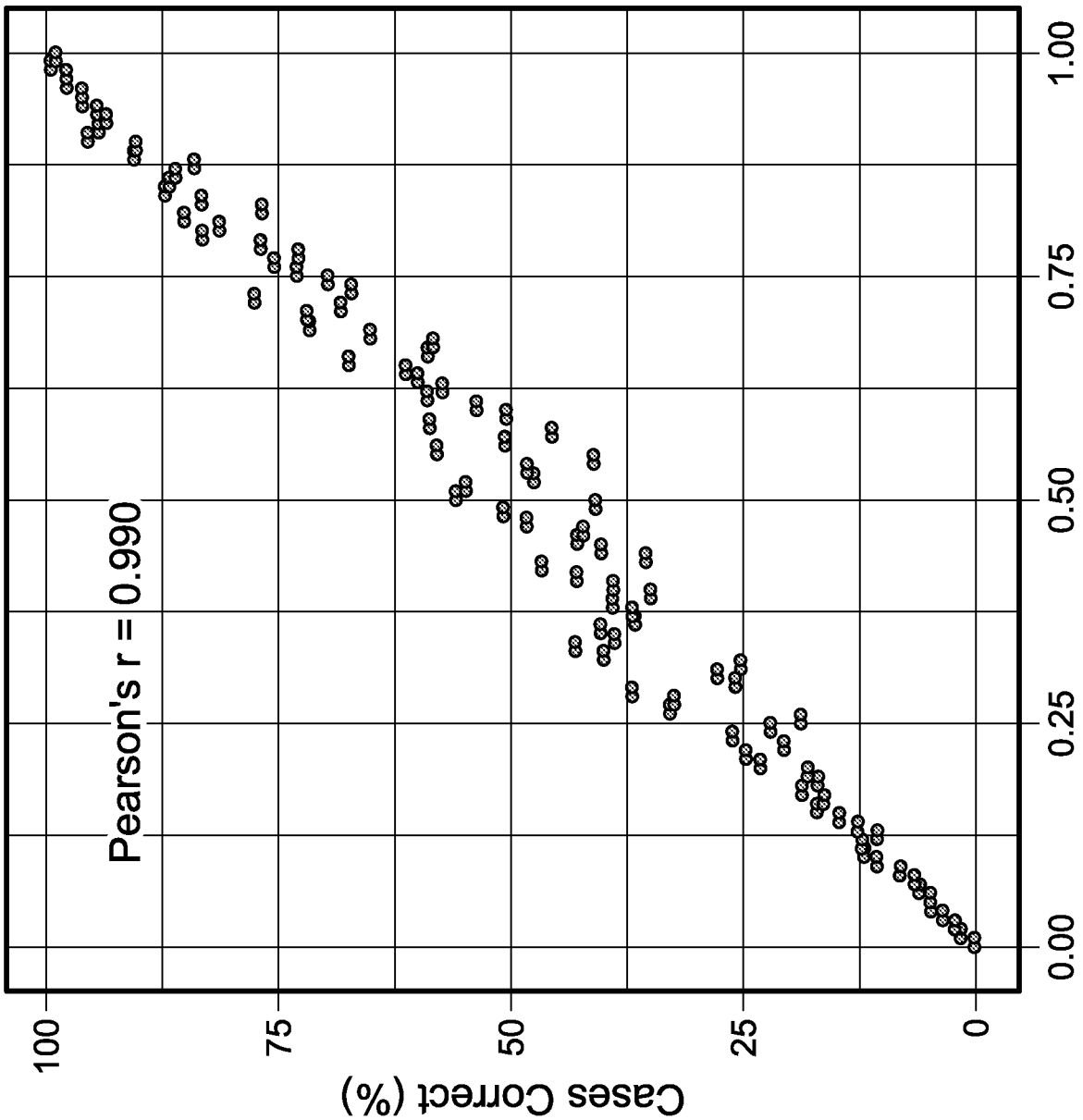

Pearson's r = 0.990
Cases Correct (%)
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Molecular GPS Score

FIG. 5C

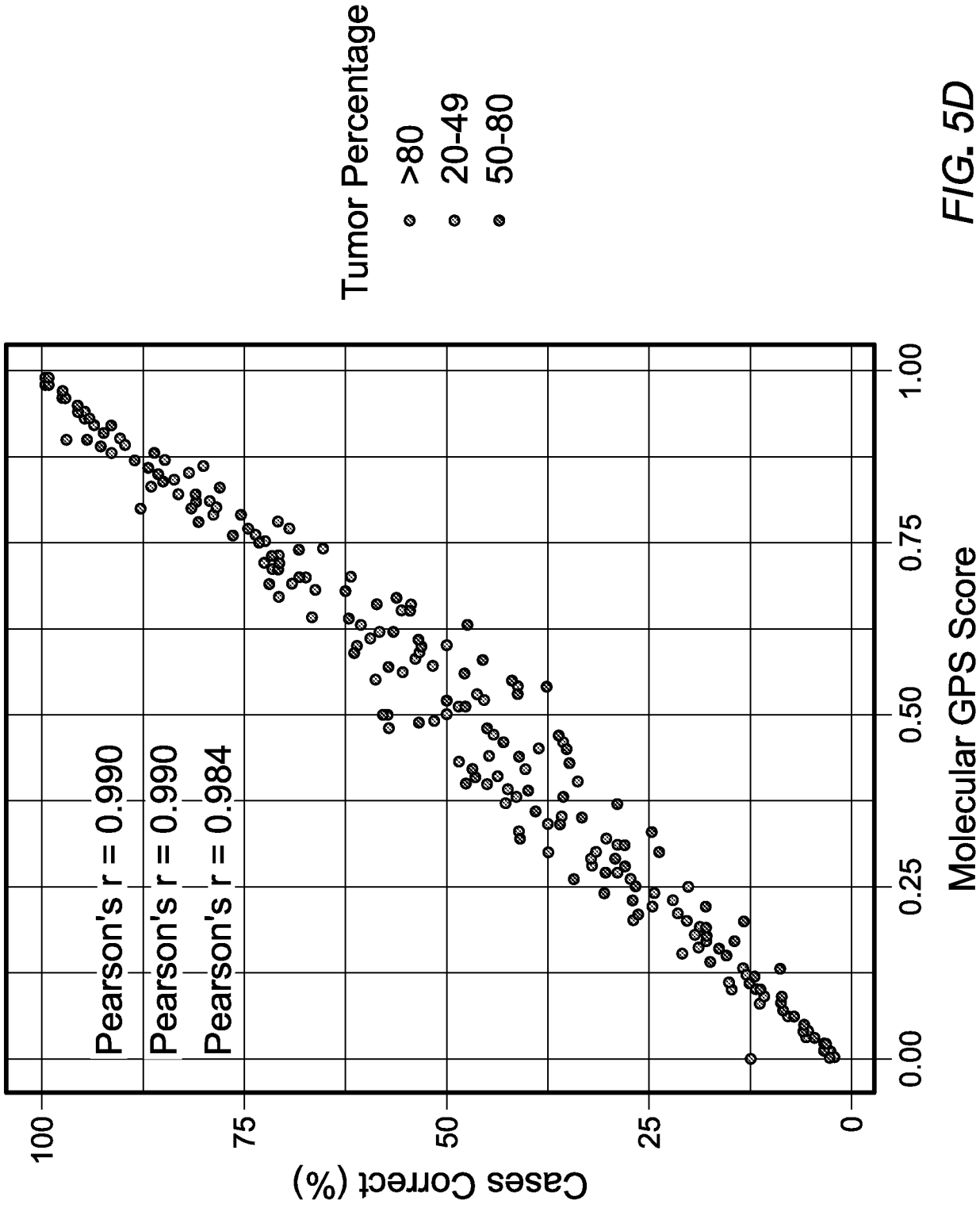

Tumor Percentage
>80
20-49
50-80
Pearson's r = 0.990
Pearson's r = 0.990
Pearson's r = 0.984
Cases Correct (%)
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Molecular GPS Score

FIG. 5D

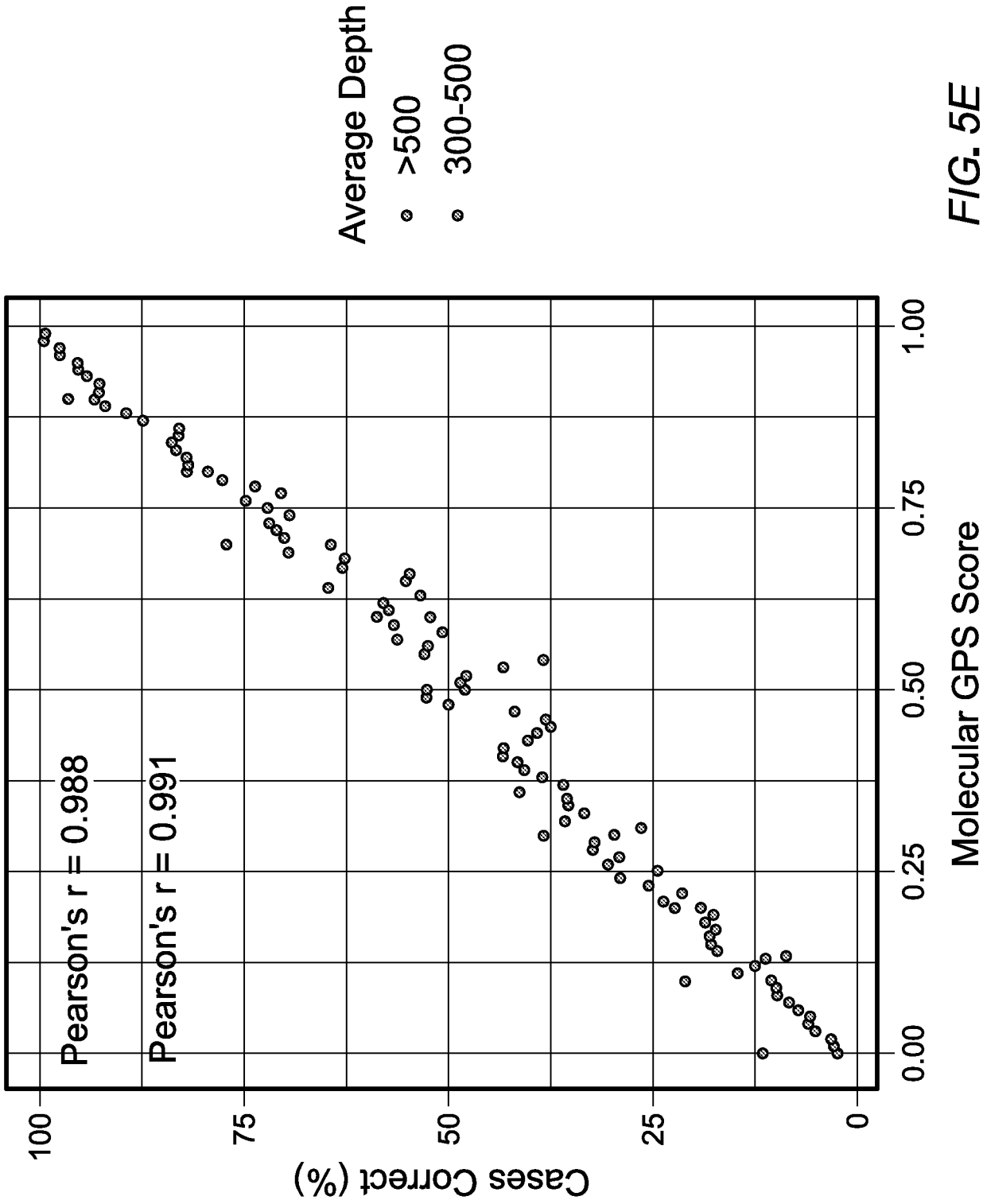

Average Depth
>500
300-500
Pearson's r = 0.988
Pearson's r = 0.991
Cases Correct (%)
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Molecular GPS Score

FIG. 5E

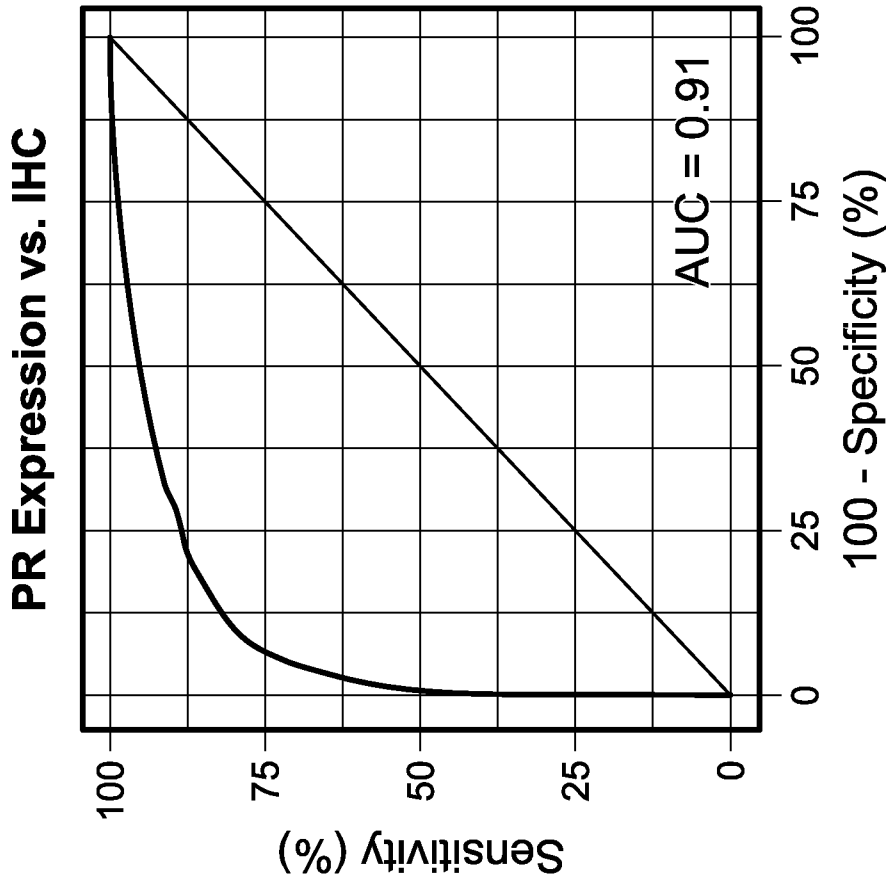

PR Expression vs. IHC
AUC = 0.91
Sensitivity (%)
100 - Specificity (%)
0
25
50
75
100

FIG. 6B

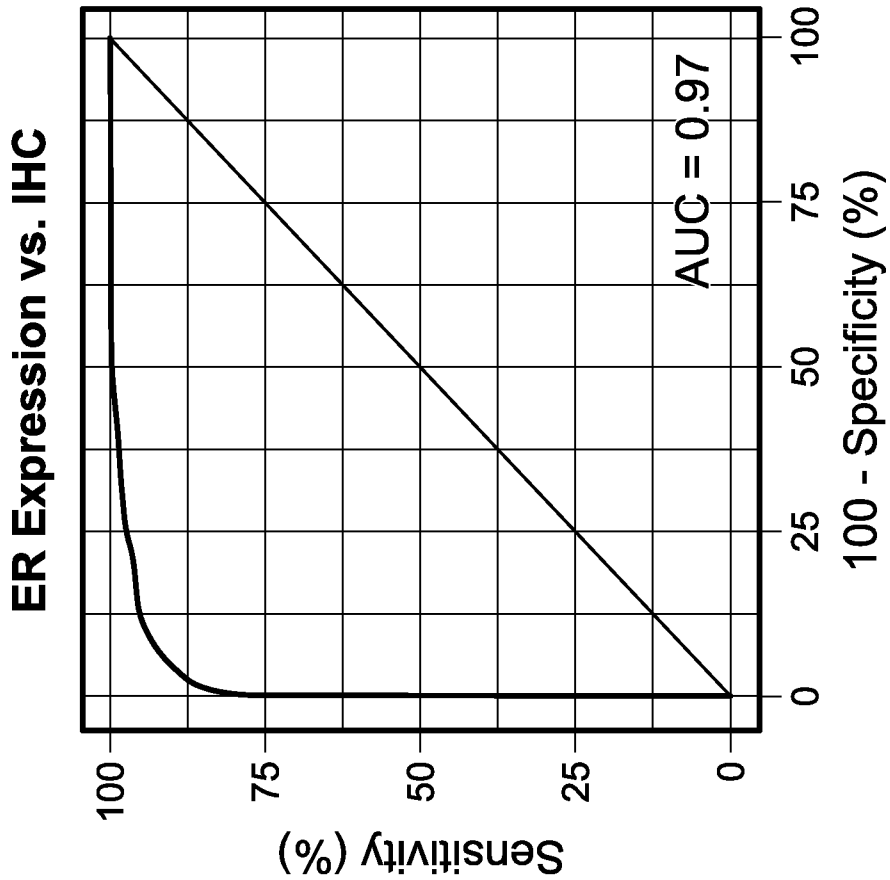

ER Expression vs. IHC
AUC = 0.97
Sensitivity (%)
100 - Specificity (%)
0
25
50
75
100

FIG. 6A

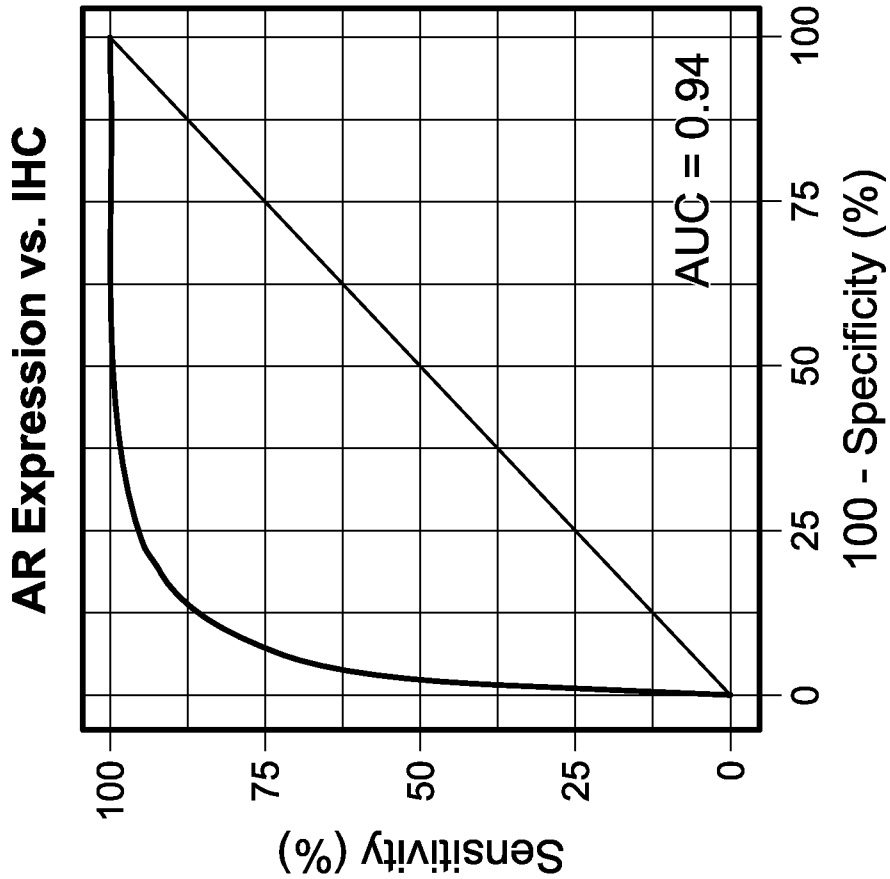

AR Expression vs. IHC
AUC = 0.94
Sensitivity (%)
100 - Specificity (%)
0
25
50
75
100

FIG. 6D

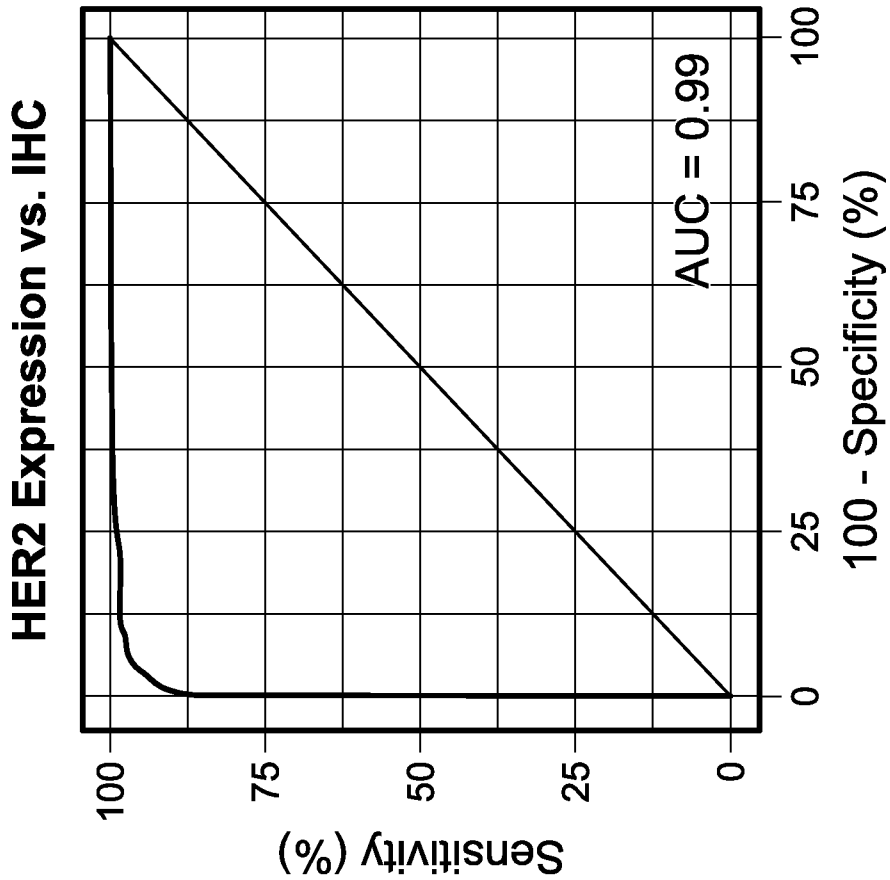

HER2 Expression vs. IHC
AUC = 0.99
Sensitivity (%)
100 - Specificity (%)
0
25
50
75
100

FIG. 6C

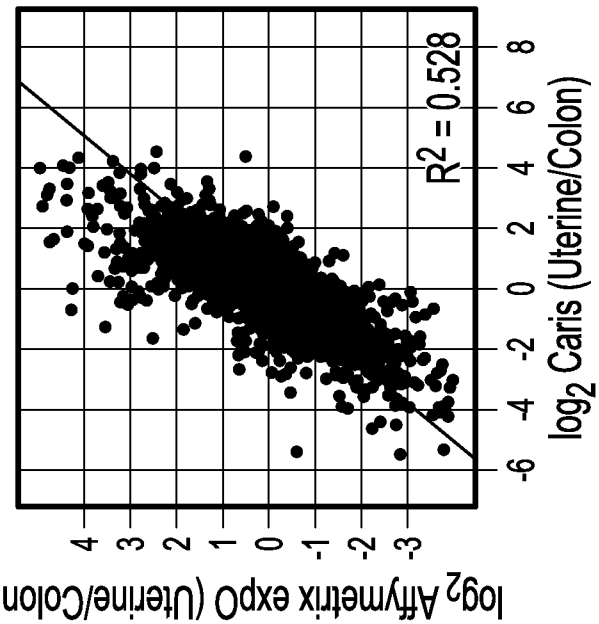


FIG. 6G

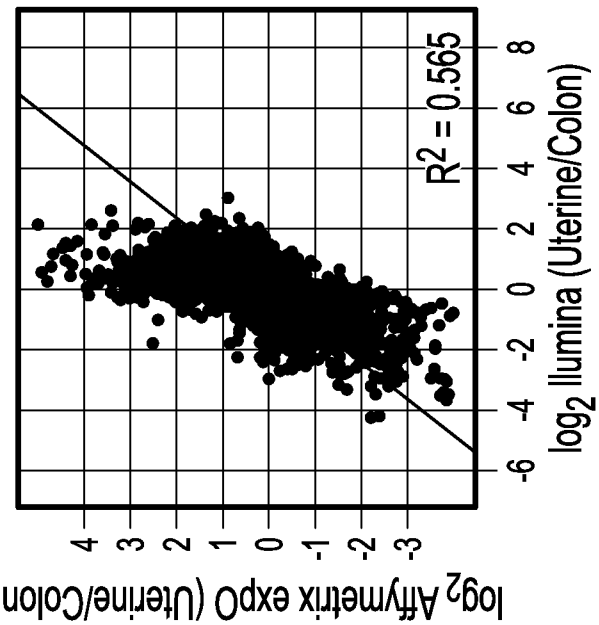


FIG. 6F

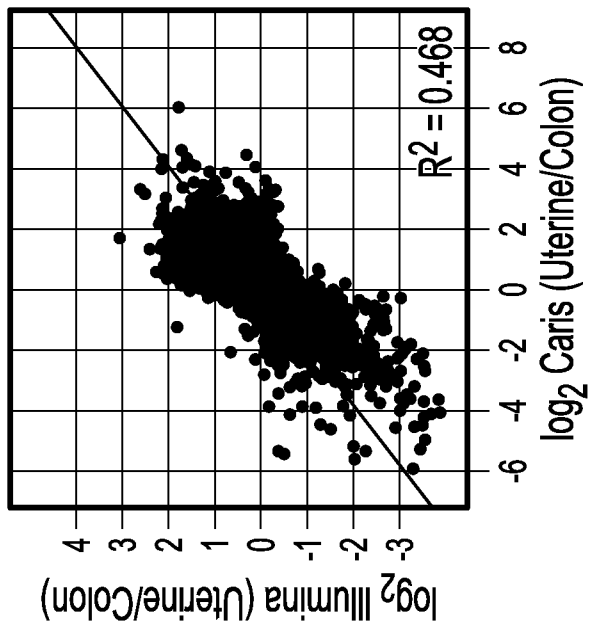


FIG. 6E

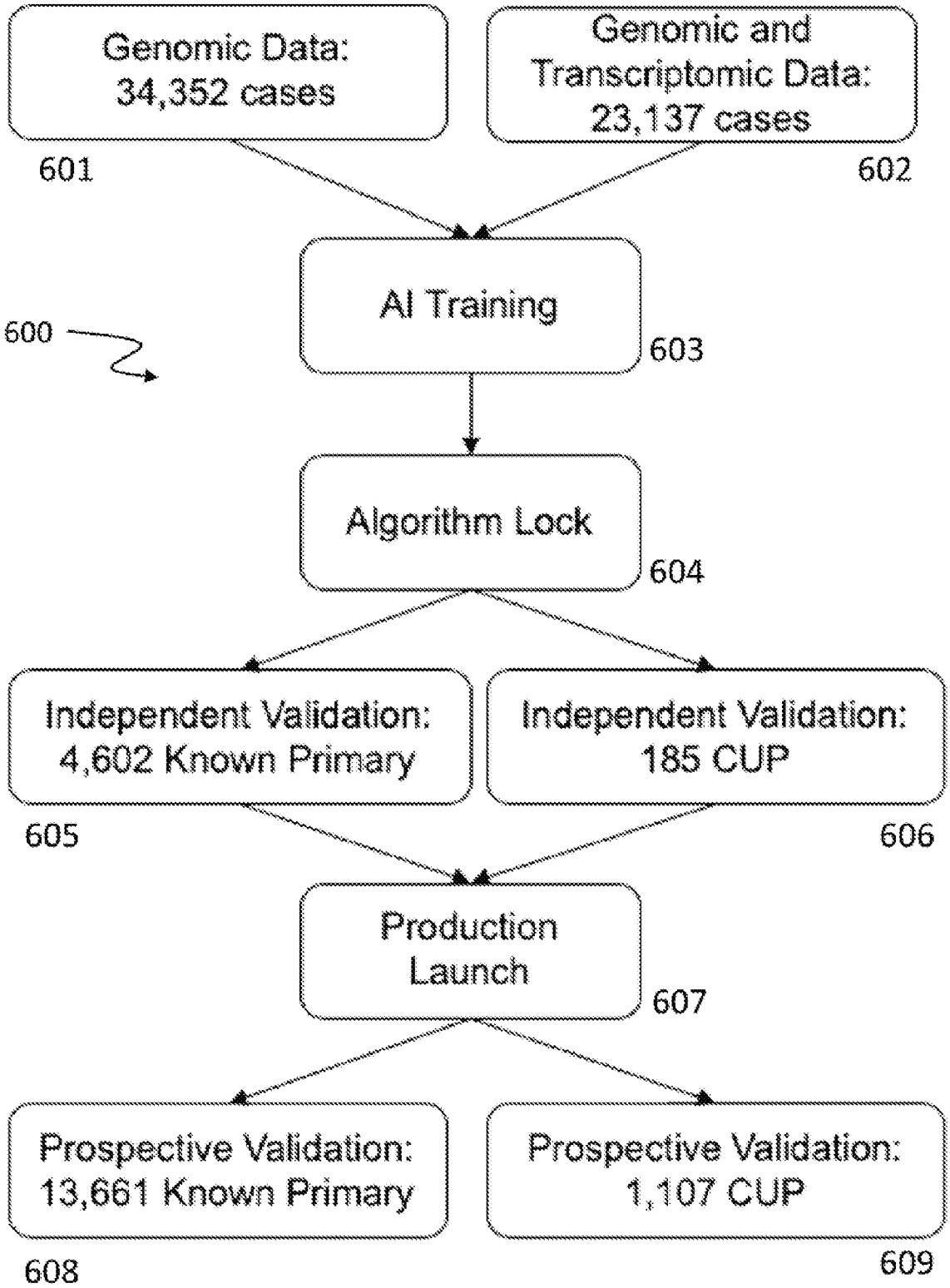

Genomic Data:
34,352 cases
601
Genomic and
Transcriptomic Data:
23,137 cases
602
600
AI Training
603
Algorithm Lock
604
Independent Validation:
4,602 Known Primary
605
Independent Validation:
185 CUP
606
Production
Launch
607
Prospective Validation:
13,661 Known Primary
608
Prospective Validation:
1,107 CUP
609

FIG. 6H

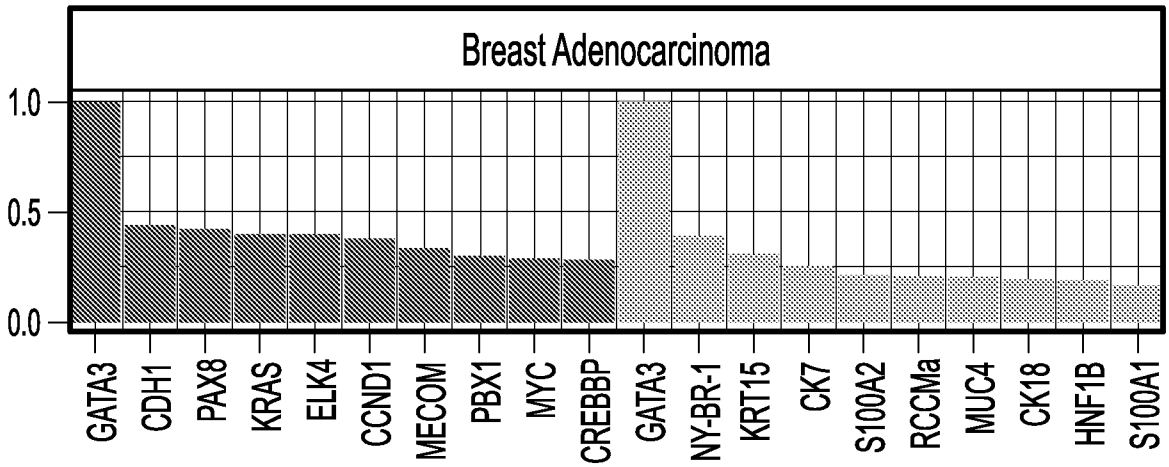


FIG. 6I

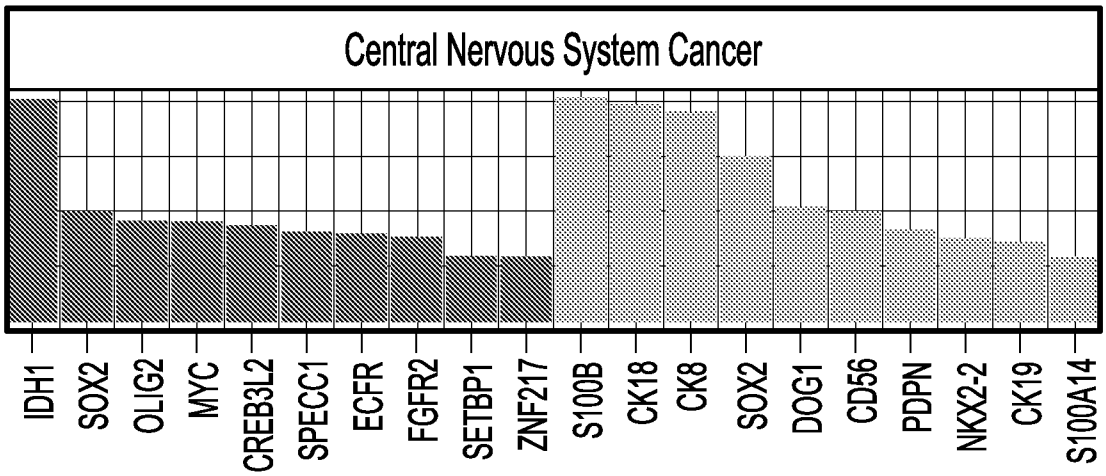


FIG. 6J

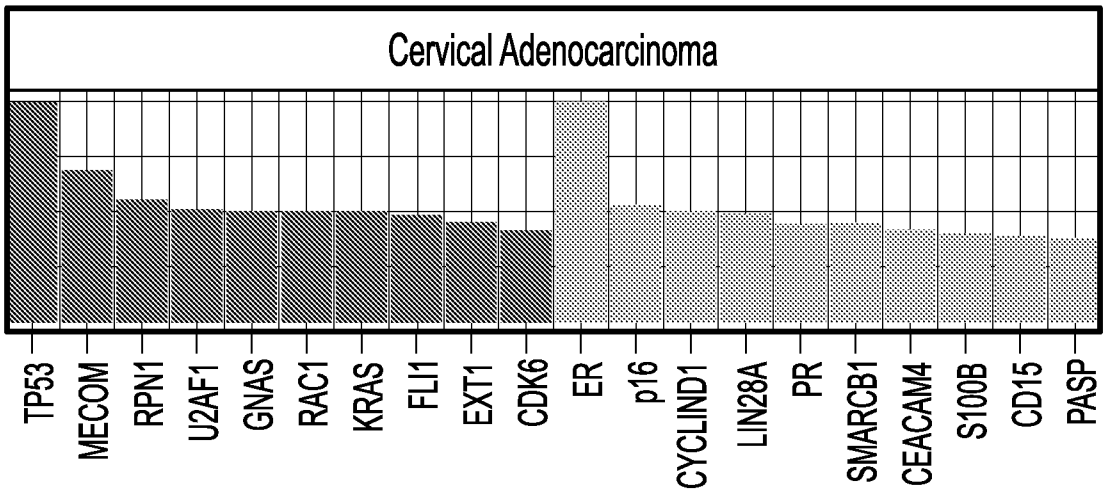


FIG. 6K

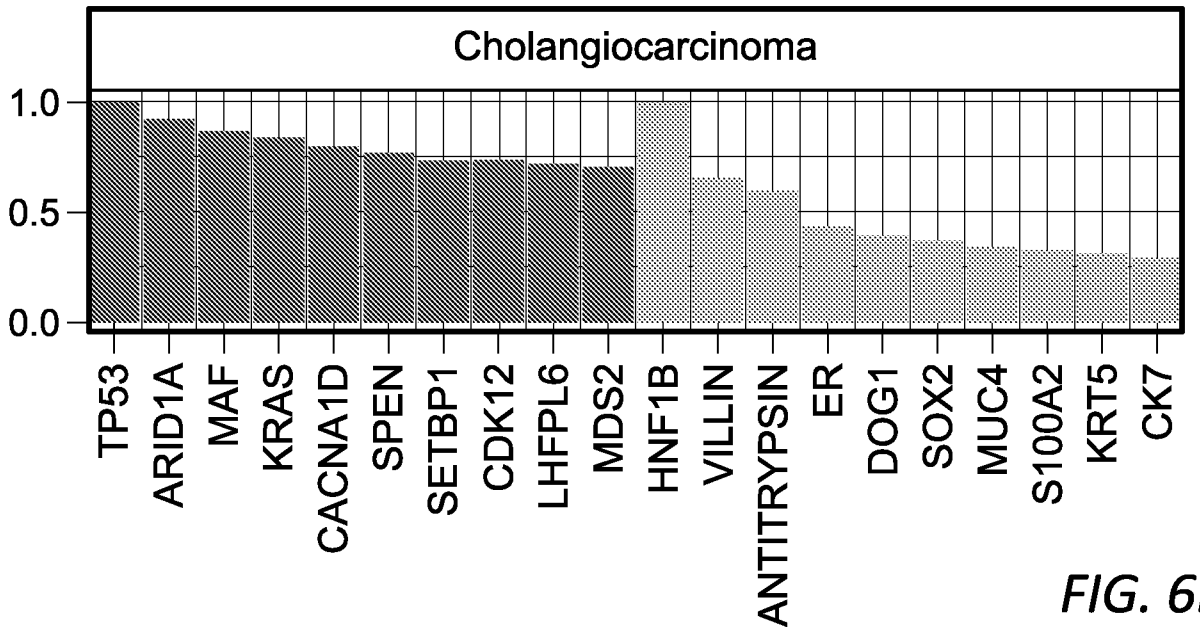


FIG. 6L

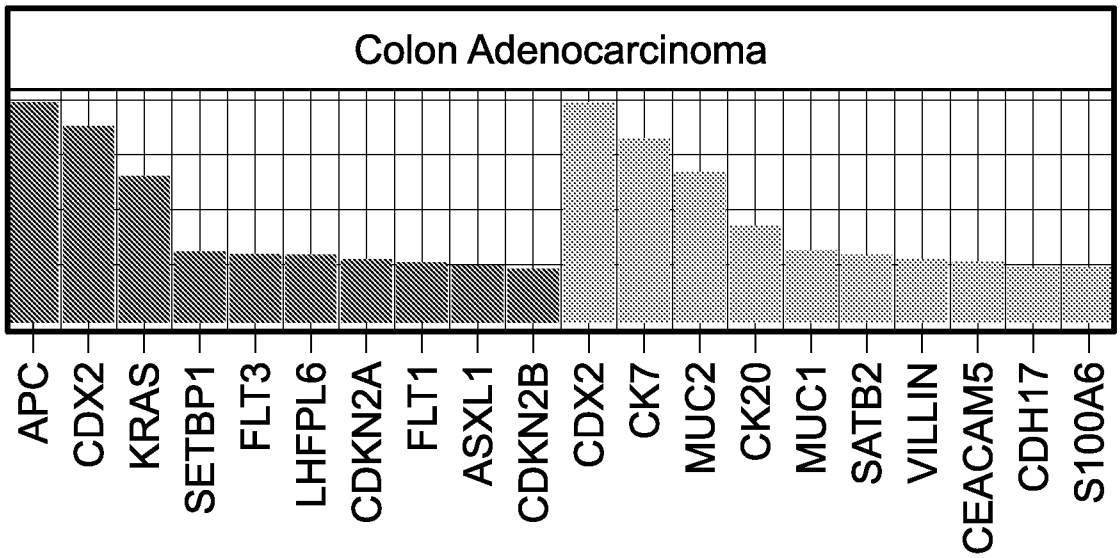


FIG. 6M

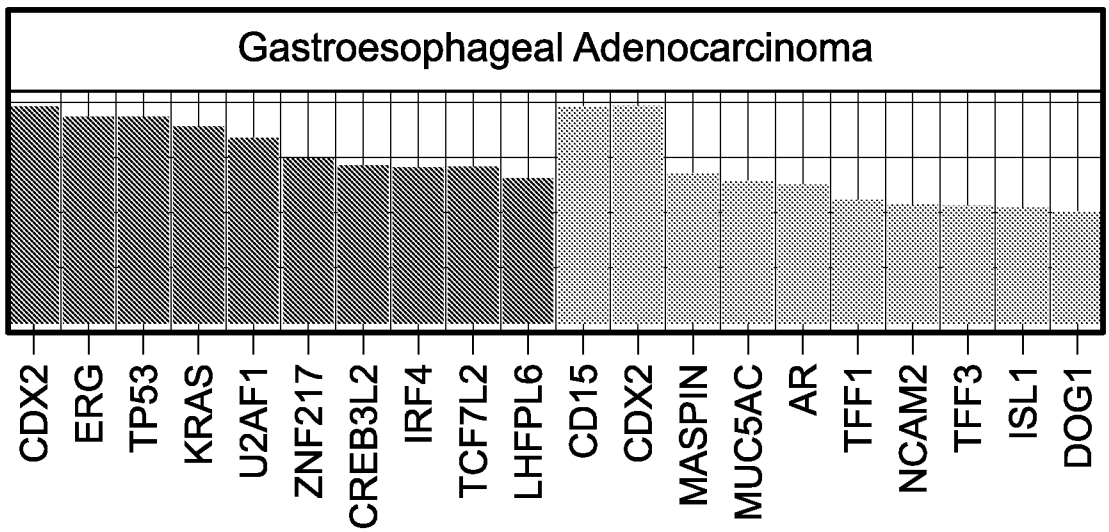


FIG. 6N

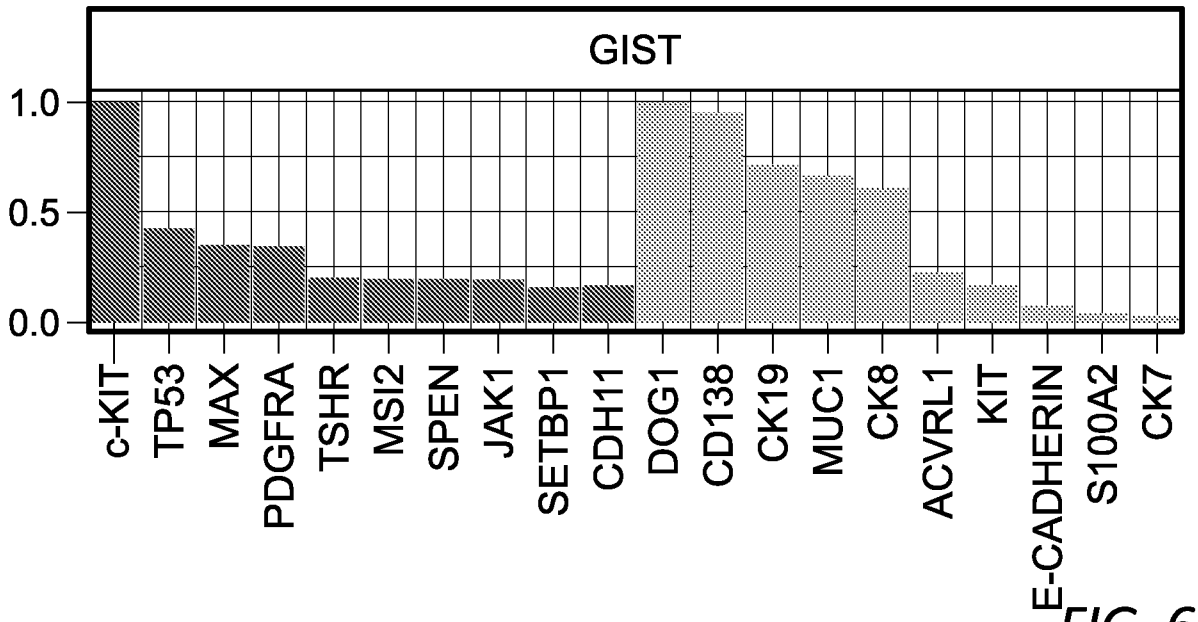


FIG. 6O

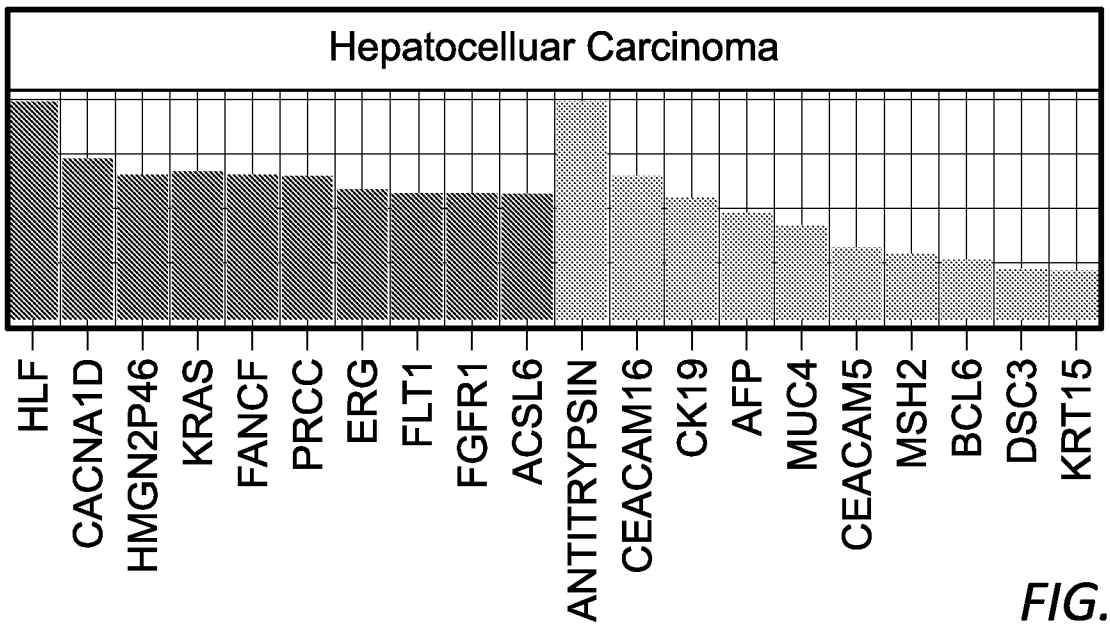


FIG. 6P

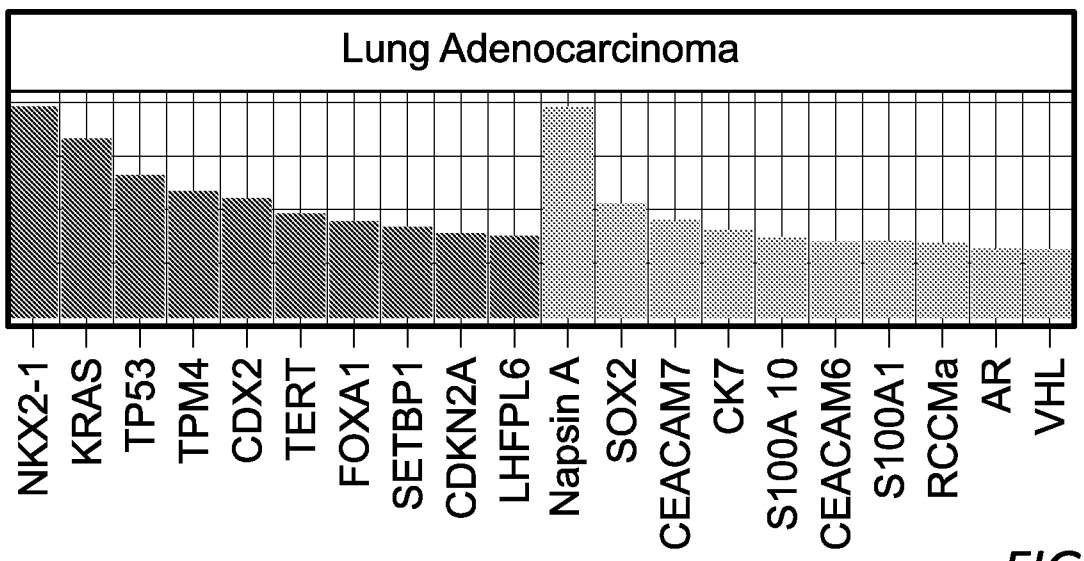


FIG. 6Q

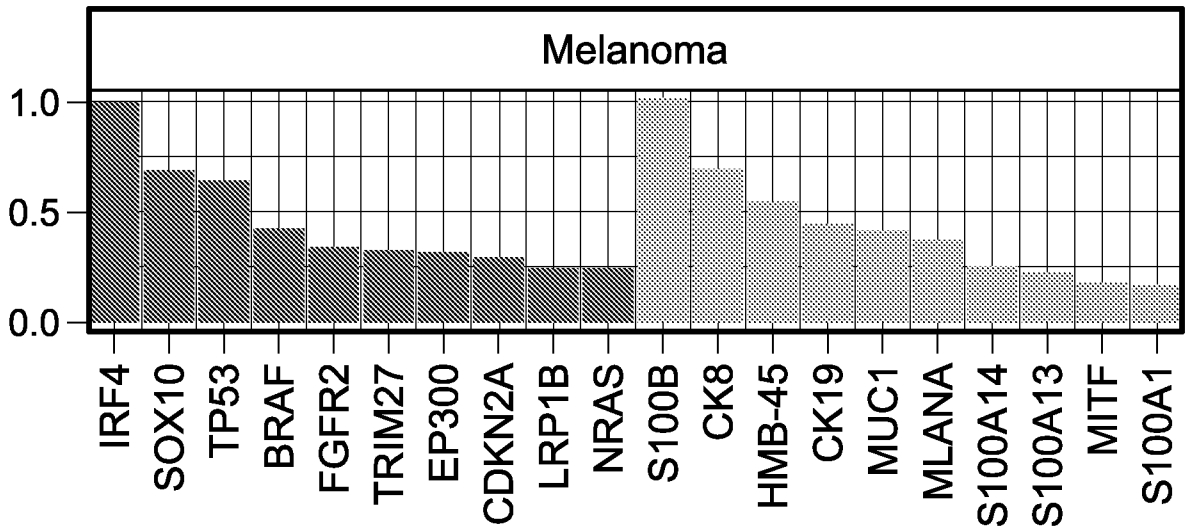


FIG. 6R

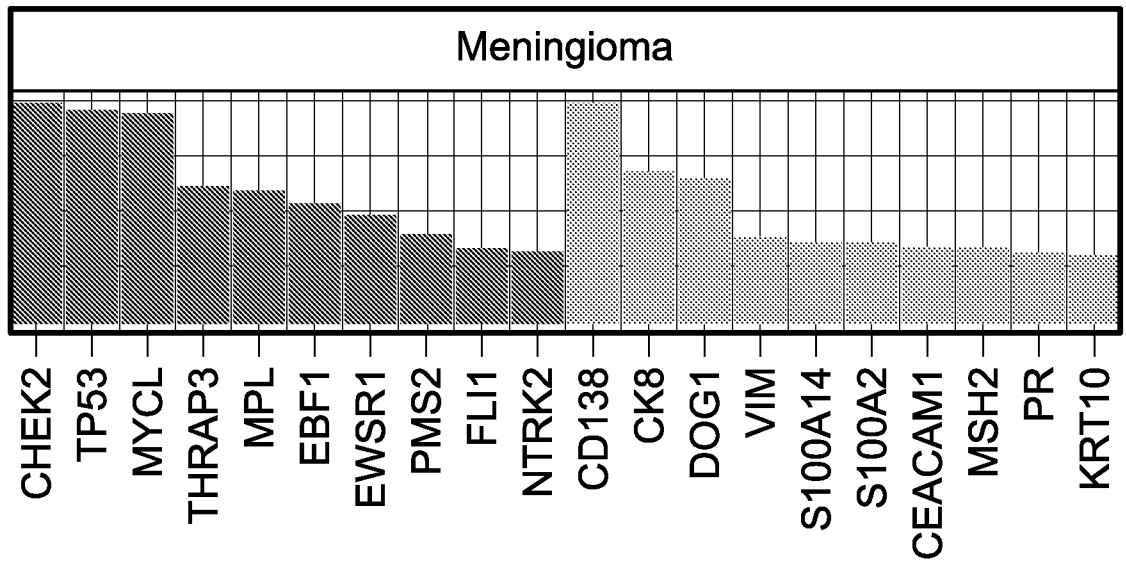


FIG. 6S

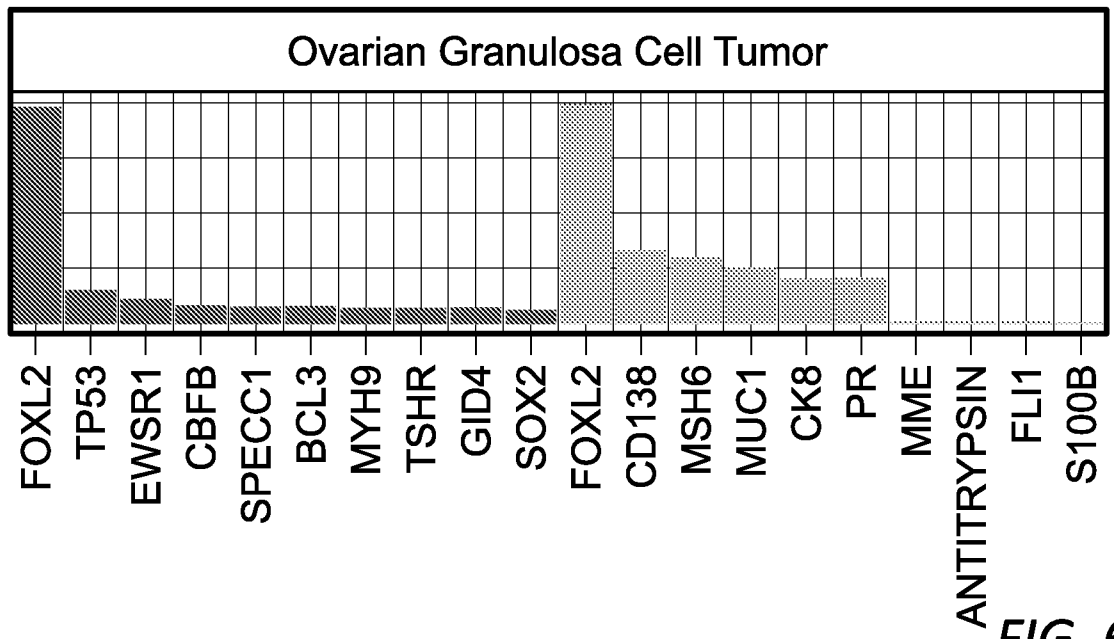


FIG. 6T

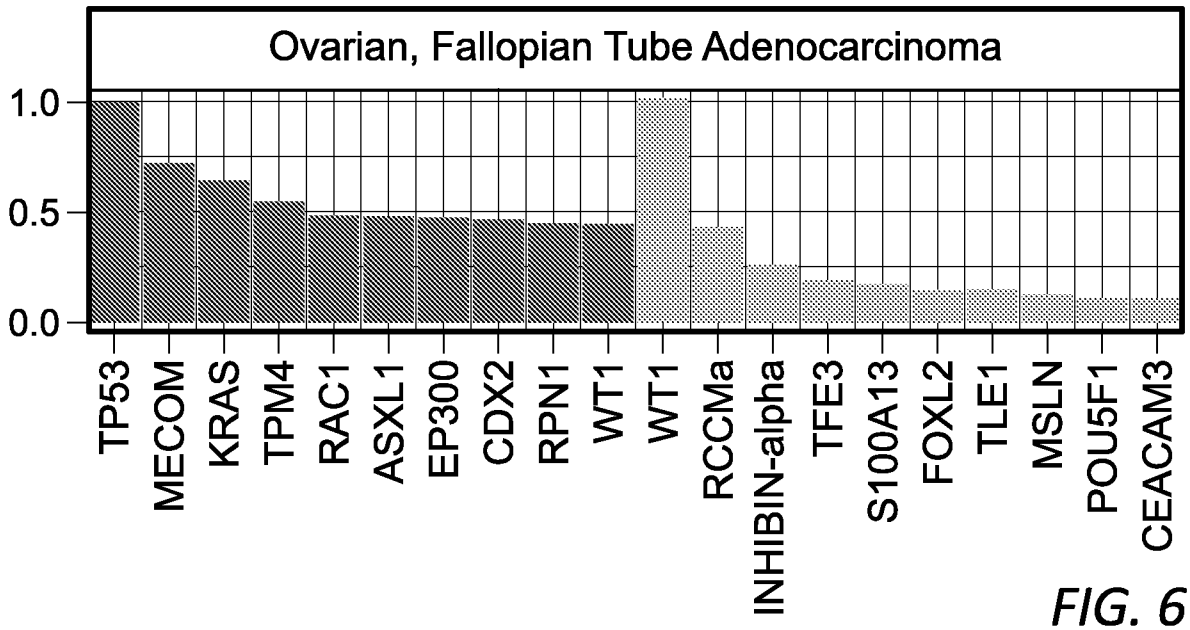


FIG. 6U

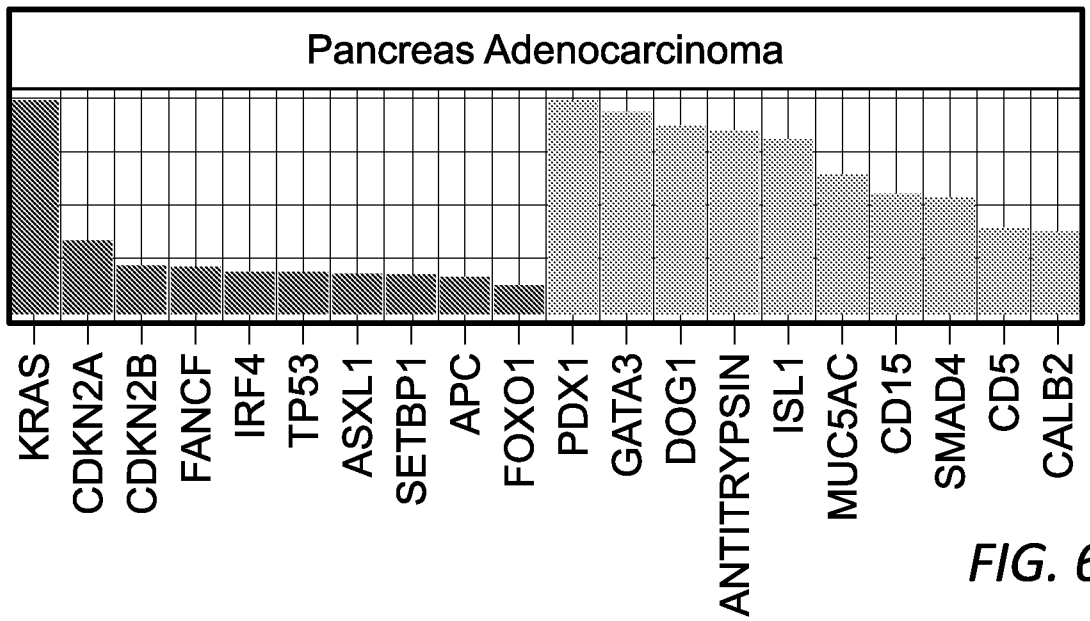


FIG. 6V

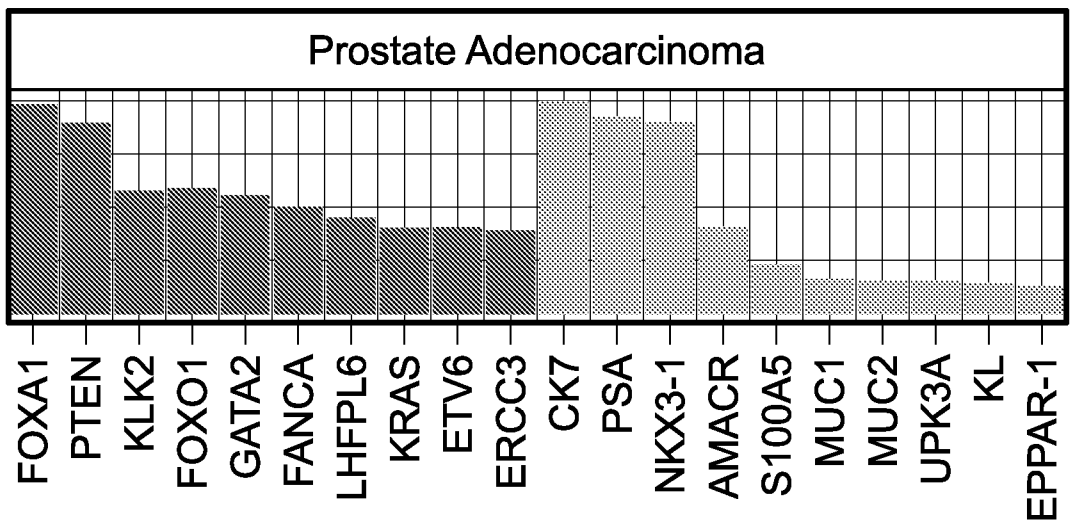


FIG. 6W

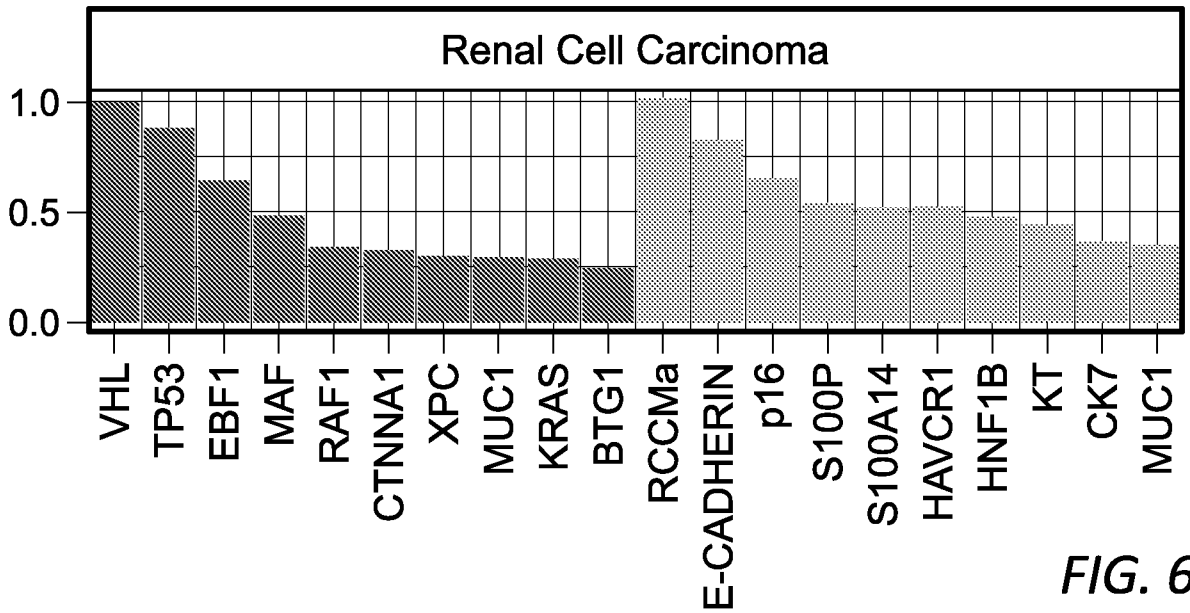


FIG. 6X

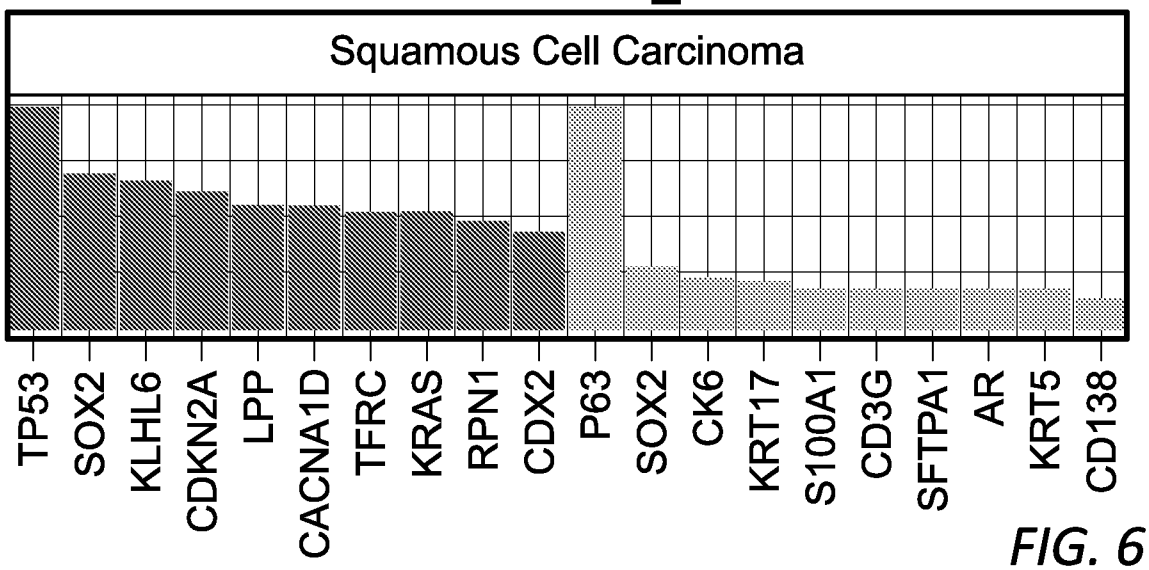


FIG. 6Y

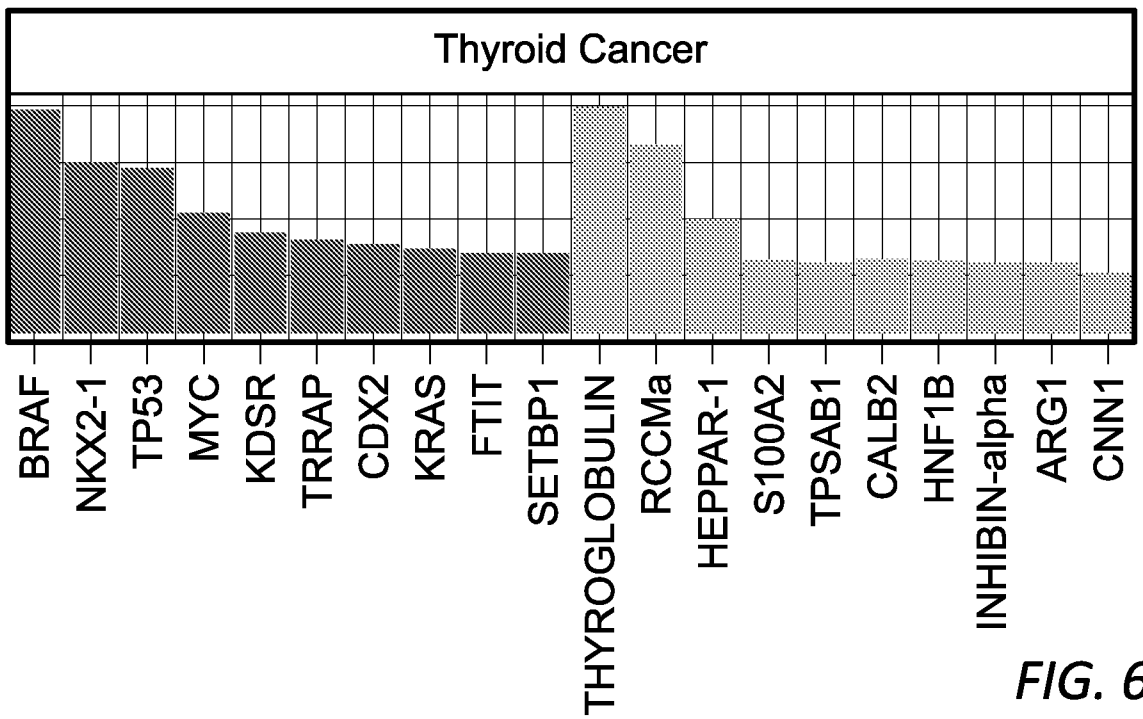


FIG. 6Z

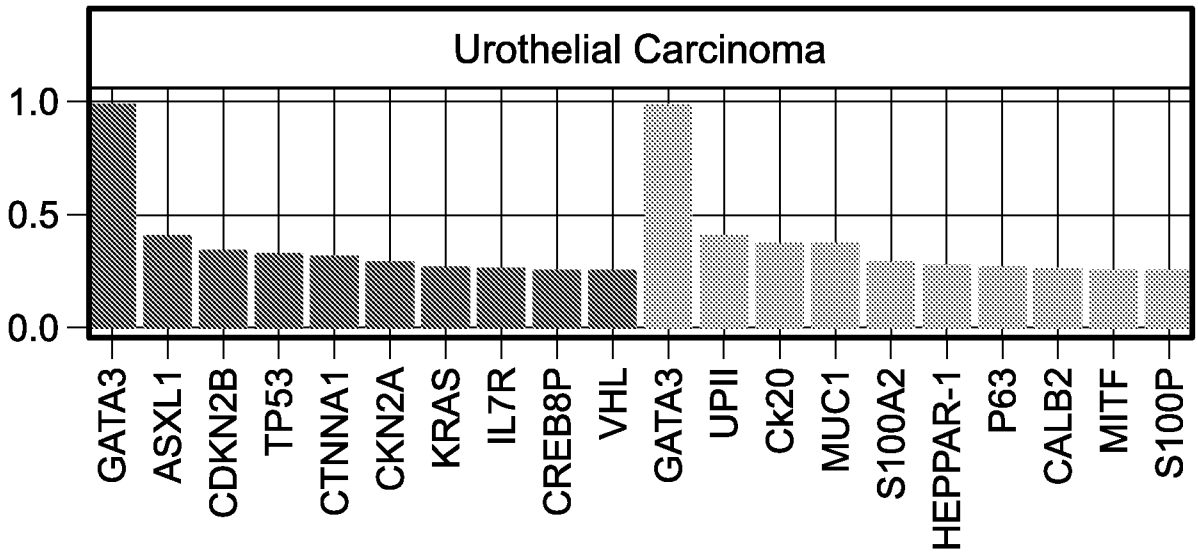


FIG. 6AA

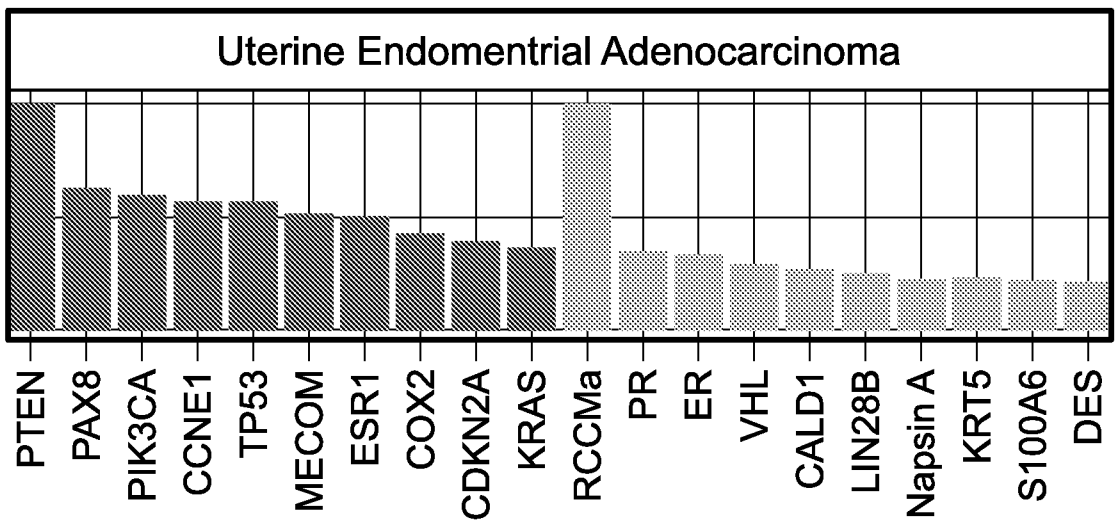


FIG. 6AB

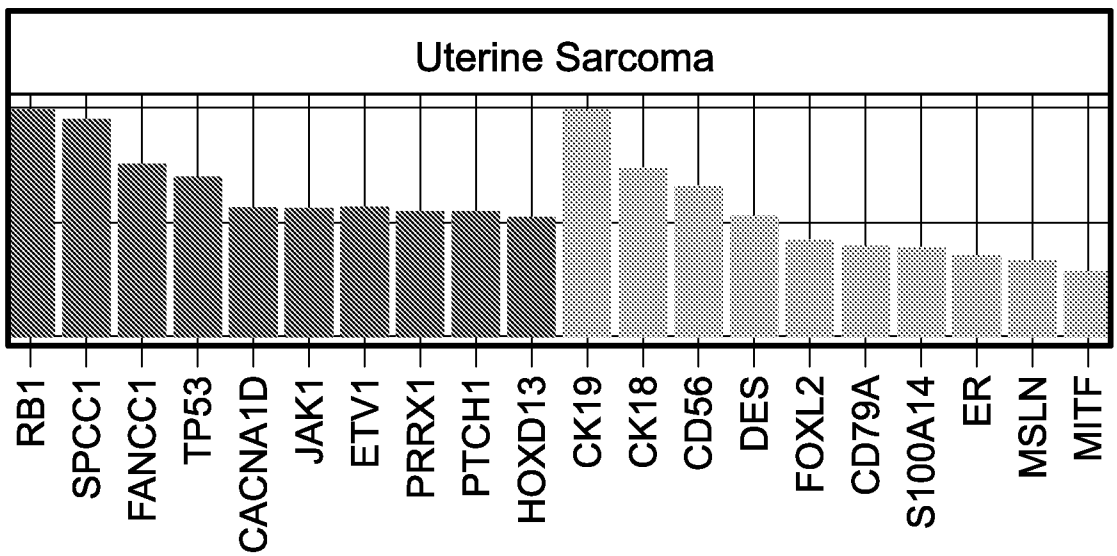


FIG. 6AC

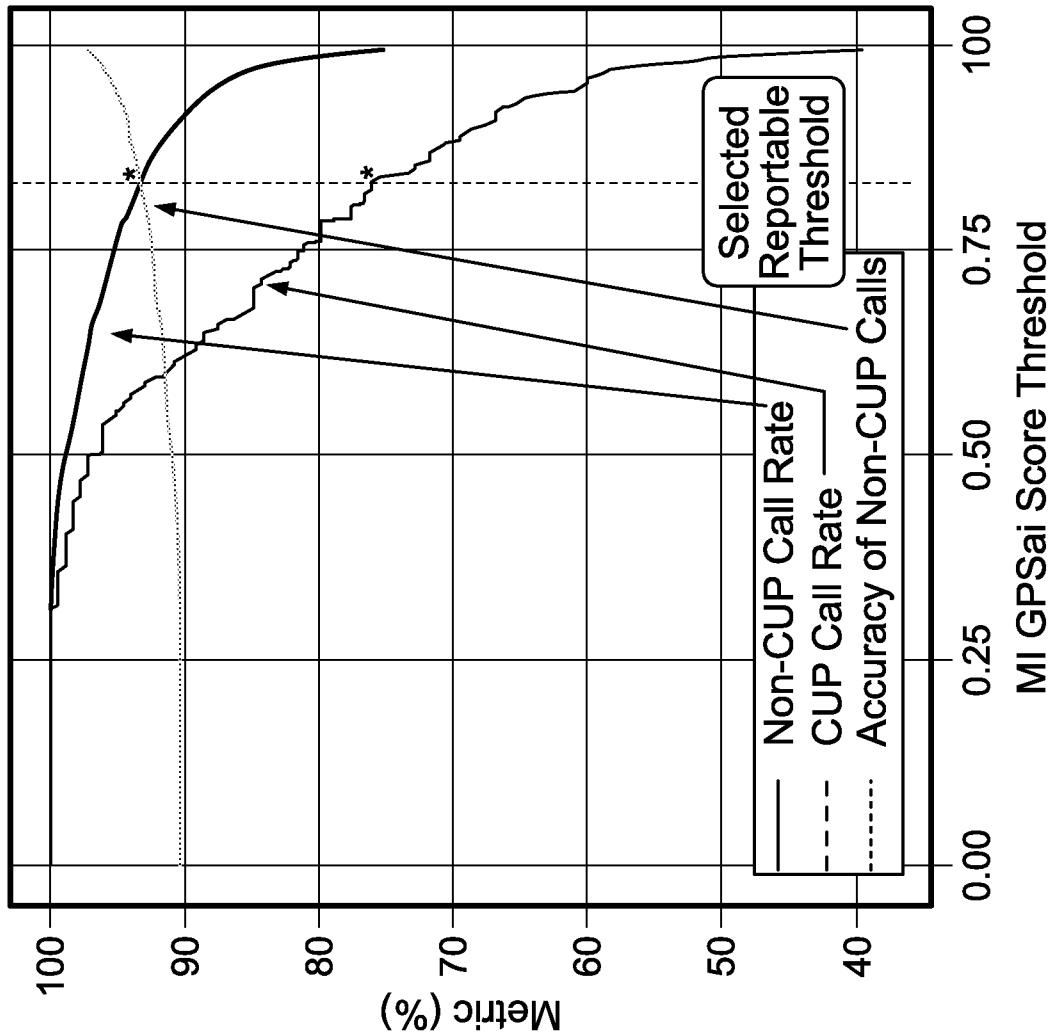

Metric (%)
100
90
80
70
60
50
40
0.00
0.25
0.50
0.75
100
MI GPSai Score Threshold
Non-CUP Call Rate
CUP Call Rate
Accuracy of Non-CUP Calls
Selected Reportable Threshold

FIG. 6AD

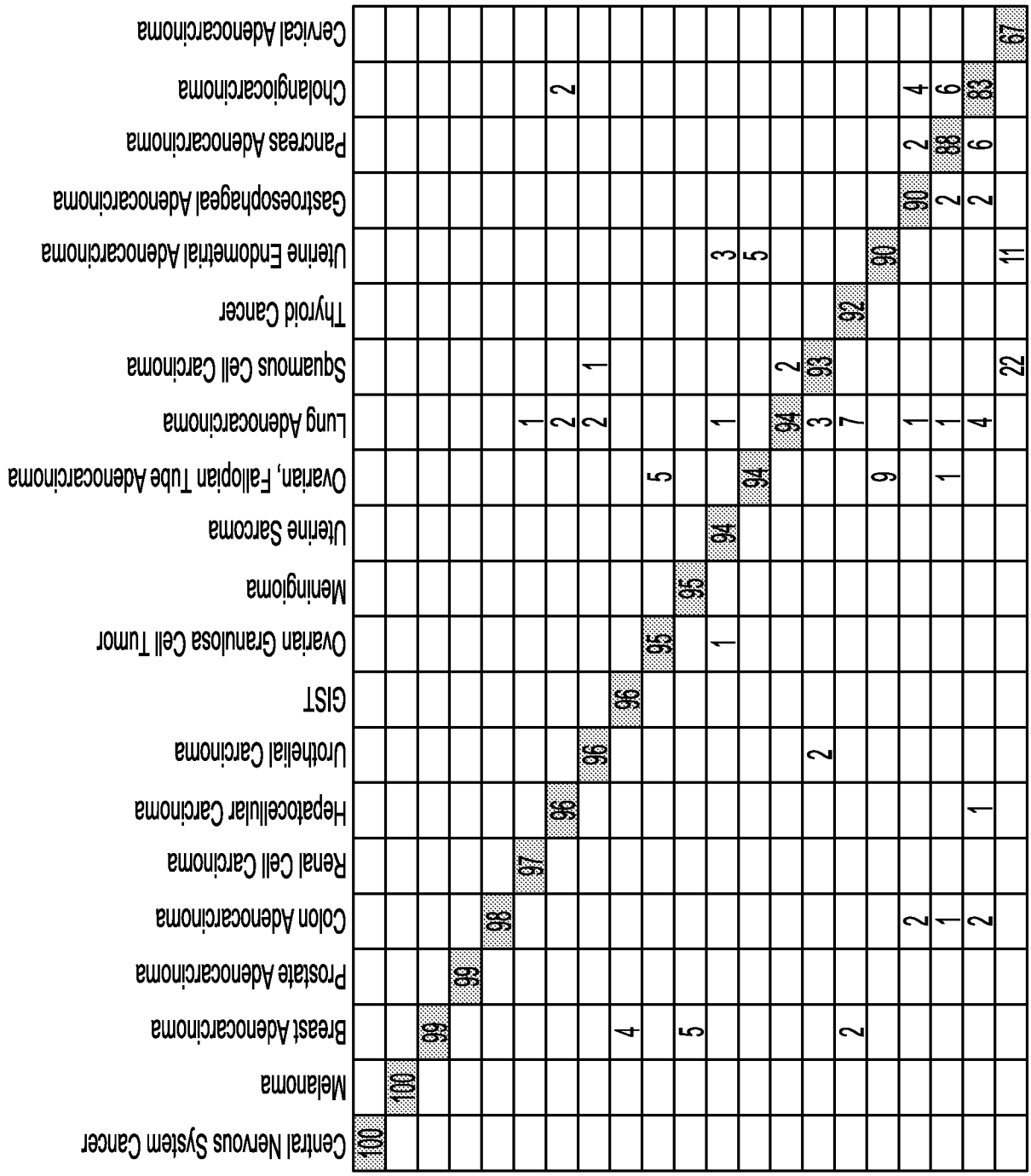

Central Nervous System Cancer
Melanoma
Breast Adenocarcinoma
Prostate Adenocarcinoma
Colon Adenocarcinoma
Renal Cell Carcinoma
Hepatocellular Carcinoma
Urothelial Carcinoma
GIST
Ovarian Granulosa Cell Tumor
Meningioma
Uterine Sarcoma
Ovarian, Fallopian Tube Adenocarcinoma
Lung Adenocarcinoma
Squamous Cell Carcinoma
Thyroid Cancer
Uterine Endometrial Adenocarcinoma
Gastroesophageal Adenocarcinoma
Pancreas Adenocarcinoma
Cholangiocarcinoma
Cervical Adenocarcinoma

FIG. 6AE

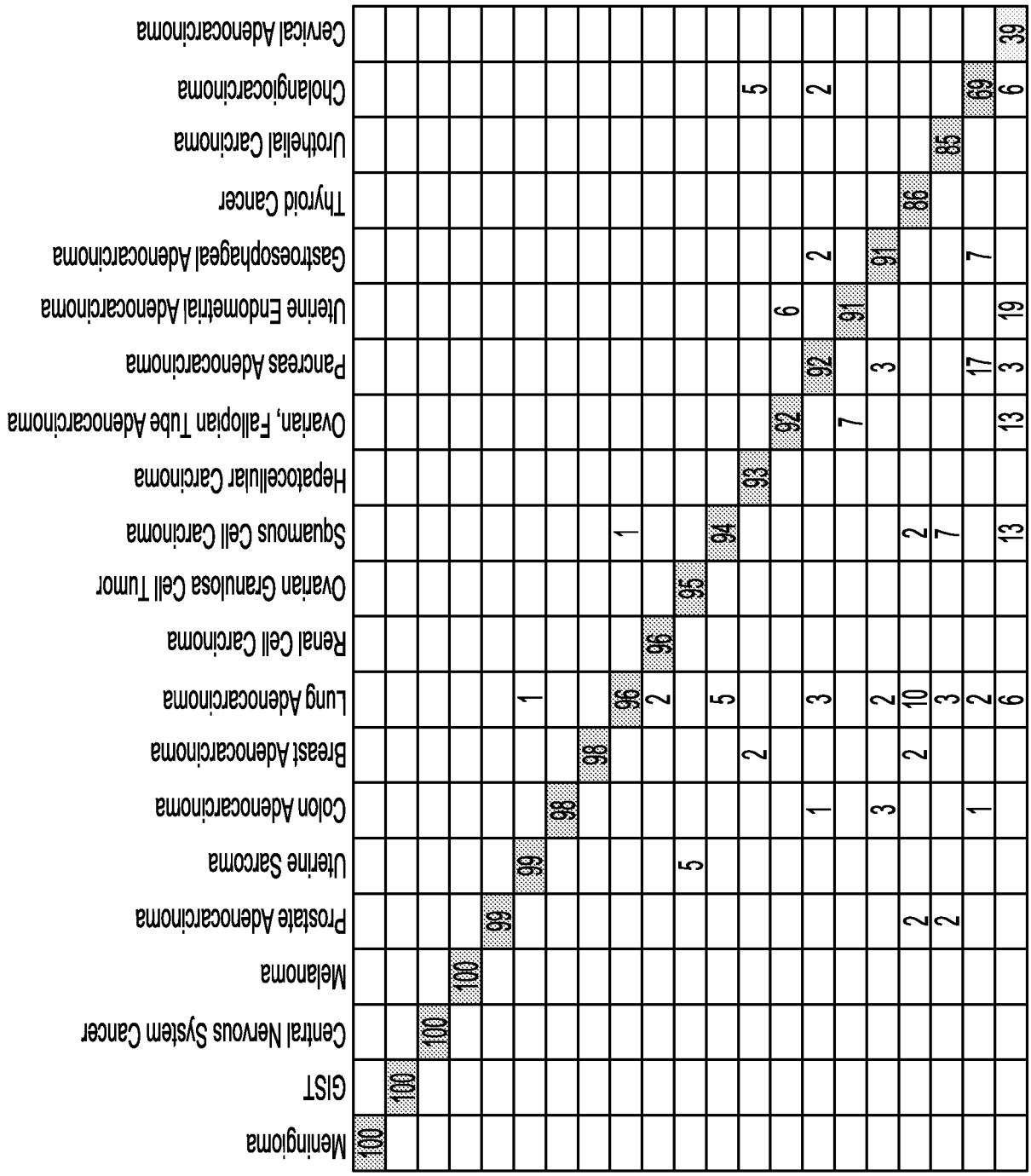

Meningioma
GIST
Central Nervous System Cancer
Melanoma
Prostate Adenocarcinoma
Uterine Sarcoma
Colon Adenocarcinoma
Breast Adenocarcinoma
Lung Adenocarcinoma
Renal Cell Carcinoma
Ovarian Granulosa Cell Tumor
Squamous Cell Carcinoma
Hepatocellular Carcinoma
Ovarian, Fallopian Tube Adenocarcinoma
Pancreas Adenocarcinoma
Uterine Endometrial Adenocarcinoma
Gastroesophageal Adenocarcinoma
Thyroid Cancer
Urothelial Carcinoma
Cholangiocarcinoma
Cervical Adenocarcinoma

FIG. 6AF

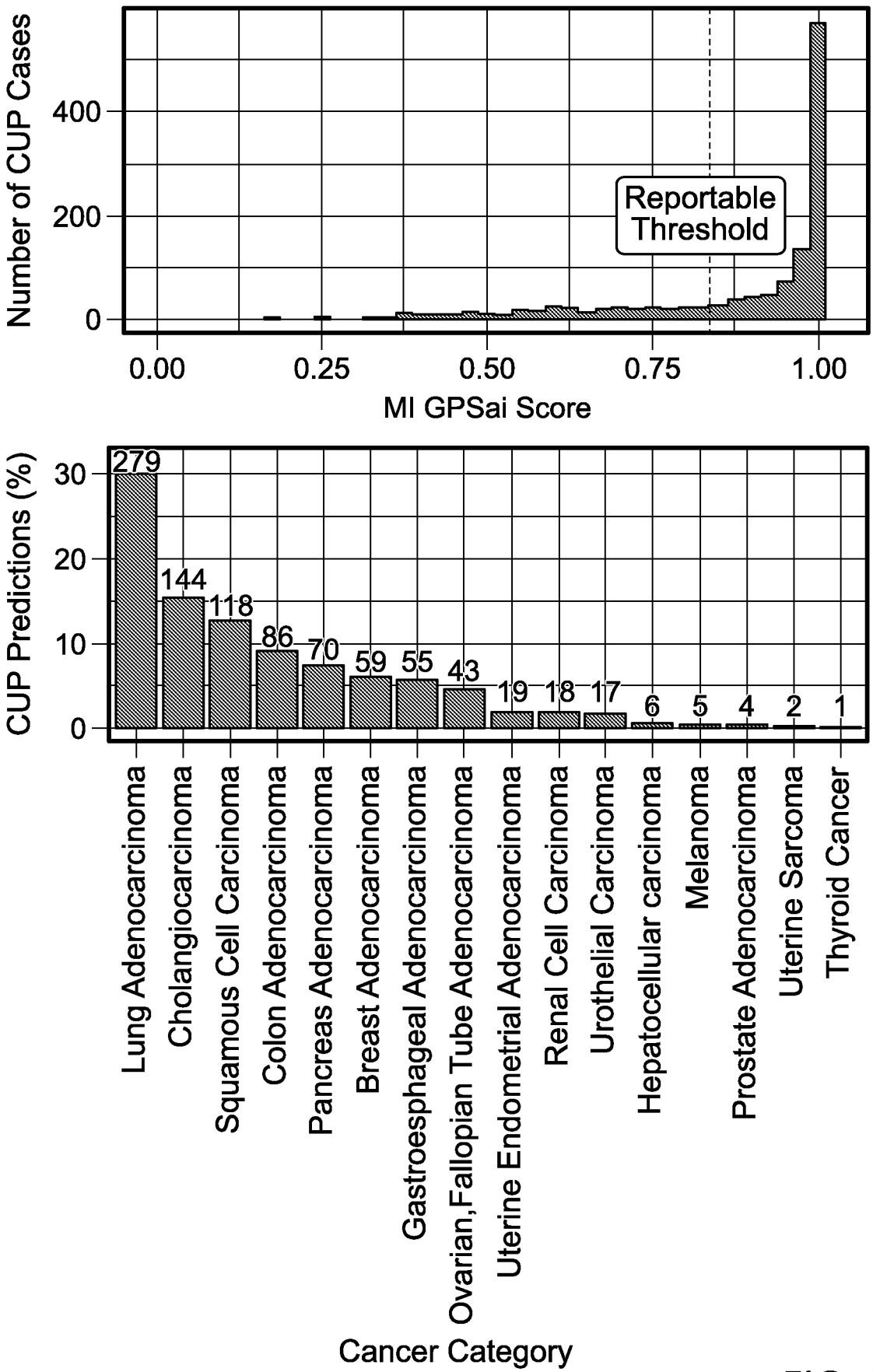

Number of CUP Cases
400
200
0
Reportable Threshold
0.00
0.25
0.50
0.75
1.00
MI GPSai Score
CUP Predictions (%)
30
20
10
0
279
144
118
86
70
59
55
43
19
18
17
6
5
4
2
1
Lung Adenocarcinoma
Cholangiocarcinoma
Squamous Cell Carcinoma
Colon Adenocarcinoma
Pancreas Adenocarcinoma
Breast Adenocarcinoma
Gastroesphageal Adenocarcinoma
Ovarian,Fallopian Tube Adenocarcinoma
Uterine Endometrial Adenocarcinoma
Renal Cell Carcinoma
Urothelial Carcinoma
Hepatocellular carcinoma
Melanoma
Prostate Adenocarcinoma
Uterine Sarcoma
Thyroid Cancer
Cancer Category

FIG. 6AG

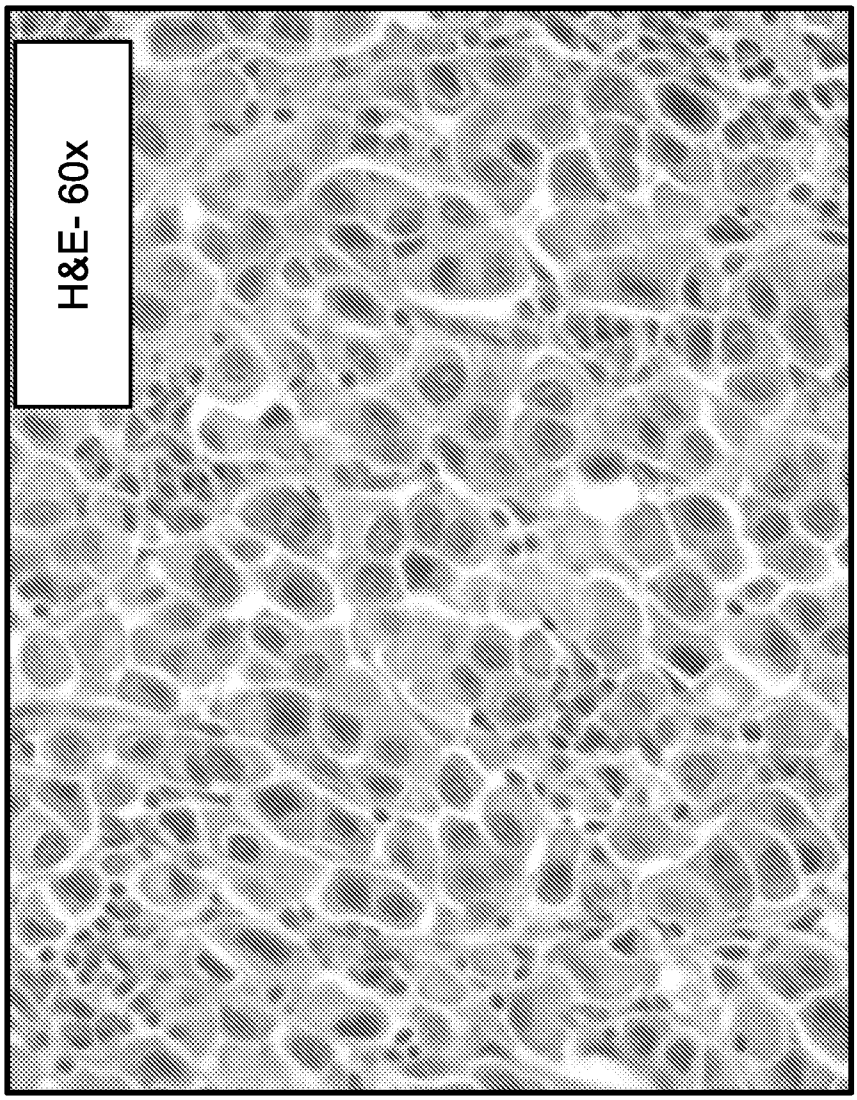


FIG. 6AH

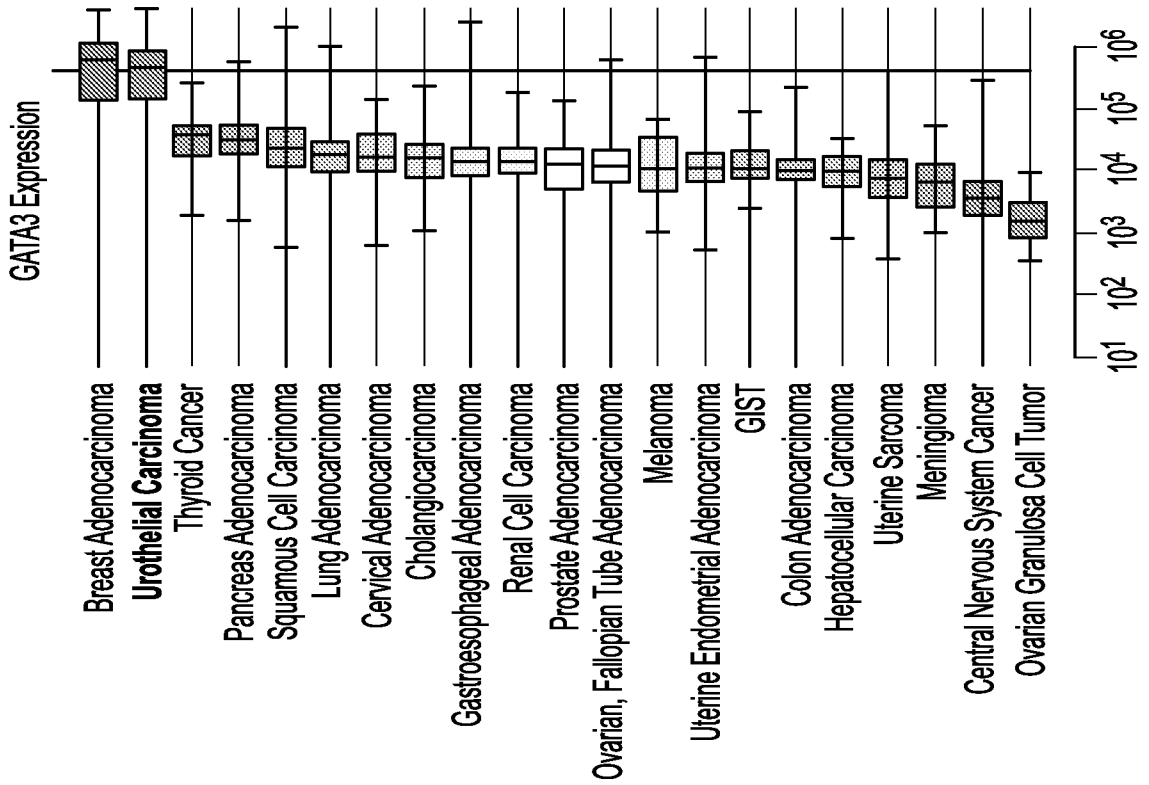


FIG. 6AJ

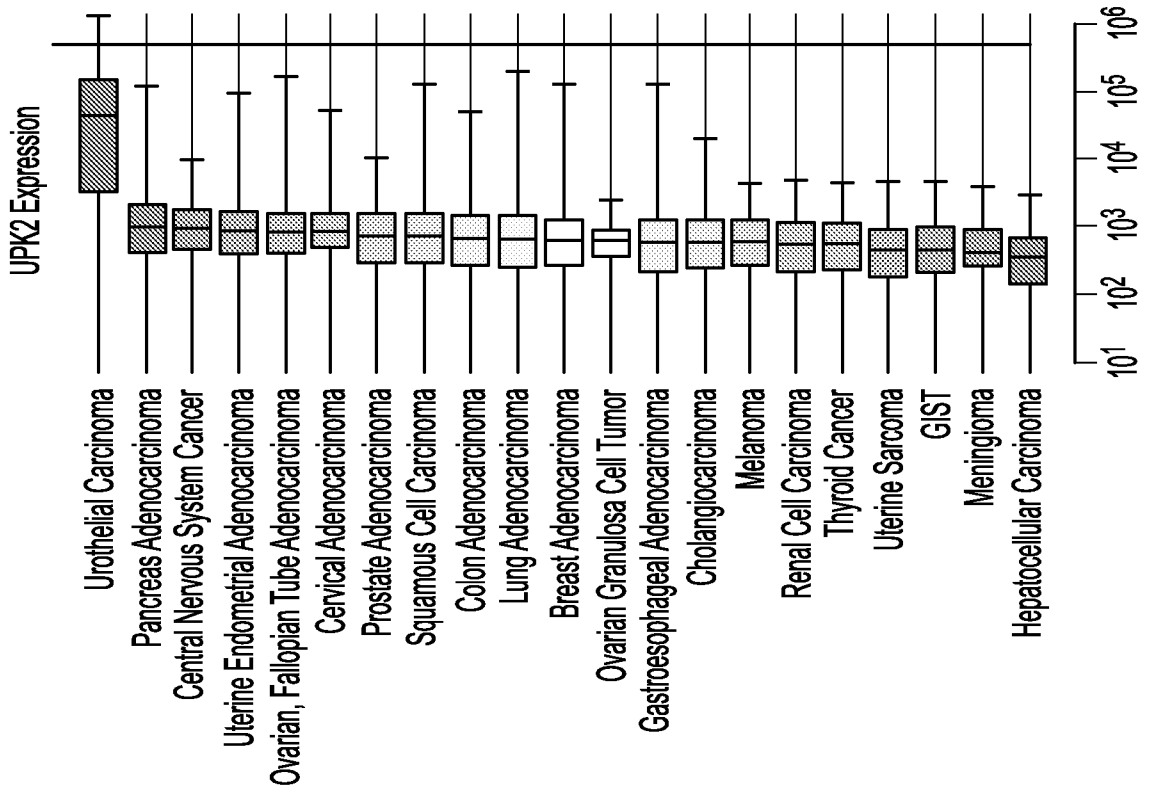


FIG. 6AI

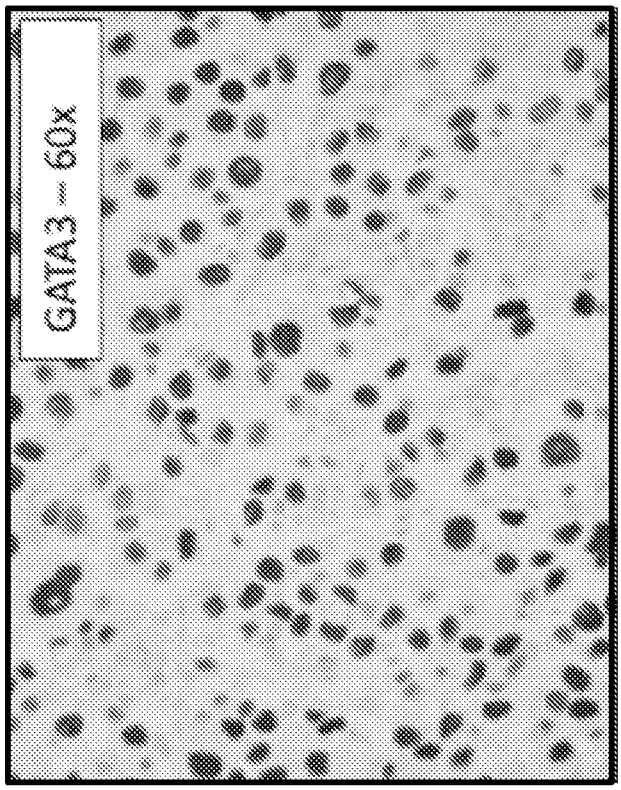


FIG. 6AL

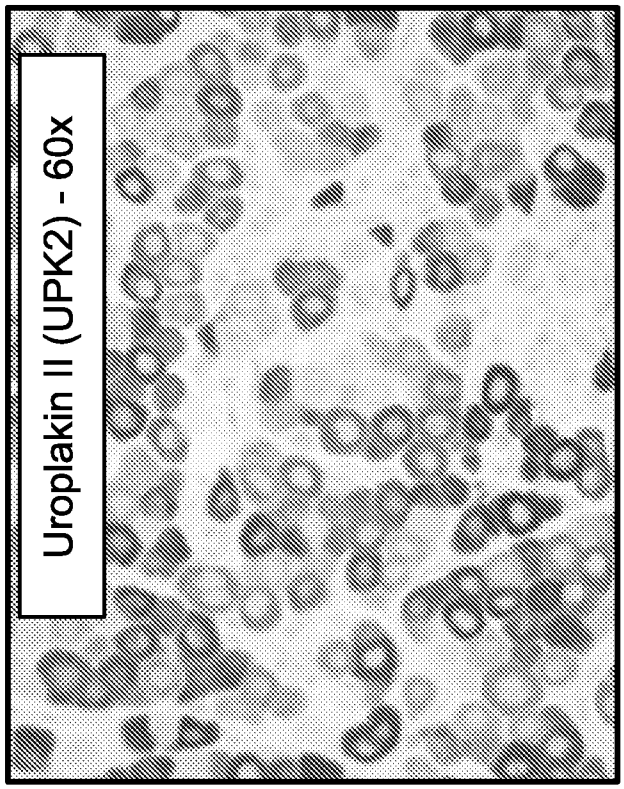


FIG. 6AK

Patient

Name: PATIENT 7, MALE
Date of Birth: 11/09/1940
Sex: Male
Case Number:
Diagnosis: Adenocarcinoma, metastatic, NOS

Specimen Information

Primary Tumor Site: Ascending colon
Specimen Site: Liver
Specimen ID: TestID123
Specimen Collected: 28-May-2019
All Testing Completed: 13-Dec-2019

Ordered By

Test Test Physician 2
Test Account 1
1234 Test Avenue
Test, AZ 00000 USA
(000) 000-0000

Results with Therapy Associations

Biomarker	Method	Alteration	Result	Therapy Association		Biomarker Tier*
KRAS	Seq	DNA-Tumor	Mutation Not Detected	Resistant	cetuximab, panitumumab	Level 1
NRAS	Seq	DNA-Tumor	Mutation Not Detected			
BRAF	Seq	DNA-Tumor	Mutation Not Detected			Level 2
ERBB2 (Her2/Neu)	IHC	Protein	Negative [0]	Sensitive / Response	lapatinib, pertuzumab, trastuzumab	Level 3A

* Biomarker reporting classification: Level 1 - highest level of clinical evidence and/or biomarker association included on the assay label; Level 2 - strong evidence of clinical significance and is endorsed by standard clinical guidelines; Level 3 - potential clinical significance (3A - evidence exists in patient's tumor type; 3B - evidence exists in another tumor type).

Cancer Type Relevant Biomarkers

Biomarker	Method	Alteration	Result
MSI	Seq	DNA-Tumor	Stable
Mismatch Repair Status		Protein	Proficient
NTRK1/2/3	Seq	RNA-Tumor	Fusion Not Detected
Tumor Mutational Burden		DNA-Tumor	Low [6 Mutations/Mb]
ERBB2 (Her2/Neu)	CNA-Seq	DNA-Tumor	Amplification Not Detected

Biomarker	Method	Alteration	Result
PIK3CA	Seq	DNA-Tumor	Mutation Not Detected
PTEN	IHC	Protein	Positive [1+, 100%]
PD-L1 (SP142)	IHC	Protein	Negative [0]

Genomic Signatures

Biomarker	Method	Alteration	Result
Microsatellite Instability (MSI)	Seq	DNA-Tumor	Stable
Tumor Mutational Burden (TMB)	Seq	DNA-Tumor	Result: Low 6 Low 7 Intermediate 17 High

FIG. 7A

Genes Tested With Mutations/Alterations

Gene	Assay	Result	Interpretation	Exon	Nucleotide	Amino Acid	Coverage
APC	Seq	DNA-Tumor	Mutated, Pathogenic	p.E1309*	16	c.3961G>T	99
IFN3	Seq	DNA-Tumor	Mutated, Pathogenic	p.S12Cfs	1	c.35delC	140

Unclassified alterations for DNA sequencing can be found in the Appendix.
Normal nucleotide reviewed below and gene reference sequences can be found in the Appendix of this report.

Immunohistochemistry Results

Test	Result	Test	Result
ERBB2 (HER2/Neu)	Negative (0	PD-L1 (SP142)	Negative (0
MLH1	Positive (1+, 95%	PMS2	Positive (1+, 90%
MSH2	Positive (1+, 100%	PTEN	Positive (1+, 100%
MSH6	Positive (1+, 95%		

Other Findings

ARID1A	ASXL1	ATRX	DNMT3A	KMT2D	MUTYH	NF1	TERT[1]

[1]TERT: Promoter region not sequenced

* Genes in this table were called indeterminate due to low coverage for some or all exons. Please see Appendix for a complete list of indeterminate genes.

FIG. 7B

Other Findings

ABL1	BRAF	CDK6	ERBB3	FGFR3	HGF/MET	KMT2C	MDM2	NTRK2	PPARG	RUNX1	SPEN
AKT1	BRCA1	CDKN1B	ERBB4	FGFR4	HNF1A	KMT2D	MED12	NTRK3	PPP2R1A	SERPINB3	SRC
ALK	BRCA2	CDKN2A	EZH2	FH	HRAS	KRAS	MITF	PAK3	PREX2	SETD2	STAG2
AMER1	BRD4	CHD4	ESR1	FLCN	IDH1	LCK	MLH1	PBRM1	PRKAR1A	SF3B1	STAT3
AR	BTK	CHEK2	EZH2	FLT1	IDH2	MAP2K1 (MEK1)	MNX1	PDGFRA	PRKDC	SMAD4	STK11
ARAF	CARD11	CIC	FANCA	FLT3	IGF1R	MAP3K1 (MEKK1)	MTOR	PDGFRB	PTCH1	SMARCA4	SUFU
ARID1A	CBFB	CREBBP	FANCC	FOXL2	JAK1	MAP2K4	NBN	PIK3CA	PTEN	SMARCB1	SYK
ATM	CCND1	CSF1R	FAT1	FUBP1	JAK2	MAP3K1	NF1	PIK3CB	PTPN11	SMO	TBX3
ATR	CCND2	CTNNB1	FBXW7	GATA3	JAK3	MAX	NFE2L2	PIK3R1	RAC1	SOCS1	TERC
ATRX	CCND3	CYLD	FGF3	GNA11	KDM5C	MCL1	NF1	PMS1	RAF1	SOX17	TERT
AURKA	CDH1	DDR2	FGF4	GNAQ	KDM6A	MDM2	NOTCH1	PMS2	RB1	SPEN	TET2
AURKB	CDK12	DICER1	FGF6	GNAS	KEAP1	MDM4	NOTCH2	POLD1	RBM10	SPOP	TGFBR2
AXIN1	CDK4	DNMT3A	FGFR1	GPS2	KIT	MED12	NOTCH3	POLE	REL	STAG2	TNFAIP3
AXIN2	CDK6	EGFR	FGFR2	GSK3B	KLF4	MEF2B	NPM1	PPP2R1A	RET	STAT3	TNFRSF14
B2M	CDKN1A	EP300	FGFR3	H3F3A	KMT2A	MEN1	NRAS	PPP6C	RHOB	STK11	TP53
BAP1	CDKN1B	ERBB2	FGFR4	HGF	KMT2C	MET	NSD1	PREX2	RHOT1	SUFU	TSC1
BCL2	CDKN2A	ERBB3	FLCN	HNF1A	KMT2D	MITF	NTRK1	PRKAR1A	RNF43	SYK	TSC2
BCL6	CEBPA	ERBB4	FLT3	HRAS	KRAS	MLH1	NTRK2	PTCH1	ROS1	TBX3	VHL
BCOR	CHD4	ESR1	FOXL2	IDH1	LCK	MPL	NTRK3	PTEN	RUNX1	TERC	WT1
BRAF	CHEK2	EZH2	FUBP1	IDH2	LMO1	MTOR	PAK3	PTPN11	SDHA	TERT	
BRCA1	CIC	FAM46C	GATA3	IGF1R	MAP2K1 (MEK1)	MUTYH	PAX5	RAC1	SDHB	TET2	
BRCA2	CREBBP	FANCA	GNA11	IKZF1	MAP2K4	MYC	PBRM1	RAD21	SDHC	TGFBR2	
BRD4	CRKL	FANCC	GNAQ	IL7R	MAP3K1	MYCL	PDGFRA	RAF1	SDHD	TNFAIP3	
BTG1	CSF1R	FAT1	GNAS	JAK1	MCL1	MYCN	PDGFRB	RB1	SERPINB3	TNFRSF14	
BTK	CTCF	FBXW7	GPS2	JAK2	MDM2	NBN	PIK3CA	RBM10	SETD2	TP53	
CARD11	CTNNB1	FGF3	GSK3B	JAK3	MDM4	NF1	PIK3CB	RET	SF3B1	TSC1	
CBFB	CYLD	FGF4	H3F3A	KDM5C	MED12	NF2	PIK3R1	RHOB	SMAD2	TSC2	
CCND1	DDR2	FGF6	HGF	KDM6A	MEF2B	NFE2L2	PMS1	RHOT1	SMAD4	U2AF1	
CCND2	DICER1	FGFR1	HNF1A	KEAP1	MEN1	NOTCH1	PMS2	RNF43	SMARCA4	VHL	
CCND3	DNMT3A	FGFR2	HRAS	KIT	MET	NOTCH2	POLD1	ROS1	SMARCB1	WT1	
CDH1	EGFR	FGFR3	IDH1	KLF4	MITF	NOTCH3	POLE	RUNX1	SMO	XPO1	
CDK12	EP300	FGFR4	IDH2	KMT2A	MLH1	NPM1	PPP2R1A	SDHA	SOCS1	YAP1	
CDK4	ERBB2	FLCN	IGF1R	KMT2C	MPL	NRAS	PPP6C	SDHB	SOX17	ZFHX3	
CDK6	ERBB3	FLT3	IKZF1	KMT2D	MTOR	NSD1	PREX2	SDHC	SPEN		
CDKN1A	ERBB4	FOXL2	IL7R	KRAS	MUTYH	NTRK1	PRKAR1A	SDHD	SPOP		
CDKN1B	ESR1	FUBP1	JAK1	LCK	MYC	NTRK2	PTCH1	SERPINB3	SRC		
CDKN2A	EZH2	GATA3	JAK2	LMO1	MYCL	NTRK3	PTEN	SETD2	STAG2		
CEBPA	FAM46C	GNA11	JAK3	MAP2K1 (MEK1)	MYCN	PAK3	PTPN11	SF3B1	STAT3		
CHD4	FANCA	GNAQ	KDM5C	MAP2K4	NBN	PAX5	RAC1	SMAD2	STK11		
CHEK2	FANCC	GNAS	KDM6A	MAP3K1	NF1	PBRM1	RAD21	SMAD4	SUFU		
CIC	FAT1	GPS2	KEAP1	MCL1	NF2	PDGFRA	RAF1	SMARCA4	SYK		
CREBBP	FBXW7	GSK3B	KIT	MDM2	NFE2L2	PDGFRB	RB1	SMARCB1	TBX3		
CRKL	FGF3	H3F3A	KLF4	MDM4	NOTCH1	PIK3CA	RBM10	SMO	TERC		
CSF1R	FGF4	HGF	KMT2A	MED12	NOTCH2	PIK3CB	RET	SOCS1	TERT		
CTCF	FGF6	HNF1A	KMT2C	MEF2B	NOTCH3	PIK3R1	RHOB	SOX17	TET2		
CTNNB1	FGFR1	HRAS	KMT2D	MEN1	NPM1	PMS1	RHOT1	SPEN	TGFBR2		
CYLD	FGFR2	IDH1	KRAS	MET	NRAS	PMS2	RNF43	SPOP	TNFAIP3		
DDR2	FGFR3	IDH2	LCK	MITF	NSD1	POLD1	ROS1	SRC	TNFRSF14		
DICER1	FGFR4	IGF1R	LMO1	MLH1	NTRK1	POLE	RUNX1	STAG2	TP53		
DNMT3A	FLCN	IKZF1	MAP2K1 (MEK1)	MPL	NTRK2	PPARG	SDHA	STAT3	TSC1		
EGFR	FLT3	IL7R	MAP2K4	MTOR	NTRK3	PPP2R1A	SDHB	STK11	TSC2		
EP300	FOXL2	JAK1	MAP3K1	MUTYH	PAK3	PREX2	SDHC	SUFU	VHL		
ERBB2	FUBP1	JAK2	MCL1	MYC	PAX5	PRKAR1A	SDHD	SYK	WT1		
ERBB3	GATA3	JAK3	MDM2	MYCL	PBRM1	PTCH1	SERPINB3	TBX3	XPO1		
ERBB4	GNA11	KDM5C	MDM4	MYCN	PDGFRA	PTEN	SETD2	TERC	YAP1		
ESR1	GNAQ	KDM6A	MED12	NBN	PDGFRB	PTPN11	SF3B1	TERT	ZFHX3		
EZH2	GNAS	KEAP1	MEF2B	NF1	PIK3CA	RAC1	SMAD2	TET2			
FAM46C	GPS2	KIT	MEN1	NF2	PIK3CB	RAD21	SMAD4	TGFBR2			
FANCA	GSK3B	KLF4	MET	NFE2L2	PIK3R1	RAF1	SMARCA4	TNFAIP3			
FANCC	H3F3A	KMT2A	MITF	NOTCH1	PMS1	RB1	SMARCB1	TNFRSF14			
FAT1	HGF	KMT2C	MLH1	NOTCH2	PMS2	RBM10	SMO	TP53			
FBXW7	HNF1A	KMT2D	MPL	NOTCH3	POLD1	RET	SOCS1	TSC1			
FGF3	HRAS	KRAS	MTOR	NPM1	POLE	RHOB	SOX17	TSC2			
FGF4	IDH1	LCK	MUTYH	NRAS	PPARG	RHOT1	SPEN	VHL			
FGF6	IDH2	LMO1	MYC	NSD1	PPP2R1A	RNF43	SPOP	WT1			
FGFR1	IGF1R	MAP2K1 (MEK1)	MYCL	NTRK1	PREX2	ROS1	SRC	XPO1			
FGFR2	IKZF1	MAP2K4	MYCN	NTRK2	PRKAR1A	RUNX1	STAG2	YAP1			
FGFR3	IL7R	MAP3K1	NBN	NTRK3	PTCH1	SDHA	STAT3	ZFHX3			
FGFR4	JAK1	MCL1	NF1	PAK3	PTEN	SDHB	STK11				
FLCN	JAK2	MDM2	NF2	PAX5	PTPN11	SDHC	SUFU				
FLT3	JAK3	MDM4	NFE2L2	PBRM1	RAC1	SDHD	SYK				
FOXL2	KDM5C	MED12	NOTCH1	PDGFRA	RAD21	SERPINB3	TBX3				
FUBP1	KDM6A	MEF2B	NOTCH2	PDGFRB	RAF1	SETD2	TERC				
GATA3	KEAP1	MEN1	NOTCH3	PIK3CA	RB1	SF3B1	TERT				
GNA11	KIT	MET	NPM1	PIK3CB	RBM10	SMAD2	TET2				
GNAQ	KLF4	MITF	NRAS	PIK3R1	RET	SMAD4	TGFBR2				
GNAS	KMT2A	MLH1	NSD1	PMS1	RHOB	SMARCA4	TNFAIP3				
GPS2	KMT2C	MPL	NTRK1	PMS2	RHOT1	SMARCB1	TNFRSF14				
GSK3B	KMT2D	MTOR	NTRK2	POLD1	RNF43	SMO	TP53				
H3F3A	KRAS	MUTYH	NTRK3	POLE	ROS1	SOCS1	TSC1				
HGF	LCK	MYC	PAK3	PPARG	RUNX1	SOX17	TSC2				
HNF1A	LMO1	MYCL	PAX5	PPP2R1A	SDHA	SPEN	VHL				
HRAS	MAP2K1 (MEK1)	MYCN	PBRM1	PPP6C	SDHB	SPOP	WT1				
IDH1	MAP2K4	NBN	PDGFRA	PREX2	SDHC	SRC	XPO1				
IDH2	MAP3K1	NF1	PDGFRB	PRKAR1A	SDHD	STAG2	YAP1				
IGF1R	MCL1	NF2	PIK3CA	PTCH1	SERPINB3	STAT3	ZFHX3				
IKZF1	MDM2	NFE2L2	PIK3CB	PTEN	SETD2	STK11					
IL7R	MDM4	NOTCH1	PIK3R1	PTPN11	SF3B1	SUFU					
JAK1	MED12	NOTCH2	PMS1	RAC1	SMAD2	SYK					
JAK2	MEF2B	NOTCH3	PMS2	RAD21	SMAD4	TBX3					
JAK3	MEN1	NPM1	POLD1	RAF1	SMARCA4	TERC					
KDM5C	MET	NRAS	POLE	RB1	SMARCB1	TERT					
KDM6A	MITF	NSD1	PPARG	RBM10	SMO	TET2					
KEAP1	MLH1	NTRK1	PPP2R1A	RET	SOCS1	TGFBR2					
KIT	MPL	NTRK2	PPP6C	RHOB	SOX17	TNFAIP3					
KLF4	MTOR	NTRK3	PREX2	RHOT1	SPEN	TNFRSF14					
KMT2A	MUTYH	PAK3	PRKAR1A	RNF43	SPOP	TP53					
KMT2C	MYC	PAX5	PTCH1	ROS1	SRC	TSC1					
KMT2D	MYCL	PBRM1	PTEN	RUNX1	STAG2	TSC2					
KRAS	MYCN	PDGFRA	PTPN11	SDHA	STAT3	VHL					
LCK	NBN	PDGFRB	RAC1	SDHB	STK11	WT1					
LMO1	NF1	PIK3CA	RAD21	SDHC	SUFU	XPO1					
MAP2K1 (MEK1)	NF2	PIK3CB	RAF1	SDHD	SYK	YAP1					
MAP2K4	NFE2L2	PIK3R1	RB1	SERPINB3	TBX3	ZFHX3					
MAP3K1	NOTCH1	PMS1	RBM10	SETD2	TERC						
MCL1	NOTCH2	PMS2	RET	SF3B1	TERT						
MDM2	NOTCH3	POLD1	RHOB	SMAD2	TET2						
MDM4	NPM1	POLE	RHOT1	SMAD4	TGFBR2						
MED12	NRAS	PPARG	RNF43	SMARCA4	TNFAIP3						
MEF2B	NSD1	PPP2R1A	ROS1	SMARCB1	TNFRSF14						
MEN1	NTRK1	PPP6C	RUNX1	SMO	TP53						
MET	NTRK2	PREX2	SDHA	SOCS1	TSC1						
MITF	NTRK3	PRKAR1A	SDHB	SOX17	TSC2						
MLH1	PAK3	PTCH1	SDHC	SPEN	VHL						
MPL	PAX5	PTEN	SDHD	SPOP	WT1						
MTOR	PBRM1	PTPN11	SERPINB3	SRC	XPO1						
MUTYH	PDGFRA	RAC1	SETD2	STAG2	YAP1						
MYC	PDGFRB	RAD21	SF3B1	STAT3	ZFHX3						
MYCL	PIK3CA	RAF1	SMAD2	STK11							
MYCN	PIK3CB	RB1	SMAD4	SUFU							
NBN	PIK3R1	RBM10	SMARCA4	SYK							
NF1	PMS1	RET	SMARCB1	TBX3							
NF2	PMS2	RHOB	SMO	TERC							
NFE2L2	POLD1	RHOT1	SOCS1	TERT							
NOTCH1	POLE	RNF43	SOX17	TET2							
NOTCH2	PPARG	ROS1	SPEN	TGFBR2							
NOTCH3	PPP2R1A	RUNX1	SPOP	TNFAIP3							
NPM1	PPP6C	SDHA	SRC	TNFRSF14							
NRAS	PREX2	SDHB	STAG2	TP53							
NSD1	PRKAR1A	SDHC	STAT3	TSC1							
NTRK1	PTCH1	SDHD	STK11	TSC2							
NTRK2	PTEN	SERPINB3	SUFU	VHL							
NTRK3	PTPN11	SETD2	SYK	WT1							
PAK3	RAC1	SF3B1	TBX3	XPO1							
PAX5	RAD21	SMAD2	TERC	YAP1							
PBRM1	RAF1	SMAD4	TERT	ZFHX3							
PDGFRA	RB1	SMARCA4	TET2								
PDGFRB	RBM10	SMARCB1	TGFBR2								
PIK3CA	RET	SMO	TNFAIP3								
PIK3CB	RHOB	SOCS1	TNFRSF14								
PIK3R1	RHOT1	SOX17	TP53								
PMS1	RNF43	SPEN	TSC1								
PMS2	ROS1	SPOP	TSC2								
POLD1	RUNX1	SRC	VHL								
POLE	SDHA	STAG2	WT1								
PPARG	SDHB	STAT3	XPO1								
PPP2R1A	SDHC	STK11	YAP1								
PPP6C	SDHD	SUFU	ZFHX3								
PREX2	SERPINB3	SYK									
PRKAR1A	SETD2	TBX3									
PTCH1	SF3B1	TERC									
PTEN	SMAD2	TERT									
PTPN11	SMAD4	TET2									
RAC1	SMARCA4	TGFBR2									
RAD21	SMARCB1	TNFAIP3									
RAF1	SMO	TNFRSF14									
RB1	SOCS1	TP53									
RBM10	SOX17	TSC1									
RET	SPEN	TSC2									
RHOB	SPOP	VHL									
RHOT1	SRC	WT1									
RNF43	STAG2	XPO1									
ROS1	STAT3	YAP1									
RUNX1	STK11	ZFHX3									
SDHA	SUFU										
SDHB	SYK										
SDHC	TBX3										
SDHD	TERC										
SERPINB3	TERT										
SETD2	TET2										
SF3B1	TGFBR2										
SMAD2	TNFAIP3										
SMAD4	TNFRSF14										
SMARCA4	TP53										
SMARCB1	TSC1										
SMO	TSC2										
SOCS1	VHL										
SOX17	WT1										
SPEN	XPO1										
SPOP	YAP1										
SRC	ZFHX3										
STAG2											
STAT3											
STK11											
SUFU											
SYK											
TBX3											
TERC											
TERT											
TET2											
TGFBR2											
TNFAIP3											
TNFRSF14											
TP53											
TSC1											
TSC2											
VHL											
WT1											
XPO1											
YAP1											
ZFHX3											

FIG. 7C

Other Findings

ABL1	ECR	ERG	ETV6	FGR	MET	NOTCH2	NUMBL	PIK3CA	PRKCB	ROS1	TFE3
AKT3	BRAF	ESR1	EWSR1	JAK2	MSMR	NRG1	OLFM4	PML	RAF1	RSPO2	TFFR
ALK	BRD3	ETV1	FGFR1	MAML2	MUSK	NTRK1	PDGFRA	PPARG	RELA	RSPO3	THADA
ARHGAP26	BRD4	ETV4	FGFR2	MAST1	MYB	NTRK2	PDGFRB	PRKCA	RET	TERT	TMPRSS2
AXL	EGFR	ETV5	FGFR3	MAST2	NOTCH1	NTRK3					

AR	EGFRvIII	MET

The complete list of unclassified alterations for RNA Whole Transcriptome Sequencing are available by request.

FIG. 7D

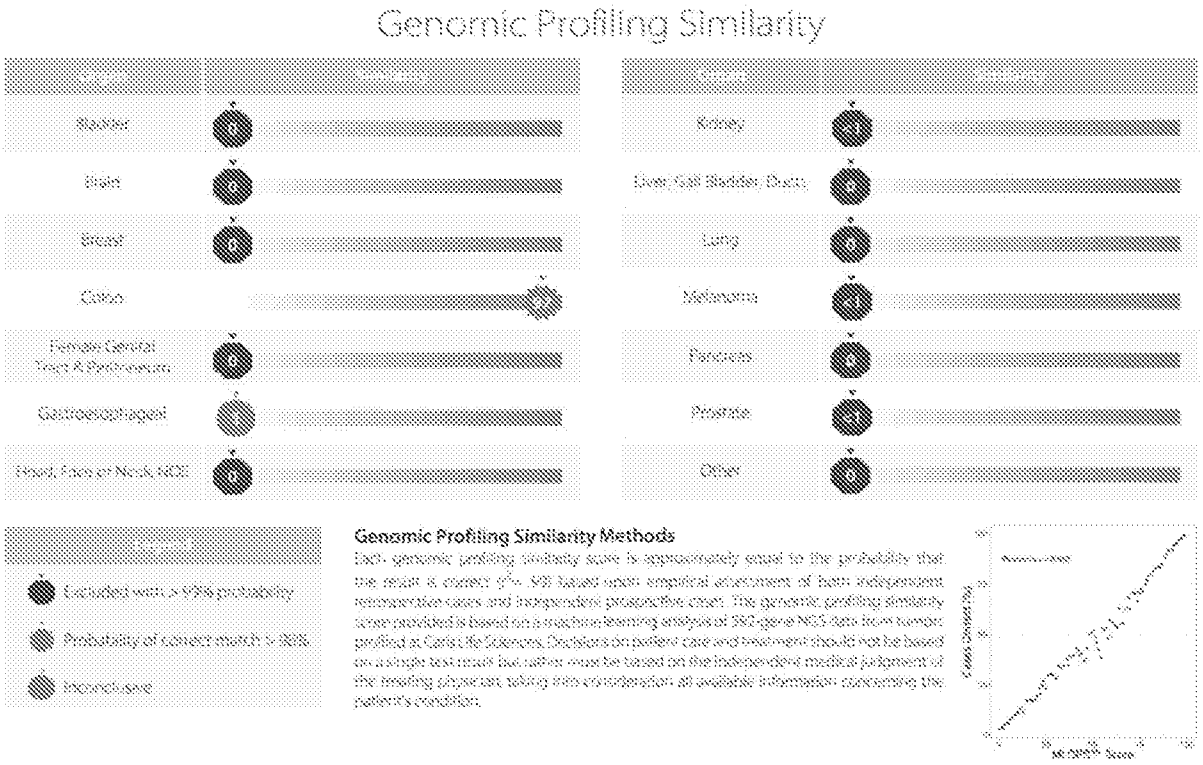

Genomic Profiling Similarity
Bladder
Brain
Breast
Colon
Female Genital Tract & Peritoneum
Gastroesophageal
Head, Face or Neck (HDS)
Kidney
Liver, Gall Bladder, Duct
Lung
Melanoma
Pancreas
Prostate
Other
Excluded with a 95% probability
Probability of tumor match > 80%
Inconclusive
Genomic Profiling Similarity Methods
Each genomic profiling similarity score is approximately equal to the probability that the result is correct (e.g., 95 based upon empirical assessment of both independent retrospective cases and independent prospective cases. The genomic profiling similarity score provided is based on a machine learning analysis of 592-gene NGS data from tumors profiled at Caris Life Sciences. Decisions on patient care and treatment should not be based on a single test result but rather must be based on the independent medical judgment of the treating physician taking into consideration all available information concerning the patient's condition.
Consultation
The MI GPS Score result for colon suggests this tumor is consistent with a colon primary. However, after review of the available clinical and pathological data, there is insufficient evidence to conclusively support this genomic similarity.

FIG. 7E

Notes of Significance

[redacted]

Clinical Trials Connector ™ opportunities based on biomarker expression: 1 Targeted Therapy Trials. See page 2 for details.

Specimen Information

Specimen ID: TestID123

Specimen Collected: 05/28/2019

Specimen Received: 12/13/2019

Other Testing Initiated: 12/13/2019

Gross Description: 1 (A) Paraffin Block - Client ID(TestID123) from Test Account 1, Test, AZ.

Pathologic Diagnosis: Liver, segment V and VII, removal: Metastatic adenocarcinoma consistent with history of colorectal primary.

Dissection Information: Molecular testing of this specimen was performed after harvesting of targeted tissues with an approved manual microdissection technique. Candidate slides were examined under a microscope and areas containing tumor cells (and separately normal cells, when necessary for testing) were circled. A laboratory technician harvested targeted tissues for extraction from the marked areas using a dissection microscope. The areas marked and extracted were microscopically reexamined on post-microdissected slides and adequacy of microdissection was reviewed by a board certified Pathologist.

FIG. 7F

Clinical Trials Connector™

For a complete list of open, enrolling clinical trials visit MI Portal to access the Clinical Trials Connector. This personalized, real-time web-based service provides additional clinical trial information and enhanced searching capabilities, including, but not limited to:

- Location: filter by geographic area
- Biomarker(s): identify specific biomarkers associated with open clinical trials to choose from
- Drug(s): search for specific therapies
- Trial Sponsor: locate trials based on the organization supporting the trial(s)

Visit www.CarisMolecularIntelligence.com to view all matched trials. Therapeutic agents listed below may or may not be currently FDA approved for the tumor type tested.

TARGETED THERAPY CLINICAL TRIALS (11)				
Drug Class	Biomarker	Biomarker	Tumor	Investigational Therapeutic
Wnt pathway inhibitors (1)	APC	WT1	GBM-Tumor	LGK974

( ) = represents the total number of clinical trials identified by the Clinical Trials Connector for the provided drug class or table.

Please refer to the "Notes of Significance" section that may contain additional information regarding therapy associations.

FIG. 7G

Disclaimer

Decisions on patient care and treatment must be based on the independent medical judgment of the treating physician, taking into consideration all available information concerning the patient's condition, the FDA prescribing information for any therapeutic, and in accordance with the applicable standard of care. Drug associations provided in this report do not guarantee that any particular agent will be effective for the treatment of any patient or for any particular condition. Caris Life Sciences expressly disclaims and makes no representation or warranty whatsoever relating, directly or indirectly, to the conclusions drawn from its review of scientific literature, including information and conclusions relating to therapies that are included or omitted from this report. Whether or not a particular patient will benefit from a selected therapy is based on many factors and can vary significantly. The selection of therapy, if any, resides solely in the discretion of the treating physician.

Individual assays that are available through Caris Molecular Intelligence® include both Laboratory Developed Tests (LDT) and U.S. Food and Drug Administration (FDA) approved or cleared tests. Caris MPI, Inc. d/b/a Caris Life Sciences® is certified under the Clinical Laboratory Improvement Amendments (CLIA) as qualified to perform high complexity clinical laboratory testing, including all of the assays that comprise the Caris Molecular Intelligence®. The LDTs were developed and their performance characteristics determined by Caris. The LDTs have not been cleared or approved by the U.S. Food and Drug Administration. Caris' CLIA certification number is located at the bottom of each page of this report. Certain tests have not been cleared or approved by the FDA. The FDA has determined that clearance or approval is not necessary for certain laboratory developed tests. Caris LDTs are used for clinical purposes. They are not investigational or for research.

The information presented in the Clinical Trials Connector™ section of this report, if applicable, is compiled from sources believed to be reliable and current. However, the accuracy and completeness of the information provided herein cannot be guaranteed. The clinical trials information present in the biomarker description was compiled from www.clinicaltrials.gov. The contents are to be used only as a guide, and health care providers should employ their best comprehensive judgment in interpreting this information for a particular patient. Specific eligibility criteria for each clinical trial should be reviewed as additional inclusion criteria may apply.

Caris Molecular Intelligence is subject to Caris' intellectual property. Patent www.carislifesciences.com/ip.

FIG. 7H

Mutational Analysis by Next-Generation Sequencing (NGS)

TUMOR MUTATIONAL BURDEN	
Mutations/Megabase (Mb)	Result
6	Low

TMB Methods

Tumor Mutational Burden was performed based on Next Generation Sequencing (NGS) analysis from genomic DNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Illumina NextSeq platform. NGS was developed and its performance characteristics determined by Caris Life Sciences.

Tumor Mutational Burden is calculated using only missense mutations that have not been previously reported as germline alterations. A high mutational burden is a potential indicator of immunotherapy response (Le et al., NEJM, 2015; Rizvi et al, Science, 2015; Rosenberg et al, Lancet, 2016; Snyder et al., NEJM, 2014). Caris Life Sciences has defined threshold levels for Tumor Mutational Burden and establish cutoff points based on published evidence and internal data/experience:

- High: greater than or equal to 17 mutations/Megabase (≥17 mutations/Mb). Approximately 7% of Caris Molecular Intelligence cases reported a High result.
- Intermediate: greater than or equal to 7 but fewer than 17 mutations/ Megabase (≥7 and <17 mutations/Mb). Approximately 34% of Caris Molecular Intelligence cases reported an Intermediate result.
- Low: less than or equal to 6 mutations/Megabase (≤6 mutations/Mb). Approximately 59% of Caris Molecular Intelligence cases reported a Low result.

MICROSATELLITE INSTABILITY (MSI-NGS)		
Test	Result	Status
MSI	No microsatellite instability detected **Procedure:** NGS	Stable

Microsatellite Instability Analysis

Microsatellite instability status by NGS (MSI-NGS) is measured by the direct analysis of known microsatellite regions sequenced in the CMI 592 gene panel. To establish clinical thresholds, MSI-NGS results were compared with results from over 2,000 matching clinical cases analyzed with traditional, PCR-based methods. Genomic variants in the microsatellite loci are detected using the same depth and frequency criteria as used for mutation detection. Only insertions and deletions resulting in a change in the number of tandem repeats are considered in this assay. Some microsatellite regions with known polymorphisms or technical sequencing issues are excluded from the analysis. The total number of microsatellite alterations in each sample are counted and grouped into three categories: High, Equivocal and Stable. MSI-Low results are reported in the Stable category. Equivocal results have a total number of microsatellite alterations in between High and Stable.

FIG. 7I

Mutational Analysis by Next-Generation Sequencing (NGS)

Gene	Assay	Result/Interpretation	Protein Change	Exon	DNA Change	Allele Frequency	Reference Sequence
APC	DNA/Tumor	Mutated, Pathogenic	p.E1306*	16	c.3916G>T	78	NM_000038

Interpretation: A pathogenic mutation, p.E1306*, was detected in APC. Germline mutations in APC are causal for Familial Adenomatous Polyposis (FAP), a disorder associated with early-onset colorectal adenocarcinoma (Maller 2016 J Pediatr Genet 5:78).

APC or adenomatous polyposis coli is a key tumor suppressor gene that encodes for a large multi-domain protein. This protein exerts its tumor suppressor function in the Wnt/β-catenin cascade mainly by controlling the degradation of β-catenin, the central activator of transcription in the Wnt signaling pathway. The Wnt signaling pathway mediates important cellular functions including intercellular adhesion, stabilization of the cytoskeleton and cell cycle regulation and apoptosis, and it is important in embryonic development and oncogenesis. Mutation in APC results in a truncated protein product with abnormal function, lacking the domains involved in β-catenin degradation. Somatic mutation in the APC gene can be detected in the majority of colorectal tumors (80%) and it is an early event in colorectal tumorigenesis. APC wild type patients have shown better disease control rate in the metastatic setting when treated with oxaliplatin, while when treated with fluoropyrimidine regimens, APC wild type patients experience more hematological toxicities. APC mutation has also been identified in oral squamous cell carcinoma, gastric cancer as well as hepatoblastoma and may contribute to cancer formation. Germline mutation in APC causes familial adenomatous polyposis, which is an autosomal dominant inherited disease that will inevitably develop to colorectal cancer if left untreated. COX-2 inhibitors including celecoxib may reduce the recurrence of adenomas and incidence of advanced adenomas in individuals with an increased risk of CRC. Turcot syndrome and Gardner's syndrome have also been associated with germline APC defects. Germline mutations of the APC have also been associated with an increased risk of developing desmoid disease, papillary thyroid carcinoma and hepatoblastoma.

Gene	Assay	Result/Interpretation	Protein Change	Exon	DNA Change	Allele Frequency	Reference Sequence
TP53	DNA/Tumor	Mutated, Pathogenic	p.S127F	5	c.380C>T	30	NM_000546

Interpretation: This TP53 mutation, and other substitutions at the same position in the protein, have been reported in numerous tumors. The frequency of this mutation in cancers indicates it is pathogenic.

TP53, or p53, plays a central role in modulating response to cellular stress through transcriptional regulation of genes involved in cell-cycle arrest, DNA repair, apoptosis, and senescence. Inactivation of the p53 pathway is essential for the formation of the majority of human tumors. Mutation in p53 (TP53) remains one of the most commonly described genetic events in human neoplasia, estimated to occur in 30-50% of all cancers. Generally, presence of a disruptive p53 mutation is associated with a poor prognosis in all types of cancers and diminished sensitivity to radiation and chemotherapy. Germline p53 mutations are associated with the Li-Fraumeni syndrome (LFS) which may lead to early-onset of several forms of cancer currently known to occur in the syndrome, including sarcomas of the bone and soft tissues, carcinomas of the breast and adrenal cortex (hereditary adrenocortical carcinoma), brain tumors and acute leukemias.

The next generation sequencing assay performed by Caris Life Sciences examines nucleic acids obtained from tumor tissue only and does not examine normal tissue such as tumor adjacent tissue or whole or peripheral blood. As such, the origin of any mutation detected may be a somatic mutation (not inherited) or a germline mutation (inherited) and will not be distinguishable by this assay. It is recommended that results be considered within the patient's clinical and health history. If a germline inheritance pattern is suspected then counseling by a board certified genetic counselor is recommended.

* Any mutations in the above genes that are known by Caris to be clinically significant (i.e., pathogenic or presumed pathogenic) are reported with interpretation in the body of the report. Any remaining mutations are listed above and have not been classified by Caris.

FIG. 7J

Mutational Analysis by Next-Generation Sequencing (NGS)

Genes tested and reported in this mutational analysis of tumor tissue by NGS*					
ARID1A	ATRX	MED12	NF1		
ASXL1	FAT1	MUC4	TERT ¶		

* Genes in this table were ruled indeterminate due to low coverage for some or all exons.

¶ TERT Promoter region not sequenced

Genes tested and reported in this mutational analysis of tumor tissue by NGS*					
ABL1	CD79B	FANCA	HNF1A	MET	PDGFRB
ART1	CDC73	FANCC	HNF1A	MITF	PIK3CB
ALK	CDH1	FANCD2	HRAS	MLH1	PIK3CA
AMER1	CDK12	FANCF	IDH1	MPL	PIK3R1
AR	CDK4	FANCG	IDH2	MRE11	PMS1
ARAF	CDK6	FANCL	IKBKE	MSH2	PMS2
ARID2	CDKN1B	FAS	IKZF1	MSH6	PNRC2
ATM	CDKN2A	FBXW7	JAK2	MTOR	POLE
BAP1	CHEK1	FGFR1	JAK3	MUTYH	PTCH1
BARD1	CHEK2	FGFR2	KDM5C	MYCN	PTEN
BCL2	CIC	FGFR3	KDM6A	MYD88	PPP2R1A
BCL9	CREBBP	FGFR4	KDR (VEGFR2)	NBN	PRDM1
BCOR	CSF1R	FH	KEAP1	NF2	PIK3AR1A
BLM	CTNNB1	FLCN	KIT	NF2L2	PRKDC
BMPR1A	CYLD	FLT1	KMT2A	NOTCH1	PTCH1
BRAF	DDR2	FLT3	KMT2C	NOTCH1	PTEN
BRCA1	DNMT3A	FOXL2	KMT2D	NPM1	PTPN11
BRCA2	EGFR	FUBP1	KRAS	NRAS	RAD50
BRIP1	EP300	GATA3	LCK	NSD1	RAF1
BTK	ERBB2 (Her2/Neu)	GNA11	MAP2K1 (MEK1)	NTRK1	RB1
CARD11	ERBB3	GNA13	MAP2K2 (MEK2)	NTRK2	RET
CBFB	ERBB4	GNAQ	MAP3K1	NTRK3	RNF43
CCND1	ERCC2	GNAS	MAP3K1	PALB2	ROS1
CCND2	ESR1	HSD3A	MAX	PBRM1	RUNX1
CCNE1	EZH2	HSP90	MEN1	PDGFRA	SDHAF2

FIG. 7K

Mutational Analysis by Next-Generation Sequencing (NGS)

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SDHB	SMAD2	SMO	STAT3	TSC1	VHL
SDHC	SMAD4	SOCS1	STK11	TSC2	
SDHD	SMARCA4	SPEN	SUFU	USP8	
SETD2	SMARCB1	SPOP	TNFAIP3	VHL	
SF3B1	SMARCE1	SRC	TNFRSF14	WRN	

For a complete list of genes tested, visit www.CarisMolecularIntelligence.com/profilemenu.

NGS Methods

Next Generation Sequencing. Direct sequence analysis was performed on genomic DNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Illumina NextSeq platform. An Agilent custom-designed SureSelect XT assay was used to enrich 592 whole-gene targets. The genes and amino acids evaluated in this report can be found at www.carislifesciences.com. All variants reported by this assay are detected with > 99% confidence based on the frequency of the mutation present and the amplicon coverage. This test is not designed to distinguish between germ line inheritance of a variant or acquired somatic mutation. This test has a sensitivity to detect as low as approximately 10% population of cells containing a mutation with an analytic sensitivity of 96.9% to detect variants with frequency greater than 5%. This may not detect insertion/deletions events that are larger than 44 bases. The Laboratory Developed Tests (LDT) Next Generation Sequencing (NGS) assays were developed and their performance characteristics determined by Caris Life Sciences. These tests have not been cleared or approved by the US Food and Drug Administration. FDA clearance or approval is not currently necessary. All performance characteristics were determined by Caris Life Sciences. Benign and non-coding variants are not included in this report but are available upon request.
The versioned reference identifier used for the transcript ID was Feb.2009 (GRCh37/hg19).

FIG. 7L

Copy Number Alterations by Next-Generation Sequencing (NGS)

AKT2	CD274 (PD-L1)	EGFR	FGFR1	MDM2	RET(CR)
ALK	CDK4	EP300	FGFR2	MET	ROS1
ARID1A	CDK6	ERBB2 (Her2/Neu)	FGFR3	MYC	TERT
AURKB	CDK8	EZH2	GATA3	NF1	VEGFA
CCND1	CDKN2A	FGF10	KDR (VEGFR2)	NFKBIA	
CCND3	CHEK2	FGF3	MAP2K1 (MEK1)	NTRK1	
CCNE1	CRKL	FGF4	MCL1	PIK3CA	

CNA Methods

The copy number alteration (CNA) of each exon is determined by a calculation using the average sequencing depth of the sample along with the sequencing depth of each exon and comparing this calculated result to a pre-calibrated value. If all exons within the gene of interest have an average of ≥ 5 copies and the average copy number of the entire gene is ≥6 copies, the gene result is reported as amplified. If an average of ≥ 4, but < 6 copies of a gene are detected, or if the average copy number of the gene is ≥6 copies, but contains exons with an average of < 3 copies, the gene result is reported as intermediate. If an average of < 4 copies of a gene are detected, the gene result is reported as no amplification detected. A complete list of copy number alteration genes are available upon request.

FIG. 7M

Gene Fusion and Transcript Variant Detection by RNA Sequencing

Fusions Detected in Solid Tumors											
ABL1	BCR	ERG	ETV6	FGR	MET	NOTCH2	NUMBL	PIK3CA	PPARG	ROS1	TFE3
AKT3	BRAF	ESR1	EWSR1	INSR	MSMB	NRG1	NUTM1	PRKCA	RAF1	RSPO2	THRB
ALK	BRCA1	ETV1	FGFR1	MAML2	MUSK	NTRK1	PDGFRA	PPARG	RELA	RSPO3	TNFAIP3
ARHGAP26	BRCA2	ETV4	FGFR2	MAST1	MYB	NTRK2	PDGFRB	PRKCA	RET	TERT	TSHR/CBL2
AXL	EGFR	ETV5	FGFR3	MAST2	NOTCH1	NTRK3					
Variants Detected in Solid Tumors											
AR	EGFRvIII	MET									

Gene Fusion Methods

Gene fusion and variant transcript detection were performed on mRNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Agilent SureSelectXT Low Input Library prep chemistry, optimized for FFPE tissue, in conjunction with the SureSelect Human All Exon V7 bait panel (48.2 MB) and the Illumina NovaSeq. This assay is designed to detect fusions occurring at known and novel breakpoints within genes. Only a portion of genes tested are included in this report. The genes included in this report represent the subset of genes most commonly associated with cancer. All results can be provided by request. Analytical validation of this test demonstrated ≥97% Positive Percent Agreement (PPA), ≥99% Negative Percent Agreement (NPA) and ≥99% Overall Percent Agreement (OPA) with a validated comparator method.
The versioned reference identifier used for the transcript ID was Feb.2009 (GRCh37/hg19).
The complete list of unclassified alterations for RNA Whole Transcriptome Sequencing are available by request.

FIG. 7N

Protein Expression by Immunohistochemistry (IHC)

Protein Name	Staining Intensity Criteria	Percentage of Positive Cells Required	Interpretation	Comments
ERBB2 (HER2/Neu)	0	NR	Negative	Intensity ≤1+ and <10% of cells stained
MLH1	1+	≥1	Negative	Intensity ≥1+ and ≥1% of cells stained
MSH2	1+	NR	Positive	Intensity ≥1+ and ≥1% of cells stained
MSH6	1+	≥1	Negative	Intensity ≥1+ and ≥1% of cells stained
PD-L1 (SP142)	0	NR	Negative	Intensity ≥2+ and ≥1% of cells stained
PMS2	1+	≥1	Positive	Intensity ≥1+ and ≥1% of cells stained
PTEN	1+	NR	Positive	Intensity ≥1+ and ≥1% of cells stained

Clones used: ERBB2 (HER2/Neu) (4B5), MLH1 (M1), MSH2 (G219-1129, MSH6 (44), PMS2 (EPR3947), PD-L1 (SP142), PTEN (6H2.1).

T. X. Signer

Electronic Signature
TN Signer, TN MD
12/13/2019

IHC Methods

The Laboratory Developed Tests (LDT) immunohistochemistry (IHC) assays were developed and their performance characteristics determined by Caris Life Sciences. These tests have not been cleared or approved by the US Food and Drug Administration. The FDA has determined that such clearance or approval is not currently necessary. Interpretations of all immunohistochemistry (IHC) assays were performed manually by a board certified pathologist.

The following IHC assays were performed using FDA-approved companion diagnostic or FDA-cleared tests consistent with the manufacturer's instructions: ALK (VENTANA ALK (D5F3) CDx Assay, Ventana), ER (CONFIRM anti-Estrogen Receptor (ER) (SP1), Ventana), PR (CONFIRM anti-Progesterone Receptor (PR) (1E2), Ventana), HER2/neu (PATHWAY anti-HER-2/neu (4B5), Ventana), Ventana), PD-L1 22c3 (pharmDx, Dako), PD-L1 SP142 (VENTANA, Ventana in urothelial carcinomas and breast carcinoma; drug association only in urothelial and triple negative breast cancers), and PD-L1 28-8 (pharmDx, Dako).

HER2 results and interpretation follow the ASCO/CAP scoring criteria.

FIG. 70

References

	Drug	Biomarker	Reference
1	cetuximab, panitumumab	BRAF	Di Nicolantonio F., A. Bardelli, et al. (2008). "Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer." J Clin Oncol. 2008 View Citation Online
2	cetuximab, panitumumab	BRAF	National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Colon Cancer Version 1.2017. 2017. National Comprehensive Cancer Network. View Citation Online
3	cetuximab, panitumumab	BRAF, NRAS	De Roock, W., S. Tejpar, et al. (2010). "Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis." Lancet Oncol. 11: 753-62. View Citation Online
4	cetuximab, panitumumab	BRAF	Sepulveda, A. R., J.A. Hamilton, et al. (2017). "Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of Clinical Oncology." J Clin Oncol. J Clin Oncol. 2017 Feb 6:JCO2016716898. View Citation Online
5	cetuximab, panitumumab	BRAF	Santini-Bianchi, A., S. Sena, et al. (2009). "Multi-centre genomic analysis of molecular alterations for predicting clinical benefit to EGFR-targeted monoclonal antibodies in colorectal cancer." PLoS ONE. 4(10): e7287. doi:10.1371/journal.pone.0007287. View Citation Online
6	cetuximab, panitumumab	BRAF	Pietrantonio, F., S. Sena, et al. (2015). "Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: a meta-analysis." Eur J Cancer. 2015 51(5):587-94 View Citation Online
7	cetuximab, panitumumab	BRAF	Van Cutsem, E., F. Ciardiello, et al. (2011). "Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status." J Clin Oncol. 29(15):2011-9. View Citation Online
8	cetuximab, panitumumab	KRAS	De Roock, W., Tejpar, S. (2010). Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer. Jama.304(16):1812-20. View Citation Online
9	cetuximab, panitumumab	KRAS, NRAS	Goodlad, J.R., S.D. Patterson, et al. (2012). "Panitumumab-FOLFOX6 Treatment and RAS Mutations in Colorectal Cancer" N Engl J Med. 366(1): 1027-1036 View Citation Online
10	cetuximab, panitumumab	KRAS	Lievre, A., Laurent-Puig, P., (2008). "KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab". J Clin Oncol. 26(2):374-9. View Citation Online
11	cetuximab, panitumumab	KRAS	Chen, J., G. Shi, et al. (2012). "Association between KRAS exon 13 mutations and clinical response to anti-EGFR treatment in patients with metastatic colorectal cancer: results from a meta-analysis"Cancer Chemother Pharmacol. 2012 Oct 23. [Epub ahead of print] View Citation Online
12	cetuximab, panitumumab	KRAS	Amado, R.G., et al. (2008). "Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer." J. Clin. Oncol. 26(1):626-1634. View Citation Online
13	cetuximab, panitumumab	KRAS	Douillard, J.Y., J. Cunsent (2010). "Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study." J. Clin. Oncol. 28(4):87-4756. View Citation Online
14	cetuximab, panitumumab	KRAS	Peeters, M., J. Cunsent, et al. (2010). "Mutant KRAS Codon 12 and 13 Alleles in Patients With Metastatic Colorectal Cancer: Assessment As Prognostic and Predictive Biomarkers of Response to Panitumumab" J. Clin. Oncol. 31:759-765. View Citation Online
15	cetuximab, panitumumab	KRAS	National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Colon Cancer Version 1.2015. View Citation Online
16	cetuximab, panitumumab	KRAS	Tejpar, S., E. Van Cutsem, et al. (2012). "Association of KRAS G13D Tumor Mutations With Outcome in Patients With Metastatic Colorectal Cancer Treated With First-Line Chemotherapy With or Without Cetuximab." J Clin Oncol. 30(28):3570-7. View Citation Online

FIG. 7P

References

#	Drug	Biomarker	Reference
17	cetuximab, panitumumab	NRAS	Peeters, M., S.D. Patterson, et al. (2013). "Massively Parallel Tumor Multigene Sequencing to Evaluate Response to Panitumumab in a Randomized Phase III Study of Metastatic Colorectal Cancer." Clin Cancer Res 19(7): 1902-1912. View Citation Online
18	lapatinib, pertuzumab, trastuzumab	ERBB2 (HER2/Neu)	Valtorta, E., M. Gambacorta, et al. (2015). "Assessment of a HER2 scoring system for colorectal cancer: results from a validation study". Modern Pathology 28: 1481-1491. View Citation Online
19	lapatinib, pertuzumab, trastuzumab	ERBB2 (HER2/Neu)	Sartore-Bianchi, A., S. Siena, (2018) "Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial." Lancet Oncol 19(6):738-746 View Citation Online
20	lapatinib, pertuzumab, trastuzumab	ERBB2 (HER2/Neu)	Hainsworth, J.D., A. Kurtzman, et al. (2018) "Targeted Therapy for Advanced Solid Tumors on the Basis of Molecular Profiles: Results From MyPathway, an Open-Label, Phase IIa Multiple Basket Study." J Clin Oncol. 36(6):536-542. View Citation Online

FIG. 7Q

Patient

Name:
Date of Birth:
Sex:
Case Number: TN20-
Diagnosis: Carcinoma, NOS

Specimen Information

Primary Tumor Site: Ovary
Specimen Site: Ascending colon
Specimen ID:
Specimen Collected:
All Testing Completed:

Ordered By

Results with Therapy Associations

Biomarker	Test Type	Result Type	Result	Therapy Association		Tier/Level of Evidence*
PD-L1 (22c3)	IHC	Protein	Positive, CPS 30	Approved	pembrolizumab	Level 3A
ERBB2 (Her2/Neu)	IHC	Protein	Negative (2+, 1%)	Approved, Off-Label	pertuzumab, trastuzumab	Level 3B

* Biomarker reporting classification: Level 1 - highest level of clinical evidence and/or biomarker association included on the drug label; Level 2 - strong evidence of clinical significance and is endorsed by standard clinical guidelines; Level 3 - potential clinical significance (3A - evidence exists in present tumor type; 3B - evidence exists in another tumor type)

Cancer Type Relevant Biomarkers

Biomarker	Test Type	Result Type	Result
MSI	Seq	DNA-Tumor	Stable
Mismatch Repair Status	IHC	Protein	Proficient
NTRK1/2/3	Seq	RNA-Tumor	Fusion Not Detected
Tumor Mutational Burden	Seq	DNA-Tumor	Low (5 Mut/Mb)
ATM	Seq	DNA-Tumor	Mutation Not Detected
BRCA1	Seq	DNA-Tumor	Mutation Not Detected
BRCA2	Seq	DNA-Tumor	Mutation Not Detected

Biomarker	Test Type	Result Type	Result
ER	IHC	Protein	Negative (0)
PR	IHC	Protein	Negative (1+, 2%)
Alteration Detected			
TP53	Seq	DNA-Tumor	Pathogenic Variant Exon 7 c.[742C>T];[=]

SAMPLE REPORT - FOR ILLUSTRATIVE PURPOSES ONLY. NOT FOR CLINICAL USE.

The selection of any, all, or none of the matched therapies and/or safety with the discretion of the treating physician. Decisions on patient care and treatment must be based on the independent medical judgment of the treating physician, taking into consideration all available information concerning the patient's condition, the FDA prescribing information for any therapeutic, and in accordance with the applicable standard of care. Whether or not a particular patient will benefit from a selected therapy is based on many factors and can vary significantly. All trademarks and registered trademarks are the property of their respective owners.

FIG. 8A

Genomic Signatures

Test	Result	Method	Status
Microsatellite Instability (MSI)	Seq	DNA-Tumor	Stable
Tumor Mutational Burden (TMB)	Seq	DNA-Tumor	Result: Low 5 Low 10 High

Genes Tested with Pathogenic or Likely Pathogenic Alterations

Gene	Test Type	Method	Result/Interpretation	Result/Interpretation	Exon	HGVS Nomenclature	Allele Frequency (%)
TP53	Seq	DNA-Tumor	Pathogenic Variant	c.782+1G>T	7	c.782+1G>T	26

Unclassified alterations for DNA and RNA sequencing can be found in the full report.
Formal nucleotide nomenclature and gene reference sequences can be found in the Appendix of this report.

Immunohistochemistry Results

Test	Result	Test	Result
ER	Negative (0)	PD-L1 (22c3)	Positive, CPS: 50
ERBB2 (Her2/Neu)	Negative (2+), 1%	PMS2	Positive (2+), 80%
MLH1	Positive (3+), 80%	PR	Negative (1+), 2%
MSH2	Positive (3+), 90%	PTEN	Positive (3+), 100%
MSH6	Positive (2+), 80%		

Genes Tested with Indeterminate Results by Tumor DNA Sequencing

ATRX	CHD4	FBXW7	FLT3	KMT2C	MRE11	NF1	RAD50	ROS1	RUNX1	SMARCB1	TERT[1]
CDC73	FANCL										

[1] TERT Promoter region not sequenced.
* Genes in this table were noted Indeterminate due to low coverage for some or all exons.

SAMPLE REPORT FOR ILLUSTRATIVE PURPOSES ONLY. NOT FOR CLINICAL USE.

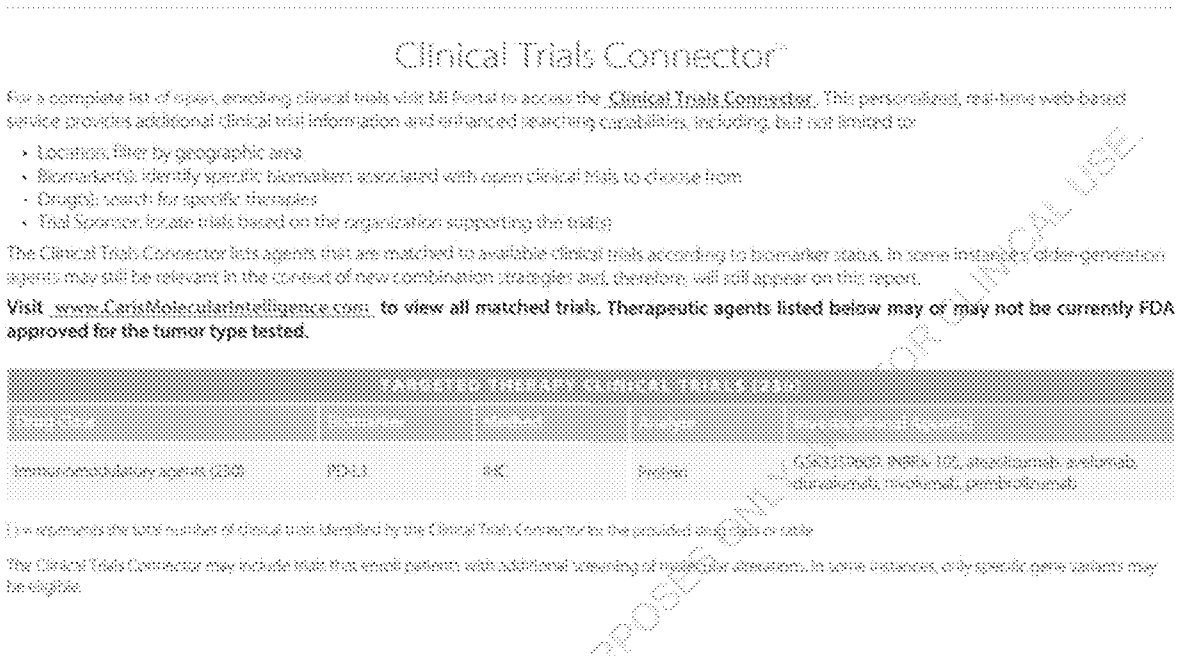
FIG. 8B


MI GPSai
The MI GPSai® (MI Genomic Prevalence Score - Artificial Intelligence) is a cancer-type similarity assessment which compares the characteristics of a patient's tumor against other tumors in the Caris database. MI GPSai analyzes a tumor's molecular signature and provides the prevalence of that signature in the Caris Life Sciences genomic and transcriptomic database across 21 distinct cancer categories.
Ovarian, Fallopian Tube Adenocarcinoma 86 %
Uterine Endometrial Adenocarcinoma 3 %
Cervical Adenocarcinoma <1 %
Renal Adenocarcinoma 0 %
Central Nervous System Cancer 0 %
Cholangiocarcinoma 0 %
Colon Adenocarcinoma 0 %
Gastroesophageal Adenocarcinoma 0 %
GIST 0 %
Hepatocellular carcinoma 0 %
Lung Adenocarcinoma 0 %
Melanoma 0 %
Meningioma 0 %
Ovarian Granulosa Cell Tumor 0 %
Pancreas Adenocarcinoma 0 %
Prostate Adenocarcinoma 0 %
Renal Cell Carcinoma 0 %
Squamous Cell Carcinoma 0 %
Thyroid Cancer 0 %
Urothelial Carcinoma 0 %
Uterine Sarcoma 0 %
Methods
MI GPSai® is a machine learning platform that was trained on genomic data from 36,521 cases and transcriptomic data from over 11,000 cases. In a validation set of over 12,000 additional cases, MI GPSai accurately predicted the tumor category in the labeled data set with an accuracy of over 93%. The accuracy increased to 97% when the second highest ranking predicted cancer type was included. The profile has been validated to diffuse data among 21 distinct cancer types. MI GPSai prevalence results were produced at or above the required confidence level for 98% of samples in the validation set. Samples that do not generate a score at or above this confidence level will not receive a MI GPSai result.
SAMPLE REPORT, FOR ILLUSTRATIVE PURPOSES ONLY. NOT FOR CLINICAL USE.

FIG. 8C


Notes of Significance
GPS pointed towards a tumor of ovarian origin. Immunostains show positive staining with WT1 and PAX8 and no staining with CK20 and CDX2. These results are consistent with a high grade tumor of ovarian origin over colonic origin.
Clinical Trials Connector™ opportunities based on biomarker expression: 230 Targeted Therapy Trials. See page 5 for details.
Specimen Information
Specimen ID:
Specimen Collected:
Specimen Received:
Other Testing Initiated:
Gross Description:
Pathologic Diagnosis:
Dissection Information: Molecular testing of this specimen was performed after harvesting of targeted tissues with an approved manual microdissection technique. Candidate slides were examined under a microscope and areas containing tumor cells (and separately normal cells, when necessary for testing) were circled. A laboratory technician harvested targeted tissues for extraction from the marked areas using a dissection microscope.
SAMPLE REPORT - FOR ILLUSTRATIVE PURPOSES ONLY - NOT FOR CLINICAL USE

FIG. 8D



Clinical Trials Connector®

For a complete list of open, enrolling clinical trials visit MI Portal to access the Clinical Trials Connector. This personalized, real-time web-based service provides additional clinical trial information and enhanced searching capabilities, including, but not limited to:

- Location: filter by geographic area
- Biomarker(s): identify specific biomarkers associated with open clinical trials to choose from
- Drug(s): search for specific therapies
- Trial Sponsor: locate trials based on the organization supporting the study

The Clinical Trials Connector lists agents that are matched to available clinical trials according to biomarker status. In some instances, older-generation agents may still be relevant in the context of new combination strategies and, therefore, will still appear on this report.

Visit www.CarisMolecularIntelligence.com to view all matched trials. Therapeutic agents listed below may or may not be currently FDA approved for the tumor type tested.

Therapy	Target	Match	Trials	Agents
Immunomodulatory agents (2/8)	PD-L1	4/4	Nivolumab	CA9-1/PD1, INBRX-105, avelumab, avelumab, durvalumab, nivolumab, pembrolizumab

( ) = represents the total number of clinical trials identified by the Clinical Trials Connector for the provided drug class or table

The Clinical Trials Connector may include trials that enroll patients with additional screening of molecular elements. In some instances, only specific gene variants may be eligible.

PROSECUTION COPY ONLY - NOT FOR COMMERCIAL USE

FIG. 8E

Disclaimer

Decisions on patient care and treatment must be based on the independent medical judgment of the treating physician, taking into consideration all available information concerning the patient's condition, the FDA prescribing information for any therapeutic, and in accordance with the applicable standard of care. Drug associations provided in the report do not guarantee that any particular agent will be effective for the treatment of any patient or for any particular condition. Caris Life Sciences expressly disclaims and makes no representation or warranty whatsoever relating, directly or indirectly, to the performance of services, including any information provided and/or conclusions drawn from therapies that are included or omitted from this report. Whether or not a particular patient will benefit from a selected therapy is based on many factors and can vary significantly. The selection of therapy, if any, resides solely in the discretion of the treating physician.

Individual assays that are available through Caris Molecular Intelligence® include both Laboratory Developed Tests (LDT) and U.S. Food and Drug Administration (FDA) approved or cleared tests. Caris MPI, Inc. d/b/a Caris Life Sciences® is certified under the Clinical Laboratory Improvement Amendments (CLIA) as qualified to perform high complexity clinical laboratory testing, including all of the assays that comprise the Caris Molecular Intelligence® The LDTs were developed and their performance characteristics determined by Caris. The LDTs have not been cleared or approved by the U.S. Food and Drug Administration. Caris' CLIA certification number is located at the bottom of each page of this report. Certain tests have not been cleared or approved by the FDA. The FDA has determined that clearance or approval is not necessary for certain laboratory developed tests. Caris LDTs are used for clinical purposes. They are not investigational or for research.

The information presented in the Clinical Trials Connector™ section of this report, if applicable, is compiled from sources believed to be reliable and current. However, the accuracy and completeness of the information provided herein cannot be guaranteed. The clinical trials information present in the biomarker description was compiled from www.clinicaltrials.gov. The contents are to be used only as a guide, and health care providers should employ their best comprehensive judgment in interpreting this information for a particular patient. Specific eligibility criteria for each clinical trial should be reviewed as additional inclusion criteria may apply.

All materials, documents, data, data software, information and/or inventions supplied to customers by or on behalf of Caris or created by either party relating to the services shall be and remain the sole and exclusive property of Caris. Customer shall not use or disclose the information provided by Caris through the services or related reports except in connection with the treatment of the patient for whom the services were ordered and shall not use such property for, or disseminate such property to, any third parties without expressed written consent from Caris. Customer shall deliver all such property to Caris immediately upon demand or upon Caris ceasing to provide the services.

Caris Molecular Intelligence is subject to Caris' intellectual property. Patent: www.carislifesciences.com/ip.

For Illustrative Purposes Only - Not Approved for Clinical Use

FIG. 8F

Mutational Analysis by Next-Generation Sequencing (NGS)

GENOMIC ALTERATIONS ANALYZED	
Genomic Alterations Analyzed	Result
5	Low

TMB Methods

Tumor Mutational Burden (TMB) analysis was performed based on Next Generation Sequencing analysis from genomic DNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Illumina platform. TMB is calculated using nonsynonymous, in-frame indel, and frameshift indel mutations that have not been previously reported as germline alterations in the Genome Aggregation Database (gnomAD) and dbSNP151 or as common benign variants identified by Caris geneticists. The cutoff point (≥10) is based on the KEYNOTE-158 pembrolizumab trial (Marabelle et al., ESMO 2019), which showed that patients with a TMB of ≥ 10 mutations per megabase across several tumor types had higher response rates than patients with fewer than 10 mutations per megabase. Caris Life Sciences is a participant in the Friends of Cancer Research TMB Harmonization Project (Merino et al., 2020).

ORGANISM IDENTIFICATION ANALYSIS		
Test	Result	Flag
MSI	No microsatellite instability detected.	Stable

Procedure: NGS

Microsatellite Instability Analysis

Microsatellite instability status by NGS (MSI-NGS) is measured by the direct analysis of known microsatellite regions sequenced in the CMI NGS panel. To establish clinical thresholds, MSI-NGS results were compared with results from over 2,000 matching clinical cases analyzed with traditional, PCR-based methods. Genomic variants in the microsatellite loci are detected using the same depth and frequency criteria as used for mutation detection. Only insertions and deletions resulting in a change in the number of tandem repeats are considered in this assay. Some microsatellite regions with known polymorphisms or technical sequencing issues are excluded from the analysis. The total number of microsatellite alterations in each sample are counted and grouped into three categories: High, Equivocal and Stable. MSI-Low results are reported in the Stable category. Equivocal results have a total number of microsatellite alterations in between High and Stable.

SOMATIC VARIANTS DETECTED							
Gene	Test	Result	Alteration	Exon	cDNA	Coverage	Transcript
TP53	DNA-Tumor	Pathogenic Variant	Intron Splice Variant	7	c.782+1G>T	26	NM_000546.5

Interpretation: A pathogenic mutation that disrupts an intron splice site was detected in TP53.

TP53, or p53, plays a critical role in modulating response to cellular stress through transcriptional regulation of genes involved in cell-cycle arrest, DNA repair, apoptosis, and senescence. Inactivation of the p53 pathway is essential for the formation of the majority of human tumors. Mutation in p53 (TP53) remains one of the most commonly described genetic events in human neoplasia, estimated to occur in 30-50% of all cancers. Generally, presence of a disruptive p53 mutation is associated with a poor prognosis in all types of cancers, and diminished sensitivity to radiation and chemotherapy. Germline p53 mutations are associated with the Li-Fraumeni syndrome (LFS) which may lead to early-onset of several forms of cancer currently known to occur in the syndrome, including sarcomas of the bone and soft tissues, carcinomas of the breast and adrenal cortex (hereditary adrenocortical carcinoma), brain tumors and acute leukemias.

SAMPLE REPORT - NOT FOR CLINICAL USE - FOR RESEARCH PURPOSES ONLY

FIG. 8G

Mutational Analysis by Next-Generation Sequencing (NGS)

Genes Tested with Sufficient Coverage for Variant Analysis*				
ATRX	FANCE	KMT2C	RAD50	SMARCE1
CDC73	FBXW7	MRE11	ROS1	TERT ¶
CHD4	FLT3	NF1	RUNX1	

* Genes in this table were ruled indeterminate due to low coverage for some or all exons.

¶ TERT Promoter region not sequenced

For a complete list of genes tested, visit www.CarisMolecularIntelligence.com/profilemenu.

NGS Methods

Next Generation Sequencing: Direct sequence analysis was performed on genomic DNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Illumina NextSeq platform. An Agilent custom-designed SureSelect XT assay was used to enrich 592 whole gene targets. The ge and amino acids evaluated in this report can be found at www.carisillea.lmce.com. All variants reported by this assay are detected with > 99% confidence based on the frequency of the mutation present and the amplicon coverage. This test is not designed to distinguish between germ inheritance of a variant or acquired somatic mutation. This test has a sensitivity to detect as low as approximately 10% population of cells contain a mutation with an analytic sensitivity of 96.9% to detect variants with frequency greater than 5%. This may not detect insertion/deletions events are longer than 44 bases. The Laboratory Developed Tests (LDT) Next Generation Sequencing (NGS) assays were developed and their performance characteristics determined by Caris Life Sciences. These tests have not been cleared or approved by the US Food and Drug Administration. FDA clearance or approval is not currently necessary. All performance characteristics were determined by Caris Life Sciences. Benign and non-coding variants are not included in this report but are available upon request.
The versioned reference identifier used for the transcript ID was Feb 2009 (GRCh37/hg19).

FOR CLINICAL USE NOT FOR CLINICAL USE

FIG. 8H

Copy Number Alterations by Next-Generation Sequencing (NGS)

Genes Interrogated for Copy Number Alterations					
AFF4	CHEK2	ETV1	KMT2C	NF1	STAG2
AKAP9	CIITA	FGFR1	KRAS	PDGFRA	STAT3
CASP8	CNBD1	GOPC	LRP1B	RAD50	SUFU
CDC73	EBF1	GPHN	MAP3K1	RECQL4	TP53
CHIC2	ERBB4	JAK2	MLLT10	ROS1	TSHR
CHN1	EPS15	KIF5B	MNX1	SMARCE1	

CNA Methods

The copy number alteration (CNA) of each exon is determined by a calculation using the average sequencing depth of the sample along with the sequencing depth of each exon and comparing this calculated result to a pre-calibrated value. If all exons within the gene of interest have an average of ≥3 copies and the average copy number of the entire gene is ≥6 copies, the gene result is reported as amplified. If an average of ≥ 4, but < 6 copies of a gene are detected, or if the average copy number of the gene is ≥6 copies, but contains exons with an average of < 3 copies, the gene result is reported as intermediate. If an average of < 4 copies of a gene are detected, the gene result is reported as no amplification detected. A complete list of copy number alteration genes are available upon request.

SAMPLE NOT FOR CLINICAL USE

FIG. 8I

Gene Fusion and Transcript Variant Detection by RNA Sequencing

Gene Fusion Methods

Gene fusion and variant transcript detection were performed on mRNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Agilent SureSelectXT Low Input Library prep chemistry, optimized for FFPE tissue, in conjunction with the SureSelect Human All Exon V7 bait panel (48.2 Mb) and the Illumina NovaSeq. This assay is designed to detect fusions occurring at known and novel breakpoints within genes. Only a portion of genes tested are included in this report. The genes included in this report represent the subset of genes most commonly associated with cancer. All results can be provided by request. Analytical validation of this test demonstrated ≥97% Positive Percent Agreement (PPA), ≥99% Negative Percent Agreement (NPA) and ≥99% Overall Percent Agreement (OPA) with a validated comparator method.
The versioned reference identifier used for the transcript ID was Feb.2009 (GRCh37/hg19).
The complete list of unclassified alterations for RNA Whole Transcriptome Sequencing are available by request.

CLINICAL

FIG. 8J

Protein Expression by Immunohistochemistry (IHC)

Biomarker	Expression Status			Interpretation
	Assay Result (0, 1+, 2+, 3+)	Percentage (%)	Status	Intensity and Percentage Criteria
ER	0	100	Negative	Intensity of ≥3+ with ≥50% or ≥2+ with ≥75% of cells stained
ERBB2 (HER2/neu)	2+	1	Negative	Intensity of ≥3+ with ≥10% or ≥2+ with ≥10% or ≥1+ with ≥10% of cells stained
MLH1	3+	80	Positive	Intensity ≥1+ and ≥1% of cells stained
MSH2	3+	80	Positive	Intensity ≥1+ and ≥1% of cells stained
MSH6	2+	90	Positive	Intensity ≥1+ and ≥1% of cells stained
PMS2	2+	80	Positive	Intensity ≥1+ and ≥1% of cells stained
PR	1+	2	Negative	Intensity ≥1+ and ≥10% of cells stained
PTEN	1+	100	Positive	Intensity ≥1+ and ≥1% of cells stained

PD-L1 22c3 pharmDx Assay		
Result	CPS	Tumor and Tumor-Associated Immune Cells
Positive	30	≥1 (%)

CPS: Combined Positive Score, calculated as the number of PD-L1 staining cells (tumor cells, lymphocytes, macrophages) divided by the total viable tumor cells, multiplied by 100.

Clones used: ER (SP1), PR (1E2), ERBB2 (4B5), MLH1 (M1), MSH2 (G219-1129), MSH6 (SP93), PMS2 (A16-4), PD-L1 (22c3), PTEN (6H2.1).

Electronic Signature

IHC Methods

The Laboratory Developed Tests (LDT) immunohistochemistry (IHC) assays were developed and their performance characteristics determined by Caris Life Sciences. These tests have not been cleared or approved by the US Food and Drug Administration. The FDA has determined that such clearance or approval is not currently necessary. Interpretations of all immunohistochemistry (IHC) assays were performed manually by a board certified pathologist using a microscope and/or digital whole slide image(s).

The following IHC assays were performed using FDA-approved companion diagnostic or FDA-cleared tests consistent with the manufacturer's instructions: ALK (VENTANA ALK (D5F3) CDx Assay, Ventana), ER (CONFIRM anti-Estrogen Receptor (ER) (SP1), Ventana), PR (CONFIRM anti-Progesterone Receptor (PR) (1E2), Ventana), HER2/neu (PATHWAY anti-HER-2/neu (4B5), Ventana), Ventana), PD-L1 22c3 (pharmDx, Dako), PD-L1 SP142 (VENTANA, Ventana in urothelial carcinomas, breast carcinoma and non-small cell lung cancer, drug association only in urothelial, triple negative breast cancers and non-small cell lung cancer), and PD-L1 28-8 (pharmDx, Dako).

HER2 results and interpretation follow the ASCO/CAP scoring criteria.

SAMPLE REPORT FOR ILLUSTRATIVE PURPOSES ONLY - NOT FOR CLINICAL USE

FIG. 8K

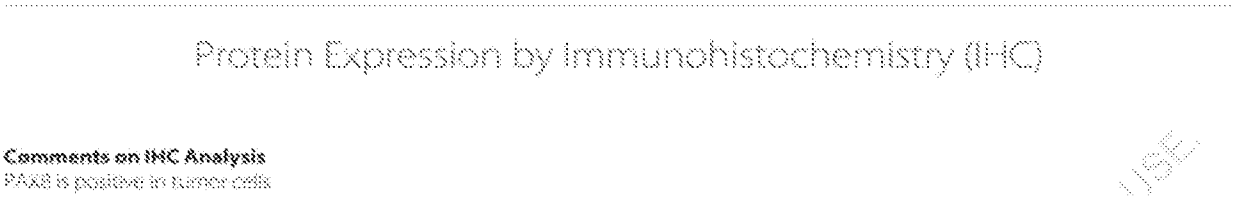

Protein Expression by Immunohistochemistry (IHC)
Comments on IHC Analysis
PAX8 is positive in tumor cells.

FIG. 8L

References

#	Drug	Target	Reference
1	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Bang, Y.-J., Xu, R.-H., Chin, K., et al. (2017). "Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial." Lancet. 376(9742): 687-697. View Citation Online
2	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Baselga, J., Im, S.A., Iwata, H., et al. (2017). "Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer." N. Engl. J. Med. 377(2): 186-187. View Citation Online
3	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Ye, M., J. Liu, et al. (2011). "Pertuzumab in adjuvant treatment (NEO) positive early breast cancer patients: A meta-analysis of published randomized controlled trials." PLoS ONE 6(9): e25161. doi:10.1371/journal.pone.0025161. View Citation Online
4	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Cortes, J., J. Baselga, et al. (2013). "Pertuzumab monotherapy after trastuzumab-based treatment and subsequent reintroduction of trastuzumab: activity and tolerability in patients with advanced human epidermal growth factor receptor 2-positive breast cancer." J Clin Oncol. 31: 2965-71. (2013). 31: 4297. View Citation Online
5	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Hainsworth, J.D., A. Nagrial, et al. (2018). "Targeted Therapy for Advanced Solid Tumors on the Basis of Molecular Profiles: Results From MyPathway, an Open-Label, Phase IIa Multiple Basket Study." J Clin Oncol. 36(6): 536-542. View Citation Online
6	pertuzumab, trastuzumab	ERBB2 (Her2/Neu)	Gianni, L., M. Pienkowski, et al. (2017). "Adjuvant pertuzumab in HER2-positive breast cancer." N. Engl. J. Med. 377(2): 113. View Citation Online
7	pembrolizumab	PD-L1 (CD274)	Mehnert, J.M., J.A. Longo-Imedio, et al. (2018). "Antitumor activity and safety of pembrolizumab in patients with advanced malignant ovarian cancer: results from the phase IB KEYNOTE-100 study." Ann Oncol. 30(7): 1080-1087. View Citation Online

FIG. 8M

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/018263

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G16B 20/00; G06N 20/00; G16H 50/20; G16B 40/00; G06N 3/08; G16H 50/70 (2021.01)
CPC - G16B 20/00; G06N 20/00; G16H 50/20; G16B 40/00; G06N 3/08; G16H 50/70 (2021.05)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
see Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2018/0068083 A1 (20/20 GENESYSTEMS INC.) 08 March 2018 (08.03.2018) entire document	1, 3, 5-7, 10-13
A	US 2018/0239949 A1 (CELLANYX DIAGNOSTICS, LLC) 23 August 2018 (23.08.2018) entire document	1, 3, 5-7, 10-13
A	US 2015/0152474 A1 (CARIS LIFE SCIENCES SWITZERLAND HOLDINGS GMBH) 04 June 2015 (04.06.2015) entire document	1, 3, 5-7, 10-13
A	US 2018/0045727 A1 (CARIS MPI, INC.) 15 February 2018 (15.02.2018) entire document	1, 3, 5-7, 10-13
A	LIBBRECHT et al., " Machine learning in genetics and genomics " , HHS Public Access, June 2015, retrieved on [02.06.2021]. Retrieved from the internet <URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5204302/pdf/nihms-839467.pdf> entire document	1, 3, 5-7, 10-13
A	US 2018/0039731 A1 (NANTOMICS, LLC) 08 February 2018 (08.02.2018) entire document	1, 3, 5-7, 10-13

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 June 2021	JUL 01 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Harry Kim Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/018263

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

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

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

2. ☒ Claims Nos.: 2, 4, 8, 9
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claims 2, 4, 8, 9 are not drafted in accordance with Rule 6.2(a) and are omnibus claims.

3. ☒ Claims Nos.: 17-29, 40-90, 95-102
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet(s).

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

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

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

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 3, 5-7, 10-13

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/018263

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-13, is drawn to a data processing apparatus for generating input data structure for use in training a machine learning model to predict at least one attribute of a biological sample.

Group II, claims 14-16, is drawn to a method for determining at least one attribute of a biological sample.

Group III, claims 30-36, is drawn to a method for classifying a biological sample.

Group IV, claims 37-39, 91-94, 103, is drawn to a method comprising: obtaining a biological sample from a subject having a cancer.

Group V, claims 104-105, is drawn to a system for training an analysis model for identifying at least one attribute of a cancer sample obtained from a body

The inventions listed as Groups I, II, III, IV, or V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group I invention: the first data structure that structures the data representing the set of one or more biomarkers associated with the sample with the second data structure representing predicted at least one attribute data for the sample as claimed therein is not present in the invention of Groups II, III, IV, or V. The special technical feature of the Group II invention: the provided input data that represents a probability or likelihood that the sample represented by the provided input data corresponds to the at least one attribute; providing, to a voting unit, the output data obtained for each of the plurality of machine learning models, wherein the provided output data includes data representing initial sample attributes determined by each of the plurality of machine learning models as claimed therein is not present in the invention of Groups I, III, IV, or V. The special technical feature of the Group III invention: obtaining, by one or more computers, first data representing one or more initial classifications for the biological sample that were previously determined based on RNA sequences of the biological sample; obtaining, by one or more computers, second data representing another initial classification for the biological sample that were previously determined based on DNA sequences of the biological sample; providing, by one or more computers as claimed therein is not present in the invention of Groups I, II, IV, or V. The special technical feature of the Group IV invention: obtaining a biological sample from a subject having a cancer; performing at least one assay on the sample to assess one or more biomarkers, thereby obtaining a biosignature for the sample; providing the biosignature into a model as claimed therein is not present in the invention of Groups I, II, III, or V. The special technical feature of the Group V invention: each model signature is configured to differentiate between at least one attribute within each of the at least one attribute; obtaining, by the system, a set of training data items, wherein each training data item represents DNA or RNA sequencing results and includes data indicating (i) whether or not a variant was detected in the sequencing results and (ii) a number of copies of a gene or transcript in the sequencing results; and training, by the system, an analysis model using the obtained set of training data items as claimed therein is not present in the invention of Groups I, II, III, or IV.

Groups I, II, III, IV, and V lack unity of invention because even though the inventions of these groups require the technical feature of training an analysis model/machine learning model for identifying at least one attribute of a cancer sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primal tumor origin, cancer/disease type, organ group, histology, and any combination thereof, this technical feature is not a special technical feature as it does not make a contribution over the prior art.

Specifically, US 2018/0039731 to Nantomics, LLC teaches training an analysis model/machine learning model for identifying at least one attribute of a cancer sample obtained from a body, wherein the at least one attribute is selected from the group consisting of a primal tumor origin, cancer/disease type, organ group, histology, and any combination thereof (Paras. [0018], [0039], [0041], [0050]).

Since none of the special technical features of the Group I, II, III, IV or V inventions are found in more than one of the inventions, unity of invention is lacking.