



(51) International Patent Classification:

C12Q 1/6876 (2018.01) C12Q 1/6883 (2018.01)

C12Q 1/6881 (2018.01) C12Q 1/6886 (2018.01)

(21) International Application Number:

PCT/US2019/017799

(22) International Filing Date:

13 February 2019 (13.02.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/629,975 13 February 2018 (13.02.2018) US

(71) Applicants: GENECENTRIC THERAPEUTICS, INC.

[US/US]; 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL** [US/US]; 109 Church Street, Chapel Hill, North Carolina 27516 (US).

(72) Inventors: FARUKI, Hawazin; c/o GeneCentric Therapeutics, Inc., 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **MAYHEW, Greg**; c/o GeneCentric Therapeutics, Inc., 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **LAI-GOLDMAN, Myla**; c/o GeneCentric Therapeutics, Inc., 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **PEROU, Charles**; c/o The University of North Carolina at Chapel Hill, 109 Church Street, Chapel Hill, North Carolina 27516 (US).

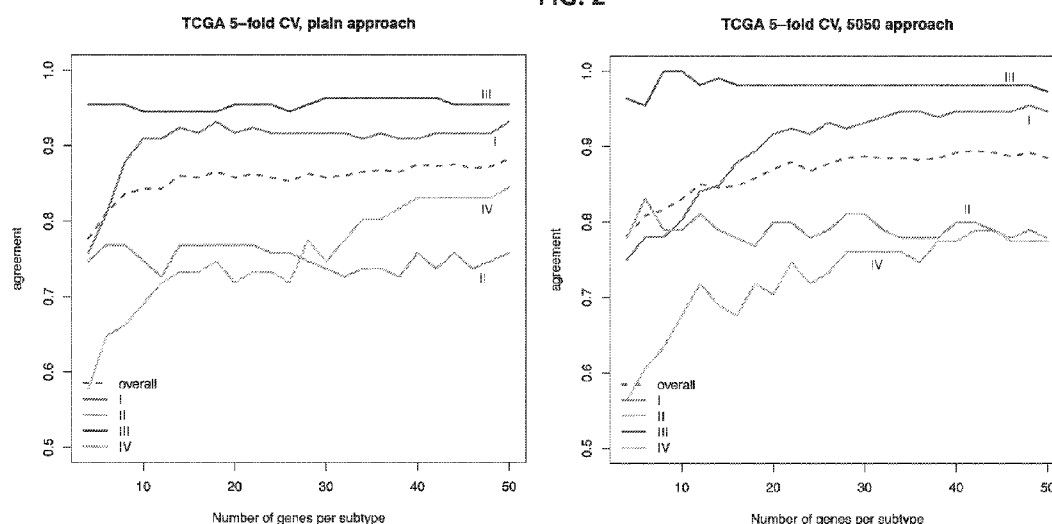
mons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **MAYHEW, Greg**; c/o GeneCentric Therapeutics, Inc., 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **LAI-GOLDMAN, Myla**; c/o GeneCentric Therapeutics, Inc., 79 TW Alexander Drive, 4401 Research Commons, Suite 105, Research Triangle Park, North Carolina 27709 (US). **PEROU, Charles**; c/o The University of North Carolina at Chapel Hill, 109 Church Street, Chapel Hill, North Carolina 27516 (US).

(74) Agent: TUSCAN, Michael et al.; Cooley LLP, 1299 Pennsylvania Avenue, Suite 700, Washington, District of Columbia 20004 (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, 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,

(54) Title: METHODS FOR SUBTYPING OF BLADDER CANCER

FIG. 2



(57) **Abstract:** Methods and compositions are provided for determining a subtype of Muscle Invasive Bladder Cancer (MIBC) in an individual by detecting the expression level of at least one classifier biomarker selected from a group of gene signatures for MIBC. Also provided herein are methods and compositions for determining the response of an individual with a MIBC subtype to a therapy such as immunotherapy.



SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *of inventorship (Rule 4.17(iv))*

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))*
- *with sequence listing part of description (Rule 5.2(a))*

**IN THE UNITED STATES PATENT & TRADEMARK
RECEIVING OFFICE**

INTERNATIONAL PCT PATENT APPLICATION

METHODS FOR SUBTYPING OF BLADDER CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Application Serial No. 62/629,975, filed February 13, 2018, which is herein incorporated by reference in its entirety for all purposes.

FIELD

[0002] The present invention relates to methods for determining a muscle invasive bladder cancer (MIBC) subtype and for predicting the prognosis of a patient inflicted with specific subtypes of bladder cancer.

STATEMENT REGARDING SEQUENCE LISTING

[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is GNCN_014_01WO_SeqList_ST25.txt. The text file is $\approx$ 401 KB, was created on February 13, 2019, and is being submitted electronically via EFS-Web.

BACKGROUND

[0004] Bladder cancer is comprised of cancers arising from the tissues of the urinary bladder. The most common type of bladder cancer is transitional cell carcinoma, which begins in urothelial cells that line the inside of the bladder. Urothelial cells are transitional cells, which are able to change shape and stretch when the bladder is full. This type of cancer is also called urothelial carcinoma. Other types of bladder cancer include squamous cell carcinoma (cancer that begins in thin, flat cells lining the bladder) and adenocarcinoma (cancer that begins in cells that make and release mucus and other fluids). More than 90% of bladder cancers are urothelial cell carcinoma and about 5% are squamous cell carcinoma. The

most significant known risk factors for bladder cancer include tobacco use, age, diet, consumption of alcohol, chronic inflammation, obesity, and the exposure to cancer-causing substances.

[0005] Common detection and diagnosis methods of bladder cancer include cystoscopy, non-invasive urine bound markers, and molecular bladder cancer detection assays. However, a challenge in clinical practice is that the pathologic assessment can be uncertain, and therefore, lacks accuracy.

[0006] The present disclosure addresses the limitations of the current methods and other needs in the field for an efficient method for improved bladder cancer tumor classification that may inform prognosis and patient management based on underlying genomic and biologic tumor characteristics.

SUMMARY

[0007] The methods disclosed herein include determination of a bladder cancer (e.g., MIBC) subtype, treatment of muscle invasive bladder cancer, prediction of overall survival of bladder cancer (e.g., MIBC) patients, and application of an algorithm for categorization of bladder cancer (e.g., MIBC) tumors into one of 4 subtypes (Type I: luminal, Type II: luminal infiltrated, Type III: basal, and Type IV: basal infiltrated/neuronal)). The algorithm can be a classification to the nearest centroid (CLaNC) algorithm.

[0008] In one aspect, provided herein is a method for determining a bladder cancer subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype. In some cases, the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set(s) comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal

sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm. In some cases, the expression level of the classifier biomarker is detected at the nucleic acid level. In some cases, the nucleic acid level is RNA or cDNA. In some cases, the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing RNAseq. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, a fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the at least one classifier biomarker comprises a plurality of classifier biomarkers. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2. In some cases, the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.

[0009] In another aspect, provided herein is a method for determining a muscle invasive bladder cancer (MIBC) subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table

1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype. In some cases, the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm. In some cases, the expression level of the classifier biomarker is detected at the nucleic acid level. In some cases, the nucleic acid level is RNA or cDNA. In some cases, the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing RNAseq. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the at least one classifier biomarker comprises a plurality of classifier biomarkers. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40

classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarkers of Table 1. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2. In some cases, the at least one classifier biomarker comprises, consists essentially of or consist of all the classifier biomarkers of Table 1 or Table 2.

[0010] In yet another aspect, provided herein is a method of detecting a biomarker in a bladder cancer sample obtained from a patient, the method comprising measuring the expression level of a plurality of classifier biomarker nucleic acids selected from Table 1 or Table 2 using an amplification, hybridization and/or sequencing assay. In some cases, the bladder cancer sample was previously diagnosed as being muscle invasive bladder cancer (MIBC). In some cases, the previous diagnosis was by histological examination. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing RNAseq. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from Table 1 or Table 2. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of

Table 2. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2.

[0011] In still another aspect, provided herein is a method of treating muscle invasive bladder cancer (MIBC) in a subject, the method comprising: measuring the expression level of at least one biomarker nucleic acid in a MIBC sample obtained from the subject, wherein the at least one biomarker nucleic acid is selected from a set of biomarkers listed in Table 1 or Table 2, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the MIBC; and administering a therapeutic agent based on the subtype of the MIBC. In some cases, the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1. In some cases, the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2. In some cases, the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2. In some cases, the method further comprises measuring the expression of at least one biomarker from an additional set of biomarkers. In some cases, the additional set of biomarkers comprises BIRC5, CCNB1, CDC20, CDCA1, CEP55, KNTC2, MKI67, PTTG1, RRM2, TYMS, UBE2C, TP53, RB1, FGFR2, FGFR3, and ERBB2. In some cases, the additional set of biomarkers comprise genes selected from Gene Expression Omnibus Dataset GSE87304, Gene Expression Omnibus Dataset GSE32894 or a combination thereof. In some cases, the additional set of biomarkers comprises at least an immune cell signature, a cell proliferation signature, or drug target genes. In some cases, the measuring the expression level is conducted using an amplification, hybridization and/or sequencing assay. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by

performing qRT-PCR. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the subject's MIBC subtype is selected from luminal, luminal infiltrated, basal, and basal infiltrated/neuronal. In some cases, the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises at least one biomarker nucleic acid listed in Table 1 or Table 2 in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the MIBC. In some cases, the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises all of the biomarker nucleic acids listed in Table 1 or Table 2, all of the biomarker nucleic acids listed in claim 45 or claim 46, in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the bladder cancer. In some cases, the publically available bladder cancer dataset is TCGA bladder cancer RNAseq dataset.

[0012] In a still further aspect, provided herein is a method of predicting overall survival in a muscularly invasive bladder cancer (MIBC) patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2 in a MIBC sample obtained from a patient, wherein the detection of the expression level of the at least one classifier biomarker specifically identifies a luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtype, and wherein identification of the subtype is predictive of the overall survival in the patient. In some cases, the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarkers of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the

results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm. In some cases, the expression level of the classifier biomarker is detected at the nucleic acid level. In some cases, the nucleic acid level is RNA or cDNA. In some cases, the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the at least one classifier biomarker comprises a plurality of classifier biomarkers. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1. In some cases, the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2. In some cases, the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] **FIG. 1** illustrates the bladder cancer subtype distribution for the 408 unique primary solid tumor samples described in Robertson, AG, et al., Cell, 171(3): 540-556 (2017) (dataset referred to herein as “Gold standard,” “GS,” or “TCGA-2017”) (upper left panel). **FIG. 1**

also provides the cluster assignment for the 129 tumor samples described in Cancer Genome Atlas Research Network, *Nature*, 507(492): 315-22 (2014) (dataset described herein as TCGA-2014) (upper right panel). The GS subtype silhouette is shown for the TCGA-2017 dataset (n=408) (bottom left panel). A solid dot indicates that a sample used to fit the GS standard subtyper was present in the TCGA-2014 dataset. Also, the GS subtype silhouette is shown for the 129 TCGA-2014 samples (bottom right panel). A solid dot indicates that the GS subtype and the TCGA-2014 subtype do not agree. Type I: Luminal; Type II: Luminal Infiltrated; Type III: Basal; Type IV: basal infiltrated/neuronal.

[0014] FIG. 2 illustrates five-fold cross validation curves using a Clanc plain (left panel) and Clanc 50:50 high: low (right panel) approach on the TCGA-2017 dataset (n=408) to guide the selection of the number of genes per subtype to include in the signature for bladder cancer subtyping provided herein.

[0015] FIG. 3 illustrates the GS subtype silhouette for the TCGA-2017 dataset (n=408), wherein non-prototype samples are shown with a solid dot. The non-prototype samples represent the bottom ~20% of each subtype, and due to their low predictive strength, were excluded from gene selection.

[0016] FIG. 4 illustrates Clanc tStats data used for gene selection. Shown are Clanc tStats data from the entire TCGA-2017 dataset, minus non-prototypes. The genes with a high-magnitude tStats value (shown in red) were selected for a 60-gene subtyper (top row; 50:50 high:low approach). Also shown are selected genes for a 56-gene subtyper (bottom row, 50:50 high:low approach).

[0017] FIG. 5 illustrates agreement and disagreement between the GS subtype (rows) and the subtype based on the 60-gene subtyper (columns) (left panel). Overall agreement was 87%. The GS subtype silhouette for the TCGA-2017 dataset (n=408) is also shown (right panel), wherein a solid dot indicates that the GS subtype and the 60-gene subtype don't agree.

[0018] FIG. 6 lists the 60 gene centroids selected for use in subtyping, and their corresponding gene expression levels (i.e., centroid classifier coefficients) by subtype.

[0019] FIG. 7 illustrates heatmaps comparing various markers (luminal markers, basal markers, squamous markers, and basal infiltrated/neuronal-differentiation markers) across defined bladder cancer subtypes (GS subtype, left panel and 60-gene subtype, right panel).

The heatmaps show a high degree of concordance between the GS subtype and the 60-gene subtype.

[0020] **FIG. 8** illustrates a heatmap of immune signatures across defined bladder cancer subtypes (60-gene subtypes). The immune signatures used to generate this data were derived from Bindea, G, et al., *Immunity*, 39(4):782-95 (2013); Faruki, H, et al., *J Thorac Oncol*, 12(6): 943-953 (2017); and Charoentong, P., et al., *Cell Reports*, 18(1): 248-262 (2017), the disclosures of which are incorporated herein by reference in their entireties.

[0021] **FIG. 9** provides survival curves of tumors from the TCGA-2017 dataset (n=402) based on clinical staging (stages II, III and IV in **FIG. 9**), showing that later-stage samples are correlated with a worse survival. Roman numerals indicate stage. Logrank $p = 8.39\text{e-}08$.

[0022] **FIG. 10A-B** provides survival curves by subtype in stage II, III and IV using the 60-gene subtype (**FIG. 10A**) and in stage II, III, and IV. Also shown are survival curves using the 60-gene subtype with the samples identified as “luminal” or “basal infiltrated/neuronal” in the TCGA-2017 data excluded (**FIG. 10B**). Survival differences were still evident even after “luminal” and “basal infiltrated/neuronal” TCGA-2017 subtypes were excluded (see **FIG. 10B**). OS-subtype association test p-value using cox model, $p=0.00262$; OS-subtype association test p-value using stratified (for stage) cox model, $p=0.0385$.

[0023] **FIG. 11** provides boxplots showing tumor mutation burden across the 60-gene signature subtypes. Single nucleotide variants are shown in the left panel, and indels are shown in the right panel. No notable differences between the subtypes were observed.

[0024] **FIG. 12** provides boxplots showing proliferation by subtype (60-gene signature) (top left panel). Subtypes III (basal) and IV (basal infiltrated/neuronal) were more proliferative than subtypes I (luminal) and II (luminal infiltrated). The gene signature used to determine proliferation is described in Martin, M., et al. *Breast Cancer Res Treat* 138:457-466 (2013), which is incorporated by reference in its entirety. **FIG. 12** also provides boxplots showing expression of potential treatment targets FGFR2, FGFR3, and ERBB2 by subtype.

[0025] **FIG. 13** illustrates tumor purity by ABSOLUTE differences across gene expression subtypes. ABSOLUTE tumor purity was measured using somatic copy number data from Robertson, AG, et al., *Cell*, 171(3): 540-556 (2017) (supplement). ABSOLUTE tumor purity was measured as described in Carter, S. L. et al. Absolute quantification of somatic DNA

alterations in human cancer. *Nat. Biotechnol.* 30, 413–421 (2012), which is incorporated herein by reference in its entirety.

[0026] FIG. 14 illustrates tumor driver (RB1, FGFR3, ERBB2) gene mutation frequencies across subtypes.

[0027] FIG. 15 illustrates frequency of TP53 tumor suppressor mutations across subtypes.

[0028] FIG. 16 illustrates the results of subtyping using the 60-gene signature in two additional datasets: the Seiler dataset (Seiler, R., et al., *Eur Urol* 72(4):544-554 (2017)) and the Sjodahl dataset (Sjodahl, G., et al., *Clin Cancer Res*, 18(12):3377-86 (2012)), both of which are incorporated by reference herein in their entireties. A comparison between the GS subtype (rows) and the 60-gene subtype (columns) was performed for the Seiler dataset (top left panel, agreement = 0.81) and the Sjodahl dataset (bottom left panel, agreement = 0.72). The GS subtype silhouette is shown for the Seiler dataset (top right panel) and for the Sjodahl dataset (bottom right panel). Solid dots indicate that the GS subtype and the 60-gene subtype do not agree.

[0029] FIG. 17A-C provides heatmaps illustrating immune profiles of the four bladder cancer gene expression subtypes (defined by the 60 gene signature) across the TCGA-2017 dataset (**FIG. 17A**), the Seiler dataset (**FIG. 17B**), and the Sjodahl dataset (**FIG. 17C**). In general, a high concordance was observed.

[0030] FIG. 18 provides boxplots showing t-cells and CD274 (PD-L1) by gene expression subtype in the various gene expression datasets. Similar patterns were observed in each dataset.

[0031] FIG. 19 provides boxplots showing proliferation and gene expression (FGFR3, FGFR2, ERBB2) across the various gene expression datasets.

[0032] FIG. 20 presents a comparison of the subtyping of the TCGA-2017 bladder cancer dataset (n=408) samples based on the 5-type classification system described in Roberston, et al. (2017), and the 4-type classification system based on the 60 gene subtype described herein.

[0033] FIG. 21 illustrates agreement and disagreement between the GS subtype (rows) and the subtype based on the 56-gene subtyper (columns) (left panel). Overall agreement was

88%. The GS subtype silhouette for the TCGA-2017 dataset (n=408) is also shown (right panel), wherein a solid dot indicates that the GS subtype and the 65-gene subtype do not agree.

[0034] FIG. 22 lists the 56 gene centroids selected for use in subtyping, and their corresponding gene expression levels (i.e., centroid classifier coefficients) by subtype.

DETAILED DESCRIPTION

Overview

[0035] The present invention provides kits, compositions and methods for identifying or diagnosing bladder cancer. That is, the methods can be useful for molecularly defining subsets of bladder cancer. The bladder cancer can be a muscle invasive bladder cancer (MIBC). The methods provide a classification of bladder cancer that can be prognostic and predictive for therapeutic response. The therapeutic response can include chemotherapy, immunotherapy, surgical intervention and radiotherapy. The methods can be also provide a prognosis of overall survival for bladder cancer patients according to their subtypes (e.g., luminal, luminal infiltrated, basal or basal infiltrated/neuronal).

[0036] While a useful term for epidemiologic purposes, "bladder cancer" can refer to cancers arising from the tissues of the urinary bladder that includes urothelial carcinoma, squamous cell carcinoma and adenocarcinoma. Bladder cancer can be described as cancers that are non-muscle invasive, muscle invasive, and transitional cell carcinoma. As used herein, bladder cancer can also refer to MIBC. Subtypes of these types of cancer as defined by underlying genomic features can have varied cell of origin, tumor drivers, proliferation, immune responses, and prognosis.

[0037] "Determining a bladder cancer subtype" can include, for example, diagnosing or detecting the presence and sub-type of bladder cancer, monitoring the progression of the disease, and identifying or detecting cells or samples that are indicative of subtypes.

[0038] In one embodiment, bladder cancer status is assessed through the evaluation of expression patterns, or profiles, of a plurality of classifier biomarkers or biomarkers in one or more subject samples. The term subject, or subject sample, may refer to an

individual regardless of health and/or disease status. A subject can be a subject, a study participant, a control subject, a screening subject, or any other class of individual from whom a sample is obtained and assessed in the context of the methods, compositions and kits provided herein. Accordingly, a subject can be diagnosed with bladder cancer (including subtypes, or grades thereof), can present with one or more symptoms of bladder cancer, or a predisposing factor, such as a family (genetic) or medical history (medical) factor, for bladder cancer, can be undergoing treatment or therapy for bladder cancer, or the like. Alternatively, a subject can be healthy with respect to any of the aforementioned factors or criteria. It will be appreciated that the term "healthy" as used herein, is relative to bladder cancer status, as the term "healthy" cannot be defined to correspond to any absolute evaluation or status. Thus, an individual defined as healthy with reference to any specified disease (e.g., bladder cancer) or disease criterion, can in fact be diagnosed with any other one or more diseases, or exhibit any other one or more disease criterion, including one or more other cancers.

[0039] As used herein, an "expression profile" or a "biomarker profile" or "gene signature" comprises one or more values corresponding to a measurement of the relative abundance, level, presence, or absence of expression of a discriminative or classifier biomarker or biomarker. An expression profile can be derived from a subject prior to or subsequent to a diagnosis of bladder cancer, can be derived from a biological sample collected from a subject at one or more time points prior to or following treatment or therapy, can be derived from a biological sample collected from a subject at one or more time points during which there is no treatment or therapy (e.g., to monitor progression of disease or to assess development of disease in a subject diagnosed with or at risk for bladder cancer), or can be collected from a healthy subject. The term subject can be used interchangeably with patient. The patient can be a human patient. The one or more biomarkers of the biomarker profiles provided herein are selected from one or more biomarkers of Table 1, Table 2 and/or any additional set of biomarker classifiers disclosed herein.

[0040] As used herein, the term "determining an expression level" or "determining an expression profile" or "detecting an expression level" or "detecting an expression profile" as used in reference to a biomarker or classifier means the application of a biomarker specific reagent such as a probe, primer or antibody and/or a method to a sample, for example a sample of the subject or patient and/or a control sample, for ascertaining or measuring

quantitatively, semi-quantitatively or qualitatively the amount of a biomarker or biomarkers, for example the amount of biomarker polypeptide or mRNA (or cDNA derived therefrom). For example, a level of a biomarker can be determined by a number of methods including for example immunoassays including for example immunohistochemistry, ELISA, Western blot, immunoprecipitation and the like, where a biomarker detection agent such as an antibody for example, a labeled antibody, specifically binds the biomarker and permits for example relative or absolute ascertaining of the amount of polypeptide biomarker, hybridization and PCR protocols where a probe or primer or primer set are used to ascertain the amount of nucleic acid biomarker, including for example probe based and amplification based methods including for example microarray analysis, RT-PCR such as quantitative RT-PCR (qRT-PCR), serial analysis of gene expression (SAGE), Northern Blot, digital molecular barcoding technology, for example Nanostring Counter Analysis, and TaqMan quantitative PCR assays. Other methods of mRNA detection and quantification can be applied, such as mRNA in situ hybridization in formalin-fixed, paraffin-embedded (FFPE) tissue samples or cells. This technology is currently offered by the QuantiGene ViewRNA (Affymetrix), which uses probe sets for each mRNA that bind specifically to an amplification system to amplify the hybridization signals; these amplified signals can be visualized using a standard fluorescence microscope or imaging system. This system for example can detect and measure transcript levels in heterogeneous samples; for example, if a sample has normal and tumor cells present in the same tissue section. As mentioned, TaqMan probe-based gene expression analysis (PCR-based) can also be used for measuring gene expression levels in tissue samples, and this technology has been shown to be useful for measuring mRNA levels in FFPE samples. In brief, TaqMan probe-based assays utilize a probe that hybridizes specifically to the mRNA target. This probe contains a quencher dye and a reporter dye (fluorescent molecule) attached to each end, and fluorescence is emitted only when specific hybridization to the mRNA target occurs. During the amplification step, the exonuclease activity of the polymerase enzyme causes the quencher and the reporter dyes to be detached from the probe, and fluorescence emission can occur. This fluorescence emission is recorded and signals are measured by a detection system; these signal intensities are used to calculate the abundance of a given transcript (gene expression) in a sample.

[0041] In one embodiment, the “expression profile” or a “biomarker profile” or “gene signature” associated with the gene signatures or classifier biomarkers described herein (e.g., Table 1, Table 2 and/or any additional set of biomarker classifiers as disclosed herein)

can be useful for distinguishing between normal and tumor samples. In another embodiment, the tumor samples are bladder cancer. In another embodiment, the bladder cancer can be further classified as luminal, luminal infiltrated, basal, or basal infiltrated/neuronal based upon an expression profile determined using the methods provided herein. In some embodiments, the bladder cancer is MIBC. Expression profiles using the classifier biomarkers disclosed herein (e.g., Table 1, Table 2 and any additional set of biomarker classifiers as disclosed herein) can provide valuable molecular tools for specifically identifying bladder cancer subtypes, and for treating bladder cancer. Accordingly, the invention provides methods for screening and classifying a subject for molecular bladder cancer subtypes.

[0042] In some instances, a single classifier biomarker or a plurality of classifier biomarkers provided herein is capable of identifying subtypes of bladder cancer with a predictive success of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, inclusive of all ranges and subranges therebetween.

[0043] In some instances, a single classifier biomarker or a plurality of classifier biomarkers as provided herein is capable of determining bladder cancer subtypes with a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, inclusive of all ranges and subranges therebetween.

[0044] The present invention also encompasses a system capable of distinguishing various subtypes of bladder cancer not detectable using current methods. This system can be

capable of processing a large number of subjects and subject variables such as expression profiles and other diagnostic criteria. The expression profile can be used in combination with other diagnostic methods including histochemical, immunohistochemical, cytologic, immunocytologic, and visual diagnostic methods including histologic or morphometric evaluation of bladder tissue.

[0045] In various embodiments, the expression profile derived from a subject is compared to a reference expression profile. A “reference expression profile” or “control expression profile” can be a profile derived from the subject prior to treatment or therapy; can be a profile produced from the subject sample at a particular time point (usually prior to or following treatment or therapy, but can also include a particular time point prior to or following diagnosis of bladder cancer); or can be derived from a healthy individual or a pooled reference from healthy individuals. A reference expression profile can be generic for bladder cancer or can be specific to different subtypes of bladder cancer. The bladder cancer reference expression profile can be from any tissues of the urinary bladder.

[0046] The reference expression profile can be compared to a test expression profile. A “test expression profile” can be derived from the same subject as the reference expression profile except at a subsequent time point (e.g., one or more days, weeks or months following collection of the reference expression profile) or can be derived from a different subject. In summary, any test expression profile of a subject can be compared to a previously collected profile from a subject that has a luminal, luminal infiltrated, basal, basal infiltrated/neuronal bladder cancer subtype.

[0047] The classifier biomarkers of the invention can include nucleic acids (RNA, cDNA, and DNA) and proteins, and variants and fragments thereof. Such biomarkers can include DNA comprising the entire or partial sequence of the nucleic acid sequence encoding the biomarker, or the complement of such a sequence. The biomarkers described herein can include RNA comprising the entire or partial sequence of any of the nucleic acid sequences of interest, or their non-natural cDNA products, obtained synthetically *in vitro* in a reverse transcription reaction. The biomarker nucleic acids can also include any expression product or portion thereof of the nucleic acid sequences of interest. A biomarker protein can be a protein encoded by or corresponding to a DNA biomarker as provided herein. A biomarker protein can comprise the entire or partial amino acid sequence of any of the biomarker proteins or polypeptides. The biomarker nucleic acid

can be extracted from a cell or can be cell free or extracted from an extracellular vesicular entity such as an exosome.

[0048] A “classifier biomarker” or “biomarker” or “classifier gene” can be any gene or protein whose level of expression in a tissue or cell is altered compared to that of a normal or healthy cell or tissue. For example, a “classifier biomarker” or “biomarker” or “classifier gene” can be any gene or protein whose level of expression in a tissue or cell is altered in a specific bladder cancer subtype. The detection of the biomarkers as provided herein can permit the determination of the specific subtype. The “classifier biomarker” or “biomarker” or “classifier gene” may be one that is up-regulated (e.g. expression is increased) or down-regulated (e.g. expression is decreased) relative to a reference or control as provided herein. The reference or control can be any reference or control as provided herein. In some embodiments, the expression values of genes that are up-regulated or down-regulated in a particular subtype of bladder cancer can be pooled into one gene signature. The overall expression level of each gene in each gene signature is referred to herein as the “expression profile” and is used to classify a test sample according to the subtype of bladder cancer. However, it is understood that independent evaluation of expression for each of the genes disclosed herein can be used to classify tumor subtypes without the need to group up-regulated and down-regulated genes into one or more gene signatures. In some cases, as shown in **Table 1**, a total of 60 biomarkers can be used for bladder cancer subtype determination. For each bladder cancer subtype, for example, 30 of the 60 biomarkers can be negatively correlated genes, while 30 of the 60 biomarkers can be positively correlated genes which can be selected as the gene signature of a specific bladder cancer subtype. In some cases, as shown in **Table 2**, a total of 56 biomarkers can be used for bladder cancer subtype determination.

[0049] The classifier biomarkers as provided herein can include any gene or protein that is selectively expressed in bladder cancer, as defined herein above. Sample biomarker genes are listed in **Table 1** or **Table 2**, below. In **Table 1**, the first column of the table represents the biomarker list selected for distinguishing luminal subtype. The second column of the table represents the biomarker list selected for distinguishing luminal infiltrated subtype. The third column of the table represents the biomarker list selected for distinguishing basal subtype. The last column of the table represents the biomarker list

selected for distinguishing basal infiltrated/neuronal subtype. In **Table 2**, the first column of the table represents the biomarker list selected for distinguishing luminal subtype. The second column of the table represents the biomarker list selected for distinguishing luminal infiltrated subtype. The third column of the table represents the biomarker list selected for distinguishing basal subtype. The last column of the table represents the biomarker list selected for distinguishing basal infiltrated/neuronal subtype.

[0050] The relative gene expression levels as represented by the centroid coefficients as described herein of the classifier biomarkers for 60-gene bladder cancer subtyping are shown in **Table 1**. In one embodiment, the gene centroids including gene expression levels (i.e., centroid coefficients) of the classifier biomarkers for bladder cancer subtyping are shown in **Table 1**. The relative gene expression levels as represented by the centroid coefficients as described herein of the classifier biomarkers for 56-gene bladder cancer subtyping are shown in **Table 2**. In one embodiment, the gene centroids including gene expression levels (i.e., centroid coefficients) of the classifier biomarkers for bladder cancer subtyping are shown in **Table 2**.

[0051] **Table 1. Gene Centroids of 60 Classifier Biomarkers for the Bladder Cancer Subtypes**

Gene Symbol	Gene Name	I (Luminal)	II (Luminal Infiltrated)	III (Basal)	IV (Neuronal/Basal Infiltrated)	SEQ ID NO:	GenBank Accession Number*
ALDH1L2	aldehyde dehydrogenase 1 family member L2	-2.95	0.51	0.59	1.31	1	NM_001034173
ANXA6	annexin A6	-2.00	0.64	0.09	1.66	2	NM_001155
ARSI	arylsulfatase family member I	-3.67	0.42	1.72	1.23	3	NM_001012301
BCAS1	breast carcinoma amplified sequence 1	1.88	1.10	-3.83	-5.58	4	NM_003657
BNC1	basonuclin 1	-1.48	-0.57	7.97	2.24	5	NM_001717
C10orf99	chromosome 10 open reading frame 99	1.69	1.15	-1.26	-7.80	6	NM_207373
C17orf28	HID1 domain containing	0.80	0.61	-3.52	-2.18	7	NM_030630
CAPN5	calpain 5	1.17	1.17	-2.49	-1.86	8	NM_004055
CCDC80	coiled-coil domain containing 80	-3.08	1.14	0.43	1.87	9	NM_199511
COL6A2	collagen type VI alpha 2 chain	-2.25	1.10	0.36	2.02	10	NM_058174
CPXM2	carboxypeptidase X, M14 family member 2	-2.14	1.78	-0.13	1.70	11	NM_198148

CTHRC1	collagen triple helix repeat containing 1	-3.30	0.61	0.64	1.15	12	NM_138455
DSG3	desmoglein 3	-2.16	-1.59	7.87	-1.24	13	NM_001944
EMILIN1	elastin microfibril interfacier 1	-1.96	1.86	-0.20	1.84	14	NM_007046
EPN3	epsin 3	0.61	0.45	-0.73	-2.60	15	NM_017957
EVPL	envoplakin	0.51	0.48	-0.65	-2.53	16	NM_001320747
FAP	fibroblast activation protein alpha	-3.87	0.94	0.84	1.46	17	NM_004460
FBN1	fibrillin 1	-2.47	1.02	0.22	1.82	18	NM_000138
FGF7	fibroblast growth factor 7	-2.23	1.98	0.12	1.44	19	NM_002009
FMO9P	flavin containing monooxygenase 9 pseudogene	2.37	3.22	-3.99	-5.26	20	NR_002925
FNDC1	fibronectin type III domain containing 1	-4.02	2.05	0.42	2.44	21	NM_032532
GABBR2	gamma-aminobutyric acid type B receptor subunit 2	0.63	5.25	-2.87	-0.72	22	NM_005458
GFPT2	glutamine-fructose-6-phosphate transaminase 2	-4.37	0.83	0.74	1.95	23	NM_005110
GGT6	gamma-glutamyltransferase 6	1.18	0.33	-2.23	-5.61	24	NM_001122890
GREM1	gremlin 1, DAN family BMP antagonist	-5.82	1.63	0.66	0.50	25	NM_013372
GRHL3	grainyhead like transcription factor 3	1.02	1.46	-1.39	-6.23	26	NM_021180
IL20RB	interleukin 20 receptor subunit beta	-0.89	-0.97	4.28	-0.59	27	NM_144717
KRT6A	keratin 6A	-2.15	-2.48	7.61	-0.59	28	NM_005554
KRT6B	keratin 6B	-1.53	-2.24	7.55	-0.39	29	NM_005555
KRT6C	keratin 6C	-1.76	-2.57	7.25	-1.05	30	NM_173086
LMOD1	leiomodien 1	-1.31	2.77	-0.22	0.96	31	NM_012134
LOC100188947	HECTD2 antisense RNA 1	3.20	2.48	-3.88	-4.38	32	NR_024467
MRVI1	murine retrovirus integration site 1 homolog	-1.19	1.65	-0.15	0.54	33	NM_001098579
NRP2	neuropilin 2	-2.34	0.40	0.62	1.43	34	NM_201266
PDLIM3	PDZ and LIM domain 3	-2.49	1.95	0.33	1.51	35	NM_014476.
PLA2G4F	phospholipase A2 group IVF	1.13	0.24	-0.66	-4.12	36	NM_213600
PODN	podocan	-1.54	1.82	-0.72	1.20	37	NM_153703
POSTN	periostin	-4.38	1.37	0.46	1.42	38	NM_006475
PRRX1	paired related homeobox 1	-3.33	0.94	0.50	2.06	39	NM_006902
PVRL4	nectin cell adhesion molecule 4	0.49	0.36	-0.39	-2.97	40	NM_030916

RAPGEFL1	Rap guanine nucleotide exchange factor like 1	1.01	0.10	-0.43	-3.66	41	NM_001303533
RHOU	ras homolog family member U	0.59	1.22	-2.78	-1.70	42	NM_021205
RHOV	ras homolog family member V	-0.17	0.20	0.89	-2.93	43	NM_133639
SCUBE2	signal peptide, CUB domain and EGF like domain containing 2	1.29	3.00	-3.51	-1.86	44	NM_020974
SDC1	syndecan 1	0.39	0.13	-0.01	-2.41	45	NM_001006946
SERPINB13	serpin family B member 13	-0.64	-2.10	5.68	-2.94	46	NM_001307923
SFRP2	secreted frizzled related protein 2	-6.65	2.50	0.68	1.68	47	NM_003013
SFRP4	secreted frizzled related protein 4	-5.67	3.09	0.38	2.19	48	NM_003014
SLC30A2	solute carrier family 30 member 2	2.17	3.26	-4.60	-3.18	49	NM_001004434
SMOC2	SPARC related modular calcium binding 2	-1.41	2.14	-0.44	1.00	50	NM_022138
SNX31	sorting nexin 31	1.40	2.05	-6.48	-7.24	51	NM_152628
SPRR2A	small proline rich protein 2A	-1.49	-0.32	5.64	-1.95	52	NM_005988
SSC5D	scavenger receptor cysteine rich family member with 5 domains	-2.38	1.82	0.09	1.78	53	NM_001144950
TBX3	T-box 3	1.48	0.55	-3.24	-2.08	54	NM_005996
TLE2	transducin like enhancer of split 2	0.96	0.69	-2.92	-1.06	55	NM_003260
TOX3	TOX high mobility group box family member 3	2.44	2.08	-6.52	-5.74	56	NM_001080430
UPK1A	uroplakin 1A	1.94	2.82	-5.92	-6.97	57	NM_007000
UPK2	uroplakin 2	1.85	2.45	-5.94	-6.04	58	NM_006760
UPK3A	uroplakin 3A	1.88	3.61	-5.65	-4.89	59	NM_006953
ZNF750	zinc finger protein 750	-0.07	0.52	0.82	-3.97	60	NM_024702

*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

[0052] Table 2. Gene Centroids of 56 Classifier Biomarkers for the Bladder Cancer Subtypes

Gene Symbol	Gene Name	I (Luminal)	II (Luminal Infiltrated)	III (Basal)	IV (Basal infiltrated /neural)	SEQ ID NO.	GenBank Accession Number*
ALDH1L2	aldehyde dehydrogenase	-2.95	0.51	0.59	1.31	1	NM_001034173

	1 family member L2						
ANXA6	annexin A6	-2.00	0.64	0.09	1.66	2	NM_001155
ARSI	arylsulfatase family member I	-3.67	0.42	1.72	1.23	3	NM_001012301
BNC1	basonuclin 1	-1.48	-0.57	7.97	2.24	5	NM_001717
C17orf28	HID1 domain containing	0.80	0.61	-3.52	-2.18	7	NM_030630
C1orf126	TMEM51 antisense RNA 1	1.65	0.18	-1.38	-1.99	61	NR_027136
CA9	carbonic anhydrase 9	0.70	-4.00	1.43	-1.19	62	NM_001216
CAPN5	calpain 5	1.17	1.17	-2.49	-1.86	8	NM_004055
CCDC80	coiled-coil domain containing 80	-3.08	1.14	0.43	1.87	9	NM_199511
CD109	CD109 molecule	-2.26	-0.32	2.62	1.52	63	NM_133493
CD163	CD163 molecule	-2.95	0.75	0.40	2.64	64	NM_004244
CMKLR1	chemerin chemokine-like receptor 1	-2.01	0.54	0.30	2.38	65	NM_001142343
CTHRC1	collagen triple helix repeat containing 1	-3.30	0.61	0.64	1.15	12	NM_138455
DSG3	desmoglein 3	-2.16	-1.59	7.87	-1.24	13	NM_001944
E2F7	E2F transcription factor 7	-0.19	-1.12	0.99	0.36	66	NM_203394
EMILIN1	elastin microfibril interfacier 1	-1.96	1.86	-0.20	1.84	14	NM_007046
EMP3	epithelial membrane protein 3	-1.89	0.34	0.70	2.16	67	NM_001425
FAM174B	family with sequence similarity 174 member B	1.46	0.57	-3.03	-1.94	68	NM_207446
FER1L4	fer-1 like family member 4, pseudogene	1.71	0.68	-3.30	-3.10	69	NR_119376
GFPT2	glutamine-fructose-6-phosphate transaminase 2	-4.37	0.83	0.74	1.95	23	NM_005110
GGT6	gamma-glutamyltransferrase 6	1.18	0.33	-2.23	-5.61	24	NM_001122890
GRHL3	grainyhead like transcription factor 3	1.02	1.46	-1.39	-6.23	26	NM_021180
HMGCS2	3-hydroxy-3-methylglutaryl-CoA synthase 2	3.62	2.44	-6.28	-6.91	70	NM_005518
IL20RB	interleukin 20 receptor subunit beta	-0.89	-0.97	4.28	-0.59	27	NM_144717
KRT6A	keratin 6A	-2.15	-2.48	7.61	-0.59	28	NM_005554
KRT6B	keratin 6B	-1.53	-2.24	7.55	-0.39	29	NM_005555

KRT6C	keratin 6C	-1.76	-2.57	7.25	-1.05	30	NM_173086
LOC100188947	HECTD2 antisense RNA 1	3.20	2.48	-3.88	-4.38	32	NR_024467
MRVI1	murine retrovirus integration site 1 homolog	-1.19	1.65	-0.15	0.54	33	NM_001098579
PIK3AP1	phosphoinositide-3-kinase adaptor protein 1	-1.83	0.19	0.50	2.18	71	NM_152309
PPFIBP2	PPFIA binding protein 2	1.22	0.73	-1.84	-2.20	72	NM_003621
PVRL4	nectin cell adhesion molecule 4	0.49	0.36	-0.39	-2.97	40	NM_030916
RAPGEFL1	Rap guanine nucleotide exchange factor like 1	1.01	0.10	-0.43	-3.66	41	NM_001303533
RHOU	ras homolog family member U	0.59	1.22	-2.78	-1.70	42	NM_021205
RHOV	ras homolog family member V	-0.17	0.20	0.89	-2.93	43	NM_133639
SARM1	sterile alpha and TIR motif containing 1	0.52	0.47	-1.87	-0.36	73	NM_015077
SDC1	syndecan 1	0.39	0.13	-0.01	-2.41	45	NM_001006946
SERPINB13	serpin family B member 13	-0.64	-2.10	5.68	-2.94	46	NM_001307923
SERPINB3	serpin family B member 3	-0.99	-4.25	5.93	-2.21	74	NM_006919
SERPINB4	serpin family B member 4	-1.12	-3.51	5.84	-1.62	75	NM_002974
SFRP2	secreted frizzled related protein 2	-6.65	2.50	0.68	1.68	47	NM_003013
SFRP4	secreted frizzled related protein 4	-5.67	3.09	0.38	2.19	48	NM_003014
SLC7A7	solute carrier family 7 member 7	-2.05	0.37	0.47	2.01	76	NR_040448
SNX31	sorting nexin 31	1.40	2.05	-6.48	-7.24	51	NM_152628
SPRR2A	small proline rich protein 2A	-1.49	-0.32	5.64	-1.95	52	NM_005988
TBX2	T-box 2	1.54	1.51	-2.76	-1.55	77	NM_005994
TBX3	T-box 3	1.48	0.55	-3.24	-2.08	54	NM_005996
TLE2	transducin like enhancer of split 2	0.96	0.69	-2.92	-1.06	55	NM_003260
TOX3	TOX high mobility group box family member 3	2.44	2.08	-6.52	-5.74	56	NM_001080430
TSSK2	testis specific serine kinase 2	0.00	0.00	0.00	0.00	78	NM_053006
UPK1A	uroplakin 1A	1.94	2.82	-5.92	-6.97	57	NM_007000
UPK2	uroplakin 2	1.85	2.45	-5.94	-6.04	58	NM_006760

UPK3A	urolakin 3A	1.88	3.61	-5.65	-4.89	59	NM_006953
VSIG4	V-set and immunoglobulin domain containing 4	-2.66	0.59	0.58	2.61	79	NM_007268
ZNF750	zinc finger protein 750	-0.07	0.52	0.82	-3.97	60	NM_024702

*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

[0053] In one embodiment, the relative gene expression levels (e.g., centroid classifier coefficients) of the classifier biomarkers for bladder cancer subtyping are shown in **Table 1** or **Table 2**. In one embodiment, all 60 genes of **Table 1**, or all 56 genes of **Table 2**, can be used to classify the subtypes of bladder cancer. In one embodiment, all 60 genes of **Table 1**, or all 56 genes of **Table 2**, can be used to classify the subtypes of MIBC. In some embodiments, the up-regulation of a classifier biomarker (e.g. expression is increased) can refer to an expression value that is positive (i.e., higher than zero) relative to a reference or control as provided herein. In some embodiments, the down-regulation of a classifier biomarker (e.g. expression is decreased) can refer to an expression value that is negative (i.e., lower than zero) relative to a reference or control as provided herein. In some embodiments, a classifier biomarker may have no specific effects on a certain bladder cancer subtype when the expression level equals to zero. For example, ALDH1L2 is down-regulated in the luminal subtype, whereas it is up-regulated in luminal infiltrated, basal, and basal infiltrated/neuronal subtypes. In another example, BCAS1 is up-regulated in both luminal and luminal infiltrated subtypes, whereas it is down-regulated in both basal and basal infiltrated/neuronal subtypes.

[0054] In some embodiments, determining bladder cancer (e.g., MIBC) subtypes can further include measuring the expression of at least one biomarker from an additional set of biomarker classifiers. In some embodiments, an additional set of biomarker classifiers can include an 11 gene signature comprising BIRC5, CCNB1, CDC20, CDCA1, CEP55, KNTC2, MKI67, PTTG1, RRM2, TYMS, and UBE2C (Martin M. et al., Breast Cancer Res Treat, 138: 457-466 (2013), which is herein incorporated by reference). In one embodiment, the 11 gene signature is related to the signature of cell proliferation. In some embodiments, an additional set of biomarker classifiers can include an 18 gene signature found in US2016115551, which is herein incorporated by reference. In one embodiment, the 18 gene

signature is related to the signature of cell proliferation. In some embodiments, an additional set of biomarker classifiers can include a 26 gene signature found in US62/789,668 filed January 8, 2019, which is herein incorporated by reference. In one embodiment, the 26 gene signature is related to the signature of cell proliferation. In some embodiments, an additional set of biomarker classifiers can include a 5 gene signature comprising tumor driver genes such as TP53 and RB1, and receptor tyrosine kinases including FGFR2, FGFR3, and ERBB2. In one embodiment, the 5 gene signature is related to the signature of tumor driver genes. In some embodiments, the biomarker classifiers can also include immune cell signatures that are known in the art such as the signatures found in Thorsson, V. et al., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830. Bindea G. et al., *Immunity*, 39(4), 782-95 (2013), Faruki H. et al., *JTO*, 12(6): 943-953 (2017), Charoentong P. et al., *Cell reports*, 18, 248-262 (2017) and/or WO2017/201165 and WO2017/201164, each of which is herein incorporated by reference). In some embodiments, an additional set of biomarker classifiers can include assessing tumor purity ABSOLUTE derived from the TCGA supplementary data. In some embodiments, an additional set of biomarker classifiers can include the bladder cancer biomarker signature described in Gene Expression Omnibus (GEO) dataset: GSE87304, Seiler R. et al., *Eur Urol*, 72(4):544-554 (2017); Gene Expression Omnibus (GEO) dataset: GSE32894, Sjödaahl G. et al., *Clin Cancer Res*, 18(12):3377-86 (2012), each of which is herein incorporated by reference).

[0055] In some embodiments, determining bladder cancer subtypes can further include assessing tumor mutation burden (TMB) and/or TMB rate. In one embodiment, the TMB value and/or rate can be calculated using any method known in the art. In one embodiment, the TMB value and/or rate can be calculated from RNA (e.g., via transcriptome profiling or RNA sequencing)) as provided in US 62/771,702 filed November 27, 2018 and US 62/743,257 filed October 9, 2018, which is herein incorporated by reference herein.

[0056] In some embodiments, determining bladder cancer (e.g., MIBC) subtypes can further include determining a cell of origin subtype of the bladder cancer sample. In one embodiment, cell of origin subtype is determined using any method known in the art such as, for example, as provided in Hoadley et al, *Cell*. 2018 Apr 5;173(2):291-304, which is herein incorporated by reference herein.

Diagnostic Uses

[0057] In one embodiment, the methods and compositions provided herein allow for the differentiation of the four subtypes of bladder cancer, or MIBC: (I) luminal, (II) luminal infiltrated, (III) basal, and (IV) basal infiltrated/neuronal.

[0058] In general, the methods provided herein are used to classify a bladder cancer sample as a particular bladder cancer subtype (e.g. subtype of bladder cancer). In one embodiment, the method comprises measuring, detecting or determining an expression level of at least one of the classifier biomarkers of any publically available bladder cancer expression dataset. In one embodiment, the method comprises detecting or determining an expression level of at least one of the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers as disclosed herein in a bladder cancer sample obtained from a patient or a subject. The bladder cancer sample for the detection or differentiation methods described herein can be a sample previously determined or diagnosed as an MIBC sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists.

[0059] In one embodiment, the measuring or detecting step is at the nucleic acid level by performing RNA-seq, a reverse transcriptase polymerase chain reaction (RT-PCR) or a hybridization assay with oligonucleotides that are substantially complementary to portions of cDNA molecules of the at least one classifier biomarker (such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers as disclosed herein) under conditions suitable for RNA-seq, RT-PCR or hybridization and obtaining expression levels of the at least one classifier biomarkers based on the detecting step. The expression levels of the at least one of the classifier biomarkers are then compared to reference expression levels of the at least one of the classifier biomarkers (such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers as disclosed herein) from at least one sample training set. The at least one sample training set can comprise, (i) expression levels of the at least one biomarker from a sample that overexpresses the at least one biomarker, (ii) expression levels from a reference luminal, luminal infiltrated, basal, and basal infiltrated/neuronal sample, or (iii) expression levels from MIBC sample, and classifying the bladder cancer sample as a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal subtype based on the results of the comparing step. In one embodiment, the comparing step can comprise applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the bladder tissue or cancer sample and the

expression data from the at least one training set(s); and classifying the bladder tissue or cancer sample as a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal sample subtype based on the results of the statistical algorithm. In one embodiment, the statistical algorithm for the comparing step can be an algorithm that comprises determining a correlation between the expression data obtained from the bladder tissue or cancer sample and centroids constructed from the expression levels or profiles measured or detected for the at least one classifier biomarkers (such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers as disclosed herein) from the at least one training set. The subtype for the bladder cancer sample can then be assigned by finding the centroid to which it is nearest from the centroids constructed from the expression data from the at least one training set, using any distance measure e.g. Euclidean distance or correlation. The centroids can be constructed using any method known in the art for generating centroids such as, for example, those found in Mullins et al. (2007) Clin Chem. 53(7):1273-9 or Dabney (2005) Bioinformatics 21(22):4148-4154 Subtype can then be assigned to the bladder cancer sample obtained from subject based on the use of a classification to the nearest centroid (CLaNC) algorithm as applied to the expression data generated from the bladder cancer sample and the centroid(s) constructed for the at least one training set. The CLaNC algorithm for use in the methods, compositions and kits provided herein can be the CLaNC algorithm implemented by the CLaNC software found in Dabney AR. CLaNC: Point-and-click software for classifying microarrays to nearest centroids. Bioinformatics. 2006;22: 122–123 or equivalents or derivatives thereof.

[0060] In one embodiment, the method comprises probing the levels of at least one of the classifier biomarkers provided herein, such as the classifier biomarkers of Table 1, Table 2 or any additional set of biomarker classifiers disclosed herein, at the nucleic acid level, in a bladder cancer sample obtained from the patient. The bladder cancer sample can be a sample previously determined or diagnosed as a MIBC sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists. The probing step, in one embodiment, comprises mixing the sample with one or more oligonucleotides that are substantially complementary to portions of cDNA molecules of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of Table 1, Table 2 or any additional set of biomarker classifiers disclosed herein under conditions suitable for hybridization of the one or more oligonucleotides to their complements or substantial complements; detecting whether hybridization occurs between

the one or more oligonucleotides to their complements or substantial complements; and obtaining hybridization values of the at least one classifier biomarkers based on the detecting step. The hybridization values of the at least one classifier biomarkers are then compared to reference hybridization value(s) from at least one sample training set. The bladder cancer sample is classified, for example, as luminal, luminal infiltrated, basal, or basal infiltrated/neuronal based on the results of the comparing step. In one embodiment, the hybridization values of the bladder cancer sample can be compared to centroids constructed from the hybridization values obtained for the training set.

[0061] The bladder tissue sample can be any sample isolated from a human subject or patient. For example, in one embodiment, the analysis is performed on bladder biopsies that are embedded in paraffin wax. In one embodiment, the sample can be a fresh frozen bladder tissue sample. In another embodiment, the sample can be a bodily fluid obtained from the patient. The bodily fluid can be blood or fractions thereof (i.e., serum, plasma), urine, saliva, sputum or cerebrospinal fluid (CSF). The sample can contain cellular as well as extracellular sources of nucleic acid for use in the methods provided herein. The extracellular sources can be cell-free DNA and/or exosomes. In one embodiment, the sample can be a cell pellet or a wash. This aspect provides a means to improve current diagnostics by accurately identifying the major histological types, even from small biopsies. The methods as provided herein, including the RT-PCR methods, are sensitive, precise and have multi-analyte capability for use with paraffin-embedded samples. See, for example, Cronin *et al.* (2004) *Am. J Pathol.* 164(1):35-42, herein incorporated by reference.

[0062] Formalin fixation and tissue embedding in paraffin wax is a universal approach for tissue processing prior to light microscopic evaluation. A major advantage afforded by formalin-fixed paraffin-embedded (FFPE) specimens is the preservation of cellular and architectural morphologic detail in tissue sections. (Fox et al. (1985) *J Histochem Cytochem* 33:845-853). The standard buffered formalin fixative in which biopsy specimens are processed is typically an aqueous solution containing 37% formaldehyde and 10-15% methyl alcohol. Formaldehyde is a highly reactive dipolar compound that results in the formation of protein-nucleic acid and protein-protein crosslinks in vitro (Clark et al. (1986) *J Histochem Cytochem* 34:1509-1512; McGhee and von Hippel (1975) *Biochemistry* 14:1281-1296, each incorporated by reference herein).

[0063] In one embodiment, the sample used herein is obtained from an individual, and comprises formalin-fixed paraffin-embedded (FFPE) tissue. However, other tissue and sample types are amenable for use herein. In one embodiment, the other tissue and sample types can be fresh frozen tissue, wash fluids, or cell pellets, or the like. In one embodiment, the sample can be a bodily fluid obtained from the individual. The bodily fluid can be blood or fractions thereof (e.g., serum, plasma), urine, sputum, saliva or cerebrospinal fluid (CSF). A biomarker nucleic acid as provided herein can be extracted from a cell or can be cell free or extracted from an extracellular vesicular entity such as an exosome.

[0064] Methods are known in the art for the isolation of RNA from FFPE tissue. In one embodiment, total RNA can be isolated from FFPE tissues as described by Bibikova et al. (2004) American Journal of Pathology 165:1799-1807, herein incorporated by reference. Likewise, the High Pure RNA Paraffin Kit (Roche) can be used. Paraffin is removed by xylene extraction followed by ethanol wash. RNA can be isolated from sectioned tissue blocks using the MasterPure Purification kit (Epicenter, Madison, Wis.); a DNase I treatment step is included. RNA can be extracted from frozen samples using Trizol reagent according to the supplier's instructions (Invitrogen Life Technologies, Carlsbad, Calif.). Samples with measurable residual genomic DNA can be resubjected to DNaseI treatment and assayed for DNA contamination. All purification, DNase treatment, and other steps can be performed according to the manufacturer's protocol. After total RNA isolation, samples can be stored at -80 °C until use.

[0065] General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al., ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York 1987-1999. Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp and Locker (Lab Invest. 56:A67, 1987) and De Andres et al. (Biotechniques 18:42-44, 1995). In particular, RNA isolation can be performed using a purification kit, a buffer set and protease from commercial manufacturers, such as Qiagen (Valencia, Calif.), according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Other commercially available RNA isolation kits include MasterPure™. Complete DNA and RNA Purification Kit (Epicentre, Madison, Wis.) and Paraffin Block RNA Isolation Kit (Ambion, Austin, Tex.). Total RNA from tissue samples can be isolated, for example, using RNA Stat-60 (Tel-Test, Friendswood, Tex.). RNA

prepared from a tumor can be isolated, for example, by cesium chloride density gradient centrifugation. Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (U.S. Pat. No. 4,843,155, incorporated by reference in its entirety for all purposes).

[0066] In one embodiment, a sample comprises cells harvested from a bladder tissue sample, for example, a MIBC sample. Cells can be harvested from a biological sample using standard techniques known in the art. For example, in one embodiment, cells are harvested by centrifuging a cell sample and resuspending the pelleted cells. The cells can be resuspended in a buffered solution such as phosphate-buffered saline (PBS). After centrifuging the cell suspension to obtain a cell pellet, the cells can be lysed to extract nucleic acid, *e.g.*, messenger RNA. All samples obtained from a subject, including those subjected to any sort of further processing, are considered to be obtained from the subject.

[0067] The sample, in one embodiment, is further processed before the detection of the biomarker levels of the combination of biomarkers set forth herein. For example, mRNA in a cell or tissue sample can be separated from other components of the sample. The sample can be concentrated and/or purified to isolate mRNA in its non-natural state, as the mRNA is not in its natural environment. For example, studies have indicated that the higher order structure of mRNA *in vivo* differs from the *in vitro* structure of the same sequence (*see, e.g.*, Rouskin *et al.* (2014). *Nature* 505, pp. 701-705, incorporated herein in its entirety for all purposes).

[0068] mRNA from the sample in one embodiment, is hybridized to a synthetic DNA probe, which in some embodiments, includes a detection moiety (*e.g.*, detectable label, capture sequence, barcode reporting sequence). Accordingly, in these embodiments, a non-natural mRNA-cDNA complex is ultimately made and used for detection of the biomarker. In another embodiment, mRNA from the sample is directly labeled with a detectable label, *e.g.*, a fluorophore. In a further embodiment, the non-natural labeled-mRNA molecule is hybridized to a cDNA probe and the complex is detected.

[0069] In one embodiment, once the mRNA is obtained from a sample, it is converted to complementary DNA (cDNA) prior to the hybridization reaction or is used in a hybridization reaction together with one or more cDNA probes. cDNA does not exist *in vivo* and therefore is a non-natural molecule. Furthermore, cDNA-mRNA hybrids are synthetic and do not exist

in vivo. Besides cDNA not existing *in vivo*, cDNA is necessarily different than mRNA, as it includes deoxyribonucleic acid and not ribonucleic acid. The cDNA is then amplified, for example, by the polymerase chain reaction (PCR) or other amplification method known to those of ordinary skill in the art. For example, other amplification methods that may be employed include the ligase chain reaction (LCR) (Wu and Wallace, *Genomics*, 4:560 (1989), Landegren *et al.*, *Science*, 241:1077 (1988), incorporated by reference in its entirety for all purposes, transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173 (1989), incorporated by reference in its entirety for all purposes), self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87:1874 (1990), incorporated by reference in its entirety for all purposes), incorporated by reference in its entirety for all purposes, and nucleic acid based sequence amplification (NASBA). Guidelines for selecting primers for PCR amplification are known to those of ordinary skill in the art. *See, e.g.*, McPherson *et al.*, *PCR Basics: From Background to Bench*, Springer-Verlag, 2000, incorporated by reference in its entirety for all purposes. The product of this amplification reaction, *i.e.*, amplified cDNA is also necessarily a non-natural product. First, as mentioned above, cDNA is a non-natural molecule. Second, in the case of PCR, the amplification process serves to create hundreds of millions of cDNA copies for every individual cDNA molecule of starting material. The numbers of copies generated are far removed from the number of copies of mRNA that are present *in vivo*.

[0070] In one embodiment, cDNA is amplified with primers that introduce an additional DNA sequence (e.g., adapter, reporter, capture sequence or moiety, barcode) onto the fragments (e.g., with the use of adapter-specific primers), or mRNA or cDNA biomarker sequences are hybridized directly to a cDNA probe comprising the additional sequence (e.g., adapter, reporter, capture sequence or moiety, barcode). Amplification and/or hybridization of mRNA to a cDNA probe therefore serves to create non-natural double stranded molecules from the non-natural single stranded cDNA, or the mRNA, by introducing additional sequences and forming non-natural hybrids. Further, as known to those of ordinary skill in the art, amplification procedures have error rates associated with them. Therefore, amplification introduces further modifications into the cDNA molecules. In one embodiment, during amplification with the adapter-specific primers, a detectable label, *e.g.*, a fluorophore, is added to single strand cDNA molecules. Amplification therefore also serves to create DNA complexes that do not occur in nature, at least because (i) cDNA does not exist *in vivo*, (i) adapter sequences are added to the ends of cDNA molecules to make DNA sequences that do

not exist *in vivo*, (ii) the error rate associated with amplification further creates DNA sequences that do not exist *in vivo*, (iii) the disparate structure of the cDNA molecules as compared to what exists in nature, and (iv) the chemical addition of a detectable label to the cDNA molecules.

[0071] In some embodiments, the expression of a biomarker of interest is detected at the nucleic acid level via detection of non-natural cDNA molecules.

[0072] In some embodiments, the method for bladder cancer subtyping includes detecting expression levels of a classifier biomarker set in a sample obtained from a subject. The method can further comprise detecting expression levels of said classifier biomarker set in one or more control or reference samples. The one or more control or reference samples can be selected from a normal or bladder cancer-free sample, a bladder cancer luminal sample, a bladder cancer luminal infiltrated sample, a bladder cancer basal sample, a bladder cancer basal infiltrated/neuronal sample, or any combination thereof. In some embodiments, the detecting includes all of the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers as disclosed herein at the nucleic acid level or protein level. In some embodiments, the detecting includes all of the classifier biomarkers of **Table 1** or **Table 2** at the nucleic acid level or protein level. In another embodiment, a single or a subset or a plurality of the classifier biomarkers of **Table 1** are detected, for example, from about 1 to about 5, from about 5 to about 15, from about 15 to about 30, from about 30 to about 45, from about 45 to about 60 of the biomarkers in **Table 1** are detected in a method to determine the bladder cancer luminal subtype. In another embodiment, each of the biomarkers from **Table 1** is detected in a method to determine the bladder cancer subtype. In another embodiment, any of 60 of the biomarkers from **Table 1** are selected as the gene signatures for a specific bladder cancer subtype. In another embodiment, a single or a subset or a plurality of the classifier biomarkers of **Table 2** are detected, for example, from about 1 to about 5, from about 5 to about 15, from about 15 to about 30, from about 30 to about 45, from about 45 to about 56 of the biomarkers in **Table 2** are detected in a method to determine the bladder cancer luminal subtype. In another embodiment, each of the biomarkers from **Table 2** is detected in a method to determine the bladder cancer subtype. In another embodiment, any of 60 of the biomarkers from **Table 2** are selected as the gene signatures for a specific bladder cancer subtype. In some embodiments, the detecting includes all of the classifier biomarkers of **Table 2** at the nucleic acid level or protein level. In another embodiment, a single or a

subset or a plurality of the classifier biomarkers of **Table 2** are detected, for example, from about 5 to about 10, from about 10 to about 20, from about 20 to about 34, from about 30 to about 56, of the biomarkers in **Table 2** are detected in a method to determine the bladder cancer subtype. In another embodiment, each of the biomarkers from **Table 2** is detected in a method to determine the bladder cancer subtype. In another embodiment, any of 56 of the biomarkers from **Table 2** are selected as the gene signatures for a specific bladder cancer subtype. The detecting can be performed by any suitable technique including, but not limited to, RNA-seq, a reverse transcriptase polymerase chain reaction (RT-PCR), a microarray hybridization assay, or another hybridization assay, *e.g.*, a NanoString assay for example, with primers and/or probes specific to the classifier biomarkers, and/or the like. In some cases, the primers useful for the amplification methods (*e.g.*, RT-PCR or qRT-PCR) are any forward and reverse primers suitable for binding to a classifier biomarker provided herein, such as the classifier biomarkers of Table 1, Table 2 or any additional set of biomarker classifiers as disclosed herein.

[0073] The biomarkers described herein include RNA comprising the entire or partial sequence of any of the nucleic acid sequences of interest, or their non-natural cDNA product, obtained synthetically *in vitro* in a reverse transcription reaction. The term “fragment” is intended to refer to a portion of the polynucleotide that generally comprise at least 10, 15, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1,000, 1,200, or 1,500 contiguous nucleotides, or up to the number of nucleotides present in a full-length biomarker polynucleotide disclosed herein. A fragment of a biomarker polynucleotide will generally encode at least 15, 25, 30, 50, 100, 150, 200, or 250 contiguous amino acids, or up to the total number of amino acids present in a full-length biomarker protein as provided herein.

[0074] In some embodiments, overexpression, such as of an RNA transcript or its expression product, is determined by normalization to the level of reference RNA transcripts or their expression products, which can be all measured transcripts (or their products) in the sample or a particular reference set of RNA transcripts (or their non-natural cDNA products). Normalization is performed to correct for or normalize away both differences in the amount of RNA or cDNA assayed and variability in the quality of the RNA or cDNA used. Therefore, an assay typically measures and incorporates the expression of certain normalizing genes, including well known housekeeping genes, such as, for example, GAPDH and/or β -

Actin. Alternatively, normalization can be based on the mean or median signal of all of the assayed biomarkers or a large subset thereof (global normalization approach).

[0075] Isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays, NanoString Assays. One method for the detection of mRNA levels involves contacting the isolated mRNA or synthesized cDNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to the non-natural cDNA or mRNA biomarker.

[0076] As explained above, in one embodiment, once the mRNA is obtained from a sample, it is converted to complementary DNA (cDNA) in a hybridization reaction. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising sequence that is complementary to a portion of a specific mRNA. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising random sequence. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising sequence that is complementary to the poly(A) tail of an mRNA. cDNA does not exist *in vivo* and therefore is a non-natural molecule. In a further embodiment, the cDNA is then amplified, for example, by the polymerase chain reaction (PCR) or other amplification method known to those of ordinary skill in the art. PCR can be performed with the forward and/or reverse primers comprising sequence complementary to at least a portion of a classifier biomarker provided herein, such as the classifier biomarkers of Table 1, Table 2 or any additional set of biomarker classifiers as disclosed herein. The product of this amplification reaction, *i.e.*, amplified cDNA is necessarily a non-natural product. As mentioned above, cDNA is a non-natural molecule. Second, in the case of PCR, the amplification process serves to create hundreds of millions of cDNA copies for every individual cDNA molecule of starting material. The number of copies generated is far removed from the number of copies of mRNA that are present *in vivo*.

[0077] In one embodiment, cDNA is amplified with primers that introduce an additional DNA sequence (adapter sequence) onto the fragments (with the use of adapter-specific primers). The adaptor sequence can be a tail, wherein the tail sequence is not complementary to the cDNA. For example, the forward and/or reverse primers comprising sequence

complementary to at least a portion of a classifier biomarker provided herein, such as the classifier biomarkers of Table 1, Table 2 or any additional set of biomarker classifiers as disclosed herein can comprise tail sequence. Amplification therefore serves to create non-natural double stranded molecules from the non-natural single stranded cDNA, by introducing barcode, adapter and/or reporter sequences onto the already non-natural cDNA. In one embodiment, during amplification with the adapter-specific primers, a detectable label, *e.g.*, a fluorophore, is added to single strand cDNA molecules. Amplification therefore also serves to create DNA complexes that do not occur in nature, at least because (i) cDNA does not exist *in vivo*, (ii) adapter sequences are added to the ends of cDNA molecules to make DNA sequences that do not exist *in vivo*, (iii) the error rate associated with amplification further creates DNA sequences that do not exist *in vivo*, (iv) the disparate structure of the cDNA molecules as compared to what exists in nature, and (v) the chemical addition of a detectable label to the cDNA molecules.

[0078] In one embodiment, the synthesized cDNA (for example, amplified cDNA) is immobilized on a solid surface via hybridization with a probe, *e.g.*, via a microarray. In another embodiment, cDNA products are detected via real-time polymerase chain reaction (PCR) via the introduction of fluorescent probes that hybridize with the cDNA products. For example, in one embodiment, biomarker detection is assessed by quantitative fluorogenic RT-PCR (*e.g.*, with TaqMan® probes). For PCR analysis, well known methods are available in the art for the determination of primer sequences for use in the analysis.

[0079] Biomarkers provided herein in one embodiment, are detected via a hybridization reaction that employs a capture probe and/or a reporter probe. For example, the hybridization probe is a probe derivatized to a solid surface such as a bead, glass or silicon substrate. In another embodiment, the capture probe is present in solution and mixed with the patient's sample, followed by attachment of the hybridization product to a surface, *e.g.*, via a biotin-avidin interaction (*e.g.*, where biotin is a part of the capture probe and avidin is on the surface). The hybridization assay, in one embodiment, employs both a capture probe and a reporter probe. The reporter probe can hybridize to either the capture probe or the biomarker nucleic acid. Reporter probes *e.g.*, are then counted and detected to determine the level of biomarker(s) in the sample. The capture and/or reporter probe, in one embodiment contain a detectable label, and/or a group that allows functionalization to a surface.

[0080] For example, the nCounter gene analysis system (see, e.g., Geiss *et al.* (2008) Nat. Biotechnol. 26, pp. 317-325, incorporated by reference in its entirety for all purposes, is amenable for use with the methods provided herein.

[0081] Hybridization assays described in U.S. Patent Nos. 7,473,767 and 8,492,094, the disclosures of which are incorporated by reference in their entireties for all purposes, are amenable for use with the methods provided herein, i.e., to detect the biomarkers and biomarker combinations described herein.

[0082] Biomarker levels may be monitored using a membrane blot (such as used in hybridization analysis such as Northern, Southern, dot, and the like), or microwells, sample tubes, gels, beads, or fibers (or any solid support comprising bound nucleic acids). See, for example, U.S. Pat. Nos. 5,770,722, 5,874,219, 5,744,305, 5,677,195 and 5,445,934, each incorporated by reference in their entireties.

[0083] In one embodiment, microarrays are used to detect biomarker levels. Microarrays are particularly well suited for this purpose because of the reproducibility between different experiments. DNA microarrays provide one method for the simultaneous measurement of the expression levels of large numbers of genes. Each array consists of a reproducible pattern of capture probes attached to a solid support. Labeled RNA or DNA is hybridized to complementary probes on the array and then detected by laser scanning hybridization intensities for each probe on the array are determined and converted to a quantitative value representing relative gene expression levels. See, for example, U.S. Pat. Nos. 6,040,138, 5,800,992 and 6,020,135, 6,033,860, and 6,344,316, each incorporated by reference in their entireties. High-density oligonucleotide arrays are particularly useful for determining the gene expression profile for a large number of RNAs in a sample.

[0084] Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, for example, U.S. Pat. No. 5,384,261. Although a planar array surface is generally used, the array can be fabricated on a surface of virtually any shape or even a multiplicity of surfaces. Arrays can be nucleic acids (or peptides) on beads, gels, polymeric surfaces, fibers (such as fiber optics), glass, or any other appropriate substrate. See, for example, U.S. Pat. Nos. 5,770,358, 5,789,162, 5,708,153, 6,040,193 and 5,800,992, each incorporated by reference in their entireties. Arrays can be packaged in such a manner as to

allow for diagnostics or other manipulation of an all-inclusive device. See, for example, U.S. Pat. Nos. 5,856,174 and 5,922,591, each incorporated by reference in their entireties.

[0085] Serial analysis of gene expression (SAGE) in one embodiment is employed in the methods described herein. 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 (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*, Velculescu et al. Science 270:484-87, 1995; Cell 88:243-51, 1997, incorporated by reference in its entirety.

[0086] An additional method of biomarker level analysis at the nucleic acid level is the use of a sequencing method, for example, RNAseq, next generation sequencing, and massively parallel signature sequencing (MPSS), as described by Brenner et al. (Nat. Biotech. 18:630-34, 2000, incorporated by reference in its entirety). This is a sequencing approach that combines non-gel-based signature sequencing with *in vitro* cloning of millions of templates on separate 5 μ m diameter 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 (typically greater than 3.0×10^6 microbeads/cm²). 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 yeast cDNA library.

[0087] Another method of biomarker level analysis at the nucleic acid level is the use of an amplification method such as, for example, RT-PCR or quantitative RT-PCR (qRT-PCR). Methods for determining the level of biomarker mRNA in a sample may involve the process of nucleic acid amplification, e.g., by RT-PCR (the experimental embodiment set forth in Mullis, 1987, U.S. Pat. No. 4,683,202), ligase chain reaction (Barany (1991) Proc. Natl. Acad. Sci. USA 88:189-193), self-sustained sequence replication (Guatelli et al. (1990) Proc.

Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) Bio/Technology 6:1197), rolling circle replication (Lizardi et al., U.S. Pat. No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. Numerous different PCR or qRT-PCR protocols are known in the art and can be directly applied or adapted for use using the presently described compositions for the detection and/or quantification of expression of discriminative genes in a sample. See, for example, Fan et al. (2004) Genome Res. 14:878-885, herein incorporated by reference. Generally, in PCR, a target polynucleotide sequence is amplified by reaction with at least one oligonucleotide primer or pair of oligonucleotide primers. The primer(s) hybridize to a complementary region of the target nucleic acid and a DNA polymerase extends the primer(s) to amplify the target sequence. Under conditions sufficient to provide polymerase-based nucleic acid amplification products, a nucleic acid fragment of one size dominates the reaction products (the target polynucleotide sequence which is the amplification product). The amplification cycle is repeated to increase the concentration of the single target polynucleotide sequence. The reaction can be performed in any thermocycler commonly used for PCR.

[0088] Quantitative RT-PCR (qRT-PCR) (also referred as real-time RT-PCR) is preferred under some circumstances because it provides not only a quantitative measurement, but also reduced time and contamination. As used herein, "quantitative PCR" (or "real time qRT-PCR") refers to the direct monitoring of the progress of a PCR amplification as it is occurring without the need for repeated sampling of the reaction products. In quantitative PCR, the reaction products may be monitored via a signaling mechanism (e.g., fluorescence) as they are generated and are tracked after the signal rises above a background level but before the reaction reaches a plateau. The number of cycles required to achieve a detectable or "threshold" level of fluorescence varies directly with the concentration of amplifiable targets at the beginning of the PCR process, enabling a measure of signal intensity to provide a measure of the amount of target nucleic acid in a sample in real time. A DNA binding dye (e.g., SYBR green) or a labeled probe can be used to detect the extension product generated by PCR amplification. Any probe format utilizing a labeled probe comprising sequence complementary to the sequences as provided herein may be used.

[0089] Immunohistochemistry methods are also suitable for detecting the levels of the biomarkers. Samples can be frozen for later preparation or immediately placed in a fixative solution. Tissue samples can be fixed by treatment with a reagent, such as formalin, glutaraldehyde, methanol, or the like and embedded in paraffin. Methods for preparing slides for immunohistochemical analysis from formalin-fixed, paraffin-embedded tissue samples are well known in the art.

[0090] In one embodiment, the levels of the biomarkers provided herein, such as the classifier biomarkers of **Table 1** (or subsets thereof, for example 1 to 5, 5 to 15, 15 to 30, 30 to 45, 45 to 60 of the biomarkers biomarkers) or **Table 2** (or subsets thereof, for example 1 to 5, 5 to 15, 15 to 30, 30 to 45, 45 to 56 of the biomarkers biomarkers), are normalized against the expression levels of all RNA transcripts or their non-natural cDNA expression products, or protein products in the sample, or of a reference set of RNA transcripts or a reference set of their non-natural cDNA expression products, or a reference set of their protein products in the sample. In one embodiment, the levels of the biomarkers provided herein, such as any of the additional set of classifier biomarkers disclosed herein are normalized against the expression levels of all RNA transcripts or their non-natural cDNA expression products, or protein products in the sample, or of a reference set of RNA transcripts or a reference set of their non-natural cDNA expression products, or a reference set of their protein products in the sample. The reference set of RNA transcripts or the reference set of their non-natural cDNA expression products, or the reference set of their protein products in the sample can be one or more housekeeping genes or their protein products.

[0091] In one embodiment, bladder cancer subtypes can be evaluated using levels of protein expression of one or more of the classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers disclosed herein. The level of protein expression can be measured using an immunological detection method. Immunological detection methods which can be used herein include, but are not limited to, competitive and non-competitive assay systems using techniques such as Western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, and the like. Such assays are

routine and well known in the art (see, e.g., Ausubel *et al*, eds, 1994, *Current Protocols in Molecular Biology*, Vol. I, John Wiley & Sons, Inc., New York, which is incorporated by reference herein in its entirety).

[0092] In one embodiment, antibodies specific for biomarker proteins are utilized to detect the expression of a biomarker protein in a body sample. The method comprises obtaining a body sample from a patient or a subject, contacting the body sample with at least one antibody directed to a biomarker that is selectively expressed in bladder cancer cells, and detecting antibody binding to determine if the biomarker is expressed in the patient sample. In one embodiment, provided herein is an immunocytochemistry technique for diagnosing bladder cancer subtypes. One of skill in the art will recognize that the immunocytochemistry method described herein may be performed manually or in an automated fashion.

[0093] As provided throughout, the methods set forth herein provide a method for determining the bladder cancer subtype of a patient. Once the biomarker levels are determined, for example by measuring non-natural cDNA biomarker levels or non-natural mRNA-cDNA biomarker complexes, the biomarker levels are compared to reference values or a reference sample as provided herein, for example with the use of statistical methods or direct comparison of detected levels, to make a determination of the bladder cancer molecular subtype. Based on the comparison, the patient's bladder cancer sample is classified, e.g., as luminal, luminal infiltrated, basal, or basal infiltrated/neuronal. As provided herein, the statistical methods utilized in the methods provided herein can be centroid based statistical methods as described herein and/or known in the art.

[0094] In one embodiment, expression level values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** or **Table 2** are compared to reference expression level value(s) from at least one sample training set, wherein the at least one sample training set comprises expression level values from a reference sample(s). In a further embodiment, the at least one sample training set comprises expression level values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers disclosed herein from a bladder cancer luminal subtype, bladder cancer luminal infiltrated subtype, bladder cancer basal subtype, bladder cancer basal infiltrated/neuronal subtype, or bladder cancer-free sample or a combination thereof.

[0095] In a separate embodiment, hybridization values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers disclosed herein are compared to reference hybridization value(s) from at least one sample training set, wherein the at least one sample training set comprises hybridization values from a reference sample(s). In a further embodiment, the at least one sample training set comprises hybridization values of the at least one classifier biomarker provided herein, such as the classifier biomarkers of **Table 1**, **Table 2** or any additional set of biomarker classifiers disclosed herein from a bladder cancer luminal subtype, bladder cancer luminal infiltrated subtype, bladder cancer basal subtype, bladder cancer basal infiltrated/neuronal subtype, or bladder cancer-free sample, or a combination thereof. Methods for comparing detected levels of biomarkers to reference values and/or reference samples are provided herein. Based on this comparison, in one embodiment a correlation or distance measure (e.g., Euclidean distance) between the biomarker levels obtained from the subject's sample and the reference values is obtained. An assessment of the bladder cancer subtype is then made.

[0096] Various statistical methods can be used to aid in the comparison of the biomarker levels obtained from the patient and reference biomarker levels, for example, from at least one sample training set.

[0097] In one embodiment, a supervised pattern recognition method is employed. Examples of supervised pattern recognition methods can include, but are not limited to, the nearest centroid methods (Dabney (2005) *Bioinformatics* 21(22):4148-4154 and Tibshirani et al. (2002) *Proc. Natl. Acad. Sci. USA* 99(10):6576-6572); soft independent modeling of class analysis (SIMCA) (see, for example, Wold, 1976); partial least squares analysis (PLS) (see, for example, Wold, 1966; Joreskog, 1982; Frank, 1984; Bro, R., 1997); linear discriminant analysis (LDA) (see, for example, Nillson, 1965); K-nearest neighbor analysis (KNN) (see, for example, Brown et al., 1996); artificial neural networks (ANN) (see, for example, Wasserman, 1989; Anker et al., 1992; Hare, 1994); probabilistic neural networks (PNNs) (see, for example, Parzen, 1962; Bishop, 1995; Speckt, 1990; Broomhead et al., 1988; Patterson, 1996); rule induction (RI) (see, for example, Quinlan, 1986); and, Bayesian methods (see, for example, Bretthorst, 1990a, 1990b, 1988). In one embodiment, the classifier for identifying tumor subtypes based on gene expression data is used in a centroid based method as described in Mullins et al. (2007) *Clin Chem.* 53(7):1273-9, which is

incorporated herein by reference in its entirety. In another embodiment, the classifier for identifying tumor subtypes based on gene expression data is used in a nearest centroid based method as described in Dabney (2005) *Bioinformatics* 21(22):4148-4154, which is incorporated herein by reference in its entirety. The nearest centroid based method can be performed using CLaNC software as described in Dabney AR. ClaNC: Point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics*. 2006;22: 122–123 or equivalents or derivatives thereof.

[0098] In other embodiments, an unsupervised training approach is employed, and therefore, no training set is used.

[0099] Referring to sample training sets for supervised learning approaches again, in some embodiments, a sample training set(s) can include expression data of a plurality or all of the classifier biomarkers (e.g., all the classifier biomarkers of **Table 1** or **Table 2**, or any additional set of biomarker classifiers disclosed herein) from a bladder cancer sample. The plurality of classifier biomarkers can comprise at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of **Table 1**. In some embodiments, the plurality of classifier biomarkers comprises all 60 biomarkers of **Table 1**. The plurality of classifier biomarkers can comprise at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of **Table 2**. In some embodiments, the plurality of classifier biomarkers comprises all 56 biomarkers of **Table 2**. In some embodiments, the sample training set(s) are normalized to remove sample-to-sample variation.

[00100] In some embodiments, comparing can include applying a statistical algorithm, such as, for example, any suitable multivariate statistical analysis model, which can be parametric or non-parametric. In some embodiments, applying the statistical algorithm can include determining a correlation between the expression data obtained from the human bladder tissue sample and the expression data from the bladder cancer training set(s). In some embodiments, cross-validation is performed, such as (for example), leave-one-out cross-validation (LOOCV). In some embodiments, integrative correlation is performed. In some embodiments, a Spearman correlation is performed. In some embodiments, a centroid

based method based on gene expression data is employed for the statistical algorithm. The centroids can be constructed using any method known in the art for generating centroids such as, for example, those found in Mullins et al. (2007) Clin Chem. 53(7):1273-9 or the nearest centroid method found in Dabney (2005) Bioinformatics 21(22):4148-4154, which is herein incorporated by reference in its entirety. In one embodiment, a correlation analysis is performed on the expression data obtained from the human bladder tissue sample and the centroid(s) constructed on the expression data from the bladder cancer training set(s). The correlation analysis can be a Spearman correlation or a Pearson correlation. In one embodiment, a distance measure analysis (e.g., Euclidean distance) is performed on the expression data obtained from the human bladder tissue sample and the centroid(s) constructed on the expression data from the bladder cancer training set(s).

[00101] Results of the gene expression analysis performed on a sample (e.g., bladder sample as provided herein) from a subject (test sample) may be compared to a biological sample(s) or data derived (e.g., expression data or levels from at least one classifier biomarker provided herein, e.g., Table 1 or 2) from a biological sample(s) that is known or suspected to be normal (“reference sample” or “normal sample”, e.g., non-bladder cancer sample). In some embodiments, a reference sample or reference gene expression data (e.g., for at least one classifier biomarkers provided herein such as, for example, Table 1 or Table 2) is obtained or derived from an individual known to have a particular molecular subtype of bladder cancer, *i.e.*, luminal, luminal infiltrated, basal or basal infiltrated/neuronal. In one embodiment, the gene expression levels or profile for the at least one classifier biomarker provided herein (e.g., Table 1 or 2) measured or detected in the test sample may be compared to centroids constructed from the gene expression performed on the reference or normal sample. The centroids can be constructed using any of the methods provided herein such as, for example, using the ClaNC software described in Dabney AR. ClaNC: Point-and-click software for classifying microarrays to nearest centroids. Bioinformatics. 2006;22: 122–123 or equivalents or derivatives related thereto. Classification or determination of the subtype of the test sample can then be ascertained by determining the nearest centroid from the reference or normal sample to which the expression levels or profile from said test sample is nearest based on a distance measure or correlation. The distance measure can be a Euclidean distance.

[00102] The reference sample may be assayed at the same time, or at a different time from the test sample (e.g., bladder tissue sample obtained from a subject suffering from or suspected of suffering from bladder cancer). Alternatively, the biomarker level information from a reference sample may be stored in a database or other means for access at a later date.

[00103] The biomarker level results of an assay on the test sample may be compared to the results of the same assay on a reference sample. In some cases, the results of the assay on the reference sample are from a database, or a reference value(s). In some cases, the results of the assay on the reference sample are a known or generally accepted value or range of values by those skilled in the art. In some cases, the comparison is qualitative. In other cases, the comparison is quantitative. In some cases, qualitative or quantitative comparisons may involve but are not limited to one or more of the following: comparing expression levels of a test sample to gene centroids constructed from expression level data from a reference sample (e.g., constructed from expression level data for one or a plurality of genes from Table 1 or Table 2), fluorescence values, spot intensities, absorbance values, chemiluminescent signals, histograms, critical threshold values, statistical significance values, expression levels of the genes described herein, mRNA copy numbers.

[00104] In one embodiment, an odds ratio (OR) is calculated for each biomarker level panel measurement. Here, the OR is a measure of association between the measured biomarker values for the patient and an outcome, e.g., bladder cancer subtype. For example, see, *J. Can. Acad. Child Adolesc. Psychiatry* 2010; 19(3): 227-229, which is incorporated by reference in its entirety for all purposes.

[00105] In one embodiment, a specified statistical confidence level may be determined in order to provide a confidence level regarding the bladder cancer subtype. For example, it may be determined that a confidence level of greater than 90% may be a useful predictor of the bladder cancer subtype. In other embodiments, more or less stringent confidence levels may be chosen. For example, a confidence level of about or at least about 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, 99.5%, or 99.9% may be chosen. The confidence level provided may in some cases be related to the quality of the sample, the quality of the data, the quality of the analysis, the specific methods used, and/or the number of gene expression values (*i.e.*, the number of genes) analyzed. The specified confidence level for providing the likelihood of response may be chosen on the basis of the expected number of false positives or false negatives. Methods for choosing parameters for achieving a specified

confidence level or for identifying markers with diagnostic power include but are not limited to Receiver Operating Characteristic (ROC) curve analysis, binormal ROC, principal component analysis, odds ratio analysis, partial least squares analysis, singular value decomposition, least absolute shrinkage and selection operator analysis, least angle regression, and the threshold gradient directed regularization method.

[00106] Determining the bladder cancer subtype in some cases can be improved through the application of algorithms designed to normalize and or improve the reliability of the gene expression data. In some embodiments, the data analysis utilizes a computer or other device, machine or apparatus for application of the various algorithms described herein due to the large number of individual data points that are processed. A “machine learning algorithm” refers to a computational-based prediction methodology, also known to persons skilled in the art as a “classifier,” employed for characterizing a gene expression profile or profiles, *e.g.*, to determine the bladder cancer subtype. The biomarker levels, determined by, *e.g.*, microarray-based hybridization assays, sequencing assays, NanoString assays, etc., are in one embodiment subjected to the algorithm in order to classify the profile. Supervised learning generally involves “training” a classifier to recognize the distinctions among subtypes such as luminal positive, luminal infiltrated positive, basal positive or basal infiltrated/neuronal positive, and then “testing” the accuracy of the classifier on an independent test set. Therefore, for new, unknown samples the classifier can be used to predict, for example, the class (*e.g.*, luminal vs. luminal infiltrated vs. basal vs. basal infiltrated/neuronal) in which the samples belong. The machine learning algorithm can be a CLaNC algorithm as provided herein.

[00107] In some embodiments, a robust multi-array average (RMA) method may be used to normalize raw data. The RMA method begins by computing background-corrected intensities for each matched cell on a number of microarrays. In one embodiment, the background corrected values are restricted to positive values as described by Irizarry *et al.* (2003). *Biostatistics* April 4 (2): 249-64, incorporated by reference in its entirety for all purposes. After background correction, the base-2 logarithm of each background corrected matched-cell intensity is then obtained. The background corrected, log-transformed, matched intensity on each microarray is then normalized using the quantile normalization method in which for each input array and each probe value, the array percentile probe value is replaced with the average of all array percentile points, this method is more completely described by

Bolstad *et al.* Bioinformatics 2003, incorporated by reference in its entirety. Following quantile normalization, the normalized data may then be fit to a linear model to obtain an intensity measure for each probe on each microarray. Tukey's median polish algorithm (Tukey, J. W., Exploratory Data Analysis. 1977, incorporated by reference in its entirety for all purposes) may then be used to determine the log-scale intensity level for the normalized probe set data.

[00108] Various other software programs may be implemented. In certain methods, feature selection and model estimation may be performed by logistic regression with *lasso* penalty using *glmnet* (Friedman *et al.* (2010). *Journal of statistical software* **33**(1): 1-22, incorporated by reference in its entirety). Raw reads may be aligned using TopHat (Trapnell *et al.* (2009). *Bioinformatics* **25**(9): 1105-11, incorporated by reference in its entirety). In methods, top features (N ranging from 10 to 200) are used to train a linear support vector machine (SVM) (Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* 1999; **9**(3): 293-300, incorporated by reference in its entirety) using the *e1071* library (Meyer D. Support vector machines: the interface to libsvm in package e1071. 2014, incorporated by reference in its entirety). Confidence intervals, in one embodiment, are computed using the pROC package (Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* 2011; **12**: 77, incorporated by reference in its entirety).

[00109] In addition, data may be filtered to remove data that may be considered suspect. In one embodiment, data derived from microarray probes that have fewer than about 4, 5, 6, 7 or 8 guanosine + cytosine nucleotides may be considered to be unreliable due to their aberrant hybridization propensity or secondary structure issues. Similarly, data deriving from microarray probes that have more than about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 guanosine + cytosine nucleotides may in one embodiment be considered unreliable due to their aberrant hybridization propensity or secondary structure issues.

[00110] In some embodiments, data from probe-sets may be excluded from analysis if they are not identified at a detectable level (above background).

[00111] In some embodiments, probe-sets that exhibit no, or low variance may be excluded from further analysis. Low-variance probe-sets are excluded from the analysis via a Chi-Square test. In one embodiment, a probe-set is considered to be low-variance if its

transformed variance is to the left of the 99 percent confidence interval of the Chi-Squared distribution with $(N-1)$ degrees of freedom. $(N-1) \times \text{Probe-set Variance} / (\text{Gene Probe-set Variance})$. $\text{Chi-Sq}(N-1)$ where N is the number of input CEL files, $(N-1)$ is the degrees of freedom for the Chi-Squared distribution, and the “probe-set variance for the gene” is the average of probe-set variances across the gene. In some embodiments, probe-sets for a given mRNA or group of mRNAs may be excluded from further analysis if they contain less than a minimum number of probes that pass through the previously described filter steps for GC content, reliability, variance and the like. For example, in some embodiments, probe-sets for a given gene or transcript cluster may be excluded from further analysis if they contain less than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or less than about 20 probes.

[00112] Methods of biomarker level data analysis in one embodiment, further include the use of a feature selection algorithm as provided herein. In some embodiments, feature selection is provided by use of the LIMMA software package (Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420, incorporated by reference in its entirety for all purposes).

[00113] Methods of biomarker level data analysis, in one embodiment, include the use of a pre-classifier algorithm. For example, an algorithm may use a specific molecular fingerprint to pre-classify the samples according to their composition and then apply a correction/normalization factor. This data/information may then be fed in to a final classification algorithm that would incorporate that information to aid in the final diagnosis.

[00114] Methods of biomarker level data analysis, in one embodiment, further include the use of a classifier algorithm as provided herein. In one embodiment, a diagonal linear discriminant analysis, k-nearest neighbor algorithm, support vector machine (SVM) algorithm, linear support vector machine, random forest algorithm, or a probabilistic model-based method or a combination thereof is provided for classification of microarray data. In some embodiments, identified markers that distinguish samples (*e.g.*, of varying biomarker level profiles, and/or varying molecular subtypes of bladder cancer (*e.g.*, luminal, luminal infiltrated, basal, basal infiltrated/neuronal)) are selected based on statistical significance of the difference in biomarker levels between classes of interest. In some cases, the statistical

significance is adjusted by applying a Benjamin Hochberg or another correction for false discovery rate (FDR).

[00115] In some cases, the classifier algorithm may be supplemented with a meta-analysis approach such as that described by Fishel and Kaufman et al. 2007 *Bioinformatics* 23(13): 1599-606, incorporated by reference in its entirety for all purposes. In some cases, the classifier algorithm may be supplemented with a meta-analysis approach such as a repeatability analysis.

[00116] Methods for deriving and applying posterior probabilities to the analysis of biomarker level data are known in the art and have been described for example in Smyth, G. K. 2004 *Stat. Appi. Genet. Mol. Biol.* 3: Article 3, incorporated by reference in its entirety for all purposes. In some cases, the posterior probabilities may be used in the methods provided herein to rank the markers provided by the classifier algorithm.

[00117] A statistical evaluation of the results of the biomarker level profiling may provide a quantitative value or values indicative of one or more of the following: molecular subtype of bladder cancer (*e.g.*, luminal, luminal infiltrated, basal, basal infiltrated/neuronal); the likelihood of the success of a particular therapeutic intervention, *e.g.*, angiogenesis inhibitor therapy, chemotherapy, or immunotherapy. In one embodiment, the data is presented directly to the physician in its most useful form to guide patient care, or is used to define patient populations in clinical trials or a patient population for a given medication. The results of the molecular profiling can be statistically evaluated using a number of methods known to the art including, but not limited to: the students T test, the two sided T test, Pearson rank sum analysis, hidden Markov model analysis, analysis of q-q plots, principal component analysis, one way ANOVA, two way ANOVA, LIMMA and the like.

[00118] In some cases, accuracy may be determined by tracking the subject over time to determine the accuracy of the original diagnosis. In other cases, accuracy may be established in a deterministic manner or using statistical methods. For example, receiver operator characteristic (ROC) analysis may be used to determine the optimal assay parameters to achieve a specific level of accuracy, specificity, positive predictive value, negative predictive value, and/or false discovery rate.

[00119] In some cases, the results of the biomarker level profiling assays, are entered into a database for access by representatives or agents of a molecular profiling business, the

individual, a medical provider, or insurance provider. In some cases, assay results include sample classification, identification, or diagnosis by a representative, agent or consultant of the business, such as a medical professional. In other cases, a computer or algorithmic analysis of the data is provided automatically. In some cases, the molecular profiling business may bill the individual, insurance provider, medical provider, researcher, or government entity for one or more of the following: molecular profiling assays performed, consulting services, data analysis, reporting of results, or database access.

[00120] In some embodiments, the results of the biomarker level profiling assays are presented as a report on a computer screen or as a paper record. In some embodiments, the report may include, but is not limited to, such information as one or more of the following: the levels of biomarkers (*e.g.*, as reported by copy number or fluorescence intensity, etc.) as compared to the reference sample or reference value(s); the likelihood the subject will respond to a particular therapy, based on the biomarker level values and the bladder cancer subtype and proposed therapies.

[00121] In one embodiment, the results of the gene expression profiling may be classified into one or more of the following: luminal positive, luminal infiltrated positive, basal positive or basal infiltrated/neuronal positive, luminal negative, luminal infiltrated negative, basal negative or basal infiltrated/neuronal negative or a combination thereof.

[00122] In some embodiments, results are classified using a trained algorithm. Trained algorithms include algorithms that have been developed using a reference set of known gene expression values and/or normal samples, for example, samples from individuals diagnosed with a particular molecular subtype of bladder cancer. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype of bladder cancer. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype of bladder cancer, and are also known to possess certain immune cell signature. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype of bladder cancer, and are also known to have certain expression of tumor driver genes.

[00123] Algorithms suitable for categorization of samples include but are not limited to k-nearest neighbor algorithms, support vector machines, linear discriminant analysis, centroid

algorithms (e.g., CLaNC), diagonal linear discriminant analysis, updown, naive Bayesian algorithms, neural network algorithms, hidden Markov model algorithms, genetic algorithms, or any combination thereof.

[00124] When a binary classifier is compared with actual true values (e.g., values from a biological sample), there are typically four possible outcomes. If the outcome from a prediction is p (where “p” is a positive classifier output, such as the presence of a deletion or duplication syndrome) and the actual value is also p, then it is called a true positive (TP); however if the actual value is n then it is said to be a false positive (FP). Conversely, a true negative has occurred when both the prediction outcome and the actual value are n (where “n” is a negative classifier output, such as no deletion or duplication syndrome), and false negative is when the prediction outcome is n while the actual value is p. In one embodiment, consider a test that seeks to determine whether a person is likely or unlikely to respond to angiogenesis inhibitor therapy. A false positive in this case occurs when the person tests positive, but actually does respond. A false negative, on the other hand, occurs when the person tests negative, suggesting they are unlikely to respond, when they actually are likely to respond. The same holds true for classifying a bladder cancer subtype.

[00125] The positive predictive value (PPV), or precision rate, or post-test probability of disease, is the proportion of subjects with positive test results who are correctly diagnosed as likely or unlikely to respond, or diagnosed with the correct bladder cancer subtype, or a combination thereof. It reflects the probability that a positive test reflects the underlying condition being tested for. Its value does however depend on the prevalence of the disease, which may vary. In one example the following characteristics are provided: FP (false positive); TN (true negative); TP (true positive); FN (false negative). False positive rate (α)=FP/(FP+TN)-specificity; False negative rate (β)=FN/(TP+FN)-sensitivity; Power=sensitivity = 1- β ; Likelihood-ratio positive=sensitivity/(1-specificity); Likelihood-ratio negative=(1-sensitivity)/specificity. The negative predictive value (NPV) is the proportion of subjects with negative test results who are correctly diagnosed.

[00126] In some embodiments, the results of the biomarker level analysis of the subject methods provide a statistical confidence level that a given diagnosis is correct. In some embodiments, such statistical confidence level is at least about, or more than about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 99.5%, or more.

[00127] In some embodiments, the method further includes classifying the bladder tissue sample as a particular bladder cancer subtype based on the comparison of biomarker levels in the sample and reference biomarker levels, for example present in at least one training set. In some embodiments, the bladder tissue sample is classified as a particular subtype if the results of the comparison meet one or more criterion such as, for example, a minimum percent agreement, a value of a statistic calculated based on the percentage agreement such as (for example) a kappa statistic, a minimum correlation (*e.g.*, Pearson's correlation) and/or the like.

[00128] It is intended that the methods described herein can be performed by software (stored in memory and/or executed on hardware), hardware, or a combination thereof. Hardware modules may include, for example, a general-purpose processor, a field programmable gate array (FPGA), and/or an application specific integrated circuit (ASIC). Software modules (executed on hardware) can be expressed in a variety of software languages (*e.g.*, computer code), including Unix utilities, C, C++, Java™, Ruby, SQL, SAS®, the R programming language/software environment, Visual Basic™, and other object-oriented, procedural, or other programming language and development tools. Examples of computer code include, but are not limited to, micro-code or micro-instructions, machine instructions, such as produced by a compiler, code used to produce a web service, and files containing higher-level instructions that are executed by a computer using an interpreter. Additional examples of computer code include, but are not limited to, control signals, encrypted code, and compressed code.

[00129] Some embodiments described herein relate to devices with a non-transitory computer-readable medium (also can be referred to as a non-transitory processor-readable medium or memory) having instructions or computer code thereon for performing various computer-implemented operations and/or methods disclosed herein. The computer-readable medium (or processor-readable medium) is non-transitory in the sense that it does not include transitory propagating signals per se (*e.g.*, a propagating electromagnetic wave carrying information on a transmission medium such as space or a cable). The media and computer code (also can be referred to as code) may be those designed and constructed for the specific purpose or purposes. Examples of non-transitory computer-readable media include, but are not limited to: magnetic storage media such as hard disks, floppy disks, and magnetic tape; optical storage media such as Compact Disc/Digital Video Discs (CD/DVDs), Compact Disc-

Read Only Memories (CD-ROMs), and holographic devices; magneto-optical storage media such as optical disks; carrier wave signal processing modules; and hardware devices that are specially configured to store and execute program code, such as Application-Specific Integrated Circuits (ASICs), Programmable Logic Devices (PLDs), Read-Only Memory (ROM) and Random-Access Memory (RAM) devices. Other embodiments described herein relate to a computer program product, which can include, for example, the instructions and/or computer code discussed herein.

[00130] In some embodiments, a single biomarker, or from about 1 to about 5, from about 5 to about 15, from about 15 to about 30, from about 30 to about 45, from about 45 to about 56, from about 45 to about 60 (e.g., as disclosed in **Table 1** or **Table 2**) is capable of classifying subtypes of bladder cancer with a predictive success of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein (e.g., in **Table 1**, **Table 2**) can be used to obtain a predictive success of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between.

[00131] In some embodiments, a single biomarker, or any of the biomarkers as herein disclosed as the additional set of biomarker classifiers is capable of classifying subtypes of bladder cancer with a predictive success of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%,

at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein (e.g., as herein disclosed as the additional set of biomarker classifiers) can be used to obtain a predictive success of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between.

[00132] In some embodiments, a single biomarker, or from about 1 to about 5, from about 5 to about 15, from about 15 to about 30, from about 30 to about 45, from about 45 to about 56, from about 45 to about 60 biomarkers (e.g., as disclosed in **Table 1** or **Table 2**) is capable of classifying subtypes of bladder cancer with a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein can be used to obtain a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at

least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between.

[00133] In some embodiments, a single biomarker, as herein disclosed such as the additional set of biomarker classifiers is capable of classifying subtypes of bladder cancer with a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein can be used to obtain a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, at least about 73%, at least about 74%, at least about 75%, at least about 76%, at least about 77%, at least about 78%, at least about 79%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, up to 100%, and all values in between.

Classifier Biomarker Selection

[00134] In one embodiment, the methods and compositions provided herein are useful for determining the bladder cancer subtype of a sample (e.g., bladder tissue sample) from a patient by analyzing the expression of a set of biomarkers, whereby the set of biomarkers comprise a fewer number of biomarkers than methods known in the art for molecularly classifying bladder cancer subtype. In some cases, the set of biomarkers is less than 250, 240, 230, 220, 210, 200, 150, 100, 90, 80, 70, or 65 biomarkers. In some cases, the set of biomarkers is between 10 and 60 biomarkers. In some cases, the set of biomarkers is the set of 60 biomarkers listed in **Table 1**. In some cases, the set of biomarkers is a sub-set of biomarkers listed **Table 1** such as, for example, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 56, 57, 58, or 59 of the biomarkers listed in **Table 1**. In some cases, the set of biomarkers is

between 10 and 56 biomarkers. In some cases, the set of biomarkers is the set of 56 biomarkers listed in **Table 2**. In some cases, the set of biomarkers is a sub-set of biomarkers listed **Table 2** such as, for example, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 51, 52, 53, 54, 55, or 56 of the biomarkers listed in **Table 2**. The biomarkers or classifier biomarkers useful in the methods and compositions provided herein can be selected from one or more bladder cancer datasets from one or more databases. The databases can be public databases. In one embodiment, classifier biomarkers (e.g., one or more genes listed in **Table 1** and **Table 2**) useful in the methods and compositions provided herein for detecting or diagnosing bladder cancer subtypes were selected from a bladder cancer dataset from The Cancer Genome Atlas (TCGA). See Robertson, AG, et al., Cell, 171(3): 540-556 (2017); Cancer Genome Atlas Research Network, Nature, 507(492): 315-22 (2014), both of which are incorporated by reference in their entireties. In one embodiment, classifier biomarkers useful for the methods and compositions provided herein such as those in **Table 1** (or **Table 2**) are selected by subjecting a large set of classifier biomarkers to an *in silico* based process in order to determine the minimum number of genes whose expression profile can be used to determine an bladder cancer subtype of sample obtained from a subject. In some cases, the large set of classifier biomarkers can be a bladder cancer dataset such as, for example, from TCGA. In some cases, the large set of classifier biomarkers can be 2708-gene classifier described herein, whereby the 2708-gene classifier can serve to define gold standard (GS) subtype. The *in silico* process for selecting a gene signature as provided herein for determining bladder cancer subtype of a sample from a patient can comprise, applying or using a Classification to Nearest Centroid (CLaNC) algorithm with modification on the standard 2708 classifier biomarkers to choose an equal number of negatively and positively correlated genes for each subtype. For determination of the optimal number of genes (e.g., 60 per subtype as shown in **Table 1**, or 56 per subtype as shown in **Table 2**) to include in the signature, the process can further comprise performing a 5-fold cross validation using TCGA bladder cancer dataset as provided herein to produce cross-validation curves as shown in **FIG. 2**. To get the final list of gene classifiers, the method can further comprise applying the Classification to Nearest Centroid (CLaNC) to the entire TCGA data set minus 20% of samples with the lowest gold standard subtype prediction strength (for example, as shown in **FIG. 3**), and removing an equal number from each subtype.

[00135] In one embodiment, the method further comprises validating the gene classifiers. Validation can comprise testing the expression of the classifiers in several fresh frozen

publicly available array and RNAseq datasets and calling the subtype based on said expression levels and subsequently comparing the expression with the gold standard subtype calls as defined by the previously published 2708-gene signature. In other words, validation can comprise calling the subtypes of the several fresh frozen publicly available array and RNAseq test datasets using their expression levels and the CLaNC algorithm as described herein and comparing the subtype calls with the gold standard subtype calls as defined by the previously published 2708-gene signature. Final validation of the gene signature (e.g., **Table 1** or **Table 2**) can then be performed in a newly collected dataset of archived formalin-fixed paraffin-embedded (FFPE) bladder cancer samples to assure comparable performance in the FFPE samples. In one embodiment, the classifier biomarkers of **Table 1** were selected based on the *in silico* CLaNC process described herein. The gene symbols and official gene names are listed in **Table 1**. In one embodiment, the classifier biomarkers of **Table 2** were selected based on the *in silico* CLaNC process described herein. The gene symbols and official gene names are listed in **Table 2**. Further to the above embodiments, the *in silico* CLaNC process can entail use of the CLaNC process described in Dabney (2005) *Bioinformatics* 21(22):4148-4154. In one embodiment, the *in silico* CLaNC process can entail use of CLaNC software described in Dabney AR. CLaNC: Point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics*. 2006;22: 122–123 or equivalents or derivatives related thereto.

[00136] In one embodiment, the methods as provided herein require the detection of at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 18, at least 20, at least 22, at least 24, at least 26, at least 28, at least 30, at least 32, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, at least 50, at least 51, at least 52, at least 53, at least 54, at least 55, at least 56, at least 57, at least 58, at least 59, up to 60 classifier biomarkers in a bladder cancer cell sample obtained from a patient which expression is altered in order to identify a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal bladder cancer subtype. The same applies for other classifier biomarker expression datasets as provided herein.

[00137] In another embodiment, the methods as provided herein require the detection of a total of at least 1, at least 2, at least 5, at least 8, at least 10, at least 18, at least 36, at least 54,

at least 55, at least 56, at least 57, at least 58, at least 59, or up to 60 classifier biomarkers out of the 60 gene biomarkers of **Table 1** (or **Table 2**) in a bladder cancer cell sample obtained from a patient in order to identify a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal bladder cancer subtype. In another embodiment, the methods as provided herein require the detection of a total of at least 1, at least 2, at least 5, at least 10, at least 20, at least 40, at least 50 or up to 60 classifier biomarkers out of the 60 gene biomarkers of **Table 1** (or **Table 2**) in a bladder cancer cell sample obtained from a patient in order to identify a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal bladder cancer subtype. The same applies for other classifier biomarker expression datasets as provided herein.

[00138] In one embodiment, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 10, at least 18, at least 36, at least 54, at least 55, at least 56, at least 57, at least 58, at least 59, or up to 60 classifier biomarkers out of the 60 gene biomarkers of **Table 1** (or **Table 2**) are “up-regulated” in a specific subtype of bladder cancer. In another embodiment, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 10, at least 18, at least 36, at least 54, at least 55, at least 56, at least 57, at least 58, at least 59, or up to 60 classifier biomarkers out of the 60 gene biomarkers of **Table 1** (or **Table 2**) are “down-regulated” in a specific subtype of bladder cancer. The same applies for other classifier biomarker expression datasets as provided herein.

[00139] In one embodiment, the expression level of an “up-regulated” biomarker as provided herein is increased by about 0.5-fold, about 1-fold, about 1.5-fold, about 2-fold, about 2.5-fold, about 3-fold, about 3.5-fold, about 4-fold, about 4.5-fold, about 5-fold, and any values in between. In another embodiment, the expression level of a “down-regulated” biomarker as provided herein is decreased by about 0.8-fold, about 1.4-fold, about 2-fold, about 2.6-fold, about 3.2-fold, about 3.6-fold, about 4-fold, and any values in between.

[00140] It is recognized that additional genes or proteins can be used in the practice of methods, compositions and kits provided herein. In general, genes useful in classifying the subtypes of bladder cancer, include those that are independently capable of distinguishing between normal versus tumor, or between different classes or grades of bladder cancer. A gene is considered to be capable of reliably distinguishing between subtypes if the area under the receiver operator characteristic (ROC) curve is approximately 1.

Clinical / Therapeutic Uses

[00141] In one embodiment, a method is provided herein for determining a disease outcome or prognosis for a patient suffering from bladder cancer. In some cases, the bladder cancer is Muscle Invasive Bladder Cancer (MIBC). The disease outcome or prognosis can be measured by examining the overall survival for a period of time or intervals (e.g., 0 to 36 months or 0 to 60 months). In one embodiment, survival is analyzed as a function of bladder cancer subtype (e.g., luminal, luminal infiltrated, basal, basal infiltrated/neuronal). In one embodiment, survival is analyzed as a function of MIBC subtype (e.g., luminal, luminal infiltrated, basal, basal infiltrated/neuronal). The bladder cancer subtype (e.g., MIBC subtype) can be determined using the methods provided herein such as, for example, determining the expression of all or subsets of the genes in **Table 1** or **Table 2**. Relapse-free and overall survival can be assessed using standard Kaplan-Meier plots as well as Cox proportional hazards modeling.

[00142] In one embodiment, the methods and compositions as provided herein for determining a bladder cancer (e.g., MIBC) subtype of a patient suffering or suspected of suffering from bladder cancer is used to determine whether or not said patient is a candidate for treatment with a specific type or types of cancer therapy. The sample can be any type of sample obtained from the patient as provided herein. In one embodiment, determining the bladder cancer (e.g., MIBC) subtype is one of a number of methods that can be employed to characterize the sample obtained from the patient such that the determining the bladder cancer (e.g., MIBC) subtype alone or in combination with one or more of the number of methods can be used to determine whether or not said patient is a candidate for treatment with a specific type or types of cancer therapy. In addition to assessing or determining a bladder cancer (e.g., MIBC) subtype, the number of methods for characterizing the sample can entail determining a proliferation score, the tumor mutation burden (TMB), the level of immune activation, the cell of origin subtype or any combination thereof. In one embodiment, one or all of the methods for characterizing the sample can be performed on RNA sequencing data obtained from the sample.

[00143] In one embodiment, in addition to assessing the bladder cancer (e.g., MIBC) subtype as provided herein, the characterization entails determining proliferation or proliferation score. In one embodiment, proliferation or the proliferation score is determined

using any method known in the art such as, for example, as provided in US62/789,668 filed January 8, 2019, which is herein incorporated by reference herein.

[00144] In one embodiment, in addition to assessing the bladder cancer (e.g., MIBC) subtype as provided herein, the characterization entails determining the cell of origin subtype. In one embodiment, cell of origin subtype is determined using any method known in the art such as, for example, as provided in Hoadley et al, Cell. 2018 Apr 5;173(2):291-304, which is herein incorporated by reference herein.

[00145] In one embodiment, in addition to determining the bladder cancer (e.g., MIBC) subtype as provided herein, the characterization entails calculating a TMB value and/or rate. The TMB value and/or rate can be calculated using any method known in the art. In one embodiment, the TMB value and/or rate can be calculated from RNA (e.g., via transcriptome profiling or RNA sequencing)) as provided in US 62/771,702 filed November 27, 2018 and US 62/743,257 filed October 9, 2018, which is herein incorporated by reference herein.

[00146] The determination of whether or not said patient is a candidate for treatment with a specific type or types of cancer therapy can be based on the bladder cancer (e.g., MIBC) subtype alone or in combination with other methods known in the art for characterizing a sample obtained from a patient suffering from or suspected of suffering from cancer. The other methods for characterizing said sample can be histologically based methods, gene expression based methods or a combination thereof. The histologically based methods can include histological cancer subtyping by one or more trained pathologists as well as the histological based methods of assessing proliferation such as, for example, determining the mitotic activity index. The gene expression based methods can include bladder cancer subtyping, assessment of TMB, assessment of cell of origin subtype, immune subtyping or any combination thereof. The gene expression based methods can be assessed from DNA, RNA or a combination thereof. In one embodiment, the characterization of the sample obtained from the patient suffering from or suspected of suffering from cancer is performed on RNA obtained or isolated from the sample.

[00147] The gene expression based tissue of origin cancer subtyping can be determined using gene signatures known in the art for specific types of cancer. In one embodiment, cell of origin subtype is determined using any method known in the art such as, for example, as provided in Hoadley et al, Cell. 2018 Apr 5;173(2):291-304, which is herein incorporated by reference herein.

[00148] The gene expression based immune subtyping or immune cell activation can be determined using immune expression signatures known in the art such as, for example, the

gene signatures found in Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A. and Ziv, E., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830, which is herein incorporated by reference in its entirety. In one embodiment, immune cell activation is determined by monitoring the immune cell signatures of Bindea et al (*Immunity* 2013; 39(4); 782-795), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the method further comprises measuring single gene immune biomarkers, such as, for example, CTLA4, PDCD1 and CD274 (PD-L1), PDCDLG2(PD-L2) and/or IFN gene signatures. In one embodiment, the level of immune cell activation is determined by measuring gene expression signatures of immunomarkers. The immunomarkers can be measured in the same and/or different sample used to determine the bladder cancer (e.g., MIBC) subtype as described herein. The immunomarkers can be those found in WO2017/201165, and WO2017/201164, each of which is herein incorporated by reference in their entirety.

[00149] The gene expression based method for calculating a TMB value and/or rate can be any method known in the art. In one embodiment, the TMB value and/or rate can be calculated from RNA (e.g., via transcriptome profiling or RNA sequencing)) as provided in US 62/771,702 filed November 27, 2018 and US 62/743,257 filed October 9, 2018, which is herein incorporated by reference herein.

[00150] In one embodiment, upon determining a patient's bladder cancer subtype (e.g., by measuring the expression of all or subsets of the genes in **Table 1** or **Table 2**), the patient is selected for suitable therapy, for example, radiotherapy (radiation therapy), surgical intervention, target therapy, chemotherapy or drug therapy with an angiogenesis inhibitor or immunotherapy or combinations thereof. In some embodiments, the suitable treatment can be any treatment or therapeutic method that can be used for a bladder cancer patient. In one embodiment, upon determining a patient's bladder cancer subtype, the patient is administered a suitable therapeutic agent, for example chemotherapeutic agent(s) or an angiogenesis inhibitor or immunotherapeutic agent(s). In one embodiment, the therapy is immunotherapy, and the immunotherapeutic agent is a checkpoint inhibitor, monoclonal antibody, biological response modifier, therapeutic vaccine or cellular immunotherapy. In some embodiments, the determination of a suitable treatment can identify treatment responders. In some embodiments, the determination of a suitable treatment can identify treatment non-responders. In some embodiments, upon determining a patient's bladder cancer subtype, the bladder cancer patient can be selected for any combination of suitable therapies. For example, chemotherapy or drug therapy with a radiotherapy, a tumor dissection with an

immunotherapy or a chemotherapeutic agent with a radiotherapy. In some embodiments, immunotherapy, or immunotherapeutic agent can be a checkpoint inhibitor, monoclonal antibody, biological response modifier, therapeutic vaccine or cellular immunotherapy.

[00151] The methods of present invention are also useful for evaluating clinical response to therapy, as well as for endpoints in clinical trials for efficacy of new therapies. The extent to which sequential diagnostic expression profiles move towards normal can be used as one measure of the efficacy of the candidate therapy.

[00152] In one embodiment, the methods as provided herein also find use in predicting response to different lines of therapies based on the subtype of bladder cancer (e.g., MIBC). For example, chemotherapeutic response can be improved by more accurately assigning tumor subtypes. Likewise, treatment regimens can be formulated based on the tumor subtype. In one embodiment, a bladder cancer (i.e., MIBC) sample obtained from a subject is determined to have a luminal subtype using the methods provided herein and is selected for treatment with FGFR inhibitors (e.g., FGFR3 inhibitors). The FGFR inhibitors can be tyrosine kinase inhibitors of FGFR3. In one embodiment, a bladder cancer (i.e., MIBC) sample obtained from a subject is determined to have a luminal-infiltrated subtype using the methods provided herein and is selected for immunotherapy. The immunotherapy can be immune checkpoint therapy. The checkpoint inhibitor can be PD-1, PD-L1, CTLA-4 checkpoint inhibitors or a combination thereof. The PD-1, PD-L1, CTLA-4 checkpoint inhibitors can be any PD-1, PD-L1, CTLA-4 checkpoint inhibitors known in the art and/or provided herein. In one embodiment, the checkpoint inhibitor can be a PD-1/PD-L1 checkpoint inhibitor such as, for example, atezolizumab. In one embodiment, a bladder cancer (i.e., MIBC) sample obtained from a subject is determined to have a basal subtype using the methods provided herein and is selected for chemotherapy (e.g., cisplatin-based neoadjuvant chemotherapy) and/or immunotherapy. The immunotherapy can be immune checkpoint therapy. The checkpoint inhibitor can be PD-1, PD-L1, CTLA-4 checkpoint inhibitors or a combination thereof. The PD-1, PD-L1, CTLA-4 checkpoint inhibitors can be any PD-1, PD-L1, CTLA-4 checkpoint inhibitors known in the art and/or provided herein. In one embodiment, a bladder cancer (i.e., MIBC) sample obtained from a subject is determined to have a basal-infiltrated/neuronal subtype using the methods provided herein and is selected for chemotherapy. The chemotherapy can be etoposide-cisplatin therapy.

[00153] In one embodiment, determination of a specific subtype of bladder cancer (e.g., MIBC) using the methods provided herein can serve as a negative predictor for use of specific treatments for subjects with said bladder cancer (e.g., MIBC) subtype. For example,

a bladder cancer (i.e., MIBC) sample obtained from a subject is determined to have a luminal subtype or a luminal infiltrated subtype using the methods provided herein and thus may not be selected for treatment with chemotherapy such as, cisplatin-based neoadjuvant chemotherapy. Accordingly, validation of the luminal or luminal-infiltrated subtype of a bladder cancer (e.g., MIBC) sample obtained from a subject using the methods provided herein can serve as negative predictors for treatment with chemotherapeutic agents and thus suggest exploration of alternative treatment strategies including, for example, targeted therapies (e.g., immunotherapy).

Immunotherapy

[00154] In one embodiment, provided herein is a method for determining whether a bladder cancer patient is likely to respond to immunotherapy by determining the subtype of bladder cancer of a sample obtained from the patient and, based on the bladder cancer subtype, assessing whether the patient is likely to respond to immunotherapy. In another embodiment, provided herein is a method of selecting a patient suffering from bladder cancer for immunotherapy by determining a bladder cancer subtype of a sample from the patient and, based on the bladder cancer subtype, selecting the patient for immunotherapy. The determination of the bladder cancer subtype of the sample obtained from the patient can be performed using any method for subtyping bladder cancer known in the art. The determination of the bladder cancer subtype of the sample obtained from the patient can be performed using any method for subtyping bladder cancer provided herein. In one embodiment, the sample obtained from the patient has been previously diagnosed as being bladder cancer, and the methods provided herein are used to determine the bladder cancer subtype of the sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists. In one embodiment, the bladder cancer has been diagnosed to be MIBC. In one embodiment, the bladder cancer subtyping is performed via gene expression analysis of a set or panel of biomarkers or subsets thereof in order to generate an expression profile. The gene expression analysis can be performed on a bladder cancer sample obtained from a patient in order to determine the presence, absence or level of expression of one or more biomarkers selected from a publically available bladder cancer database described herein and/or **Tables 1** or **2** provided herein. The bladder cancer subtype (e.g., MIBC subtype) can be selected from the group consisting of luminal, luminal infiltrated, basal, and basal infiltrated/neuronal. The immunotherapy can be

any immunotherapy provided herein. In one embodiment, the immunotherapy comprises administering one or more checkpoint inhibitors. The checkpoint inhibitors can be any checkpoint inhibitor provided herein such as, for example, a checkpoint inhibitor that targets PD-1, PD-L1 or CTLA4.

[00155] As disclosed herein, the biomarkers panels, or subsets thereof, can be those disclosed in any publically available liver cancer gene expression dataset or datasets. In one embodiment, the biomarker panel or subset thereof is, for example, the cancer genome atlas bladder cancer gene expression dataset (n=412). See Robertson, AG, et al., *Cell*, 171(3): 540-556 (2017), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the biomarker panel or subset thereof is, for example, the bladder cancer gene expression dataset (n=305) disclosed in Seiler et al. *Eur Urol* 72(4):544-554 (2017), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the biomarker panel or subset thereof is, for example, the bladder cancer gene expression dataset (n=308) disclosed in Sjodahl, G., et al, *Clin Cancer Res*, 18(12):3377-86 (2012), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the biomarker panel or subset thereof is, for example, the bladder cancer gene expression dataset disclosed in **Table 1** or **Table 2** in combination with one or more biomarkers from a publically available bladder cancer expression dataset.

[00156] In one embodiment, from about 1 to about 5, about 5 to about 10, from about 5 to about 15, from about 5 to about 20, from about 5 to about 25, from about 5 to about 30, from about 5 to about 35, from about 5 to about 40, from about 5 to about 45, from about 5 to about 50, from about 5 to about 55, from about 5 to about 60, of the biomarkers in any of the bladder cancer gene expression datasets provided herein, including, for example, **Table 1** (or **Table 2**) for an bladder cancer sample are detected in a method to determine the bladder cancer subtype as provided herein. In another embodiment, each of the biomarkers from any one of the bladder cancer gene expression datasets provided herein, including, for example, **Table 1** (or **Table 2**) for a bladder cancer sample are detected in a method to determine the bladder cancer subtype as provided herein.

[00157] In one embodiment, the methods provided herein further comprise determining the presence, absence or level of immune activation in a bladder cancer subtype. The presence or level of immune cell activation can be determined by creating an expression profile or detecting the expression of one or more biomarkers associated with innate immune

cells and/or adaptive immune cells associated with each bladder cancer subtype in a sample obtained from a patient. In one embodiment, immune cell activation associated with a bladder cancer subtype is determined by monitoring the immune cell signatures of Bindea et al (Immunity 2013; 39(4); 782-795), Faruki H. et al., JTO, 12(6): 943-953 (2017), Charoentong P. et al., Cell reports, 18, 248-262 (2017), Thorsson, V. et al., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830, and/or WO2017/201165 and WO2017/201164, the contents of each of which are herein incorporated by reference in its entirety. In one embodiment, the method further comprises measuring single gene immune biomarkers, such as, for example, CTLA4, PDCD1 and CD274 (PD-L1), PDCDLG2(PD-L2) and/or IFN gene signatures. In one embodiment, immune subtyping or immune cell activation can be determined using the gene signatures found in Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A. and Ziv, E., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830, which is herein incorporated by reference in its entirety. The presence or a detectable level of immune activation (Innate and/or Adaptive) associated with a bladder cancer subtype can indicate or predict that a patient with said bladder cancer subtype may be amendable to immunotherapy. The immunotherapy can be treatment with a checkpoint inhibitor as provided herein. In one embodiment, a method is provided herein for detecting the expression of at least one classifier biomarker provided herein in a sample (e.g., bladder cancer sample) obtained from a patient further comprises administering an immunotherapeutic agent following detection of immune activation as provided herein in said sample.

[00158] In one embodiment, the method comprises determining a subtype of a bladder cancer sample and subsequently determining a level of immune cell activation of said subtype. In one embodiment, the subtype is determined by determining the expression levels of one or more classifier biomarkers using sequencing (e.g., RNASeq), amplification (e.g., qRT-PCR) or hybridization assays (e.g., microarray analysis) as described herein. The one or more biomarkers can be selected from a publically available database (e.g., TCGA bladder cancer gene expression datasets or any other publically available bladder cancer gene expression datasets provided herein). In some embodiments, the biomarkers of **Table 1** or **Table 2** can be used to specifically determine the subtype of a bladder cancer sample obtained from a patient. In one embodiment, the level of immune cell activation is determined by measuring gene expression signatures of immunomarkers. The immunomarkers can be measured in the same and/or different sample used to subtype the bladder cancer sample as described herein. The immunomarkers that can be measured can comprise, consist of, or consistently

essentially of innate immune cell (IIC) and/or adaptive immune cell (AIC) gene signatures, interferon (IFN) gene signatures, individual immunomarkers, major histocompatibility complex class II (MHC class II) genes or a combination thereof. The gene expression signatures for IICs, AICs, IFN and MHC class II can be any known gene signatures for said cell types or genes known in the art. For example, the immune gene signatures can be those from Bindea et al. (*Immunity* 2013; 39(4): 782-795), Faruki H. et al., *JTO*, 12(6): 943-953 (2017), Charoentong P. et al., *Cell reports*, 18, 248-262 (2017) and/or WO2017/201165 and WO2017/201164. The individual immunomarkers can be those found in Faruki H. et al., *JTO*, 12(6): 943-953 (2017), Charoentong P. et al., *Cell reports*, 18, 248-262 (2017) and/or WO2017/201165 and WO2017/201164. In one embodiment, the individual immunomarkers can be CTLA4 (e.g., NM_005214 (SEQ ID NO. 80), PDCD1 (e.g., NM_005018; SEQ ID NO. 81), PDL2 (e.g., AY254343; SEQ ID NO. 82) and CD274 (PD-L1; e.g., NM_014143; SEQ ID NO. 83). In one embodiment, immune subtyping or immune cell activation can be determined using the gene signatures found in Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A. and Ziv, E., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830.

[00159] In one embodiment, upon determining a patient's bladder cancer subtype using any of the methods and classifier biomarkers panels or subsets thereof as provided herein, the patient is selected for treatment with or administered an immunotherapeutic agent. The immunotherapeutic agent can be a checkpoint inhibitor, monoclonal antibody, biological response modifiers, therapeutic vaccine or cellular immunotherapy.

[00160] In another embodiment, the immunotherapeutic agent is a checkpoint inhibitor. In some cases, a method for determining the likelihood of response to one or more checkpoint inhibitors is provided. In one embodiment, the checkpoint inhibitor is a PD-1/PD-L1 checkpoint inhibitor. The PD-1/PD-L1 checkpoint inhibitor can be nivolumab, pembrolizumab, atezolizumab, durvalumab, lambrolizumab, or avelumab. In one embodiment, the checkpoint inhibitor is a CTLA-4 checkpoint inhibitor. The CTLA-4 checkpoint inhibitor can be ipilimumab or tremelimumab. In one embodiment, the checkpoint inhibitor is a combination of checkpoint inhibitors such as, for example, a combination of one or more PD-1/PD-L1 checkpoint inhibitors used in combination with one or more CTLA-4 checkpoint inhibitors.

[00161] In one embodiment, the immunotherapeutic agent is a monoclonal antibody. In some cases, a method for determining the likelihood of response to one or more monoclonal antibodies is provided. The monoclonal antibody can be directed against tumor cells or

directed against tumor products. The monoclonal antibody can be panitumumab, matuzumab, necitumunab, trastuzumab, amatuximab, bevacizumab, ramucirumab, bavituximab, patritumab, rilotumumab, cetuximab, immu-132, or demcizumab.

[00162] In yet another embodiment, the immunotherapeutic agent is a therapeutic vaccine. In some cases, a method for determining the likelihood of response to one or more therapeutic vaccines is provided. The therapeutic vaccine can be a peptide or tumor cell vaccine. The vaccine can target MAGE-3 antigens, NY-ESO-1 antigens, p53 antigens, survivin antigens, or MUC1 antigens. The therapeutic cancer vaccine can be GVAX (GM-CSF gene-transfected tumor cell vaccine), belagenpumatucel-L (allogeneic tumor cell vaccine made with four irradiated NSCLC cell lines modified with TGF-beta2 antisense plasmid), MAGE-A3 vaccine (composed of MAGE-A3 protein and adjuvant AS15), (I)-BLP-25 anti-MUC-1 (targets MUC-1 expressed on tumor cells), CimaVax EGF (vaccine composed of human recombinant Epidermal Growth Factor (EGF) conjugated to a carrier protein), WT1 peptide vaccine (composed of four Wilms' tumor suppressor gene analogue peptides), CRS-207 (live-attenuated *Listeria monocytogenes* vector encoding human mesothelin), Bec2/BCG (induces anti-GD3 antibodies), GV1001 (targets the human telomerase reverse transcriptase), TG4010 (targets the MUC1 antigen), racotumomab (anti-idiotypic antibody which mimicks the NGcGM3 ganglioside that is expressed on multiple human cancers), tecemotide (liposomal BLP25; liposome-based vaccine made from tandem repeat region of MUC1) or DRibbles (a vaccine made from nine cancer antigens plus TLR adjuvants).

[00163] In one embodiment, the immunotherapeutic agent is a biological response modifier. In some cases, a method for determining the likelihood of response to one or more biological response modifiers is provided. The biological response modifier can trigger inflammation such as, for example, PF-3512676 (CpG 7909) (a toll-like receptor 9 agonist), CpG-ODN 2006 (downregulates Tregs), *Bacillus Calmette-Guerin* (BCG), *Mycobacterium vaccae* (SRL172) (nonspecific immune stimulants now often tested as adjuvants). The biological response modifier can be cytokine therapy such as, for example, IL-2+ tumor necrosis factor alpha (TNF-alpha) or interferon alpha (induces T-cell proliferation), interferon gamma (induces tumor cell apoptosis), or Mda-7 (IL-24) (Mda-7/IL-24 induces tumor cell apoptosis and inhibits tumor angiogenesis). The biological response modifier can be a colony-stimulating factor such as, for example granulocyte colony-stimulating factor. The biological response modifier can be a multi-modal effector such as, for example, multi-target VEGFR: thalidomide and analogues such as lenalidomide and pomalidomide,

cyclophosphamide, cyclosporine, denileukin diftitox, talactoferrin, trabectedin or all-trans-retinoic acid.

[00164] In one embodiment, the immunotherapy is cellular immunotherapy. In some cases, a method for determining the likelihood of response to one or more cellular therapeutic agents. The cellular immunotherapeutic agent can be dendritic cells (DCs) (ex vivo generated DC-vaccines loaded with tumor antigens), T-cells (ex vivo generated lymphokine-activated killer cells; cytokine-induced killer cells; activated T-cells; gamma delta T-cells), or natural killer cells.

[00165] In some cases, specific subtypes of bladder cancer have different levels of immune activation (e.g., innate immunity and/or adaptive immunity) such that subtypes with elevated or detectable immune activation (e.g., innate immunity and/or adaptive immunity) are selected for treatment with one or more immunotherapeutic agents described herein. In some cases, specific subtypes of bladder cancer have high or elevated levels of immune activation. In some cases, the luminal, luminal infiltrated, basal, and/or basal infiltrated/neuronal subtype has elevated levels of immune activation (e.g., innate immunity and/or adaptive immunity) as compared to other bladder cancer subtypes. In some cases, the luminal, luminal infiltrated, basal, and/or basal infiltrated/neuronal subtype has reduced levels of immune activation (e.g., innate immunity and/or adaptive immunity) as compared to other bladder cancer subtypes. In one embodiment, bladder cancer subtypes with low levels of or no immune activation (e.g., innate immunity and/or adaptive immunity) are not selected for treatment with one or more immunotherapeutic agents described herein.

Angiogenesis Inhibitors

[00166] In one embodiment, upon determining a patient's or subject's bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.), the patient is selected for drug therapy with an angiogenesis inhibitor.

[00167] In one embodiment, the angiogenesis inhibitor is a vascular endothelial growth factor (VEGF) inhibitor, a VEGF receptor inhibitor, a platelet derived growth factor (PDGF) inhibitor or a PDGF receptor inhibitor.

[00168] In general, methods of determining whether a patient is likely to respond to angiogenesis inhibitor therapy, or methods of selecting a patient for angiogenesis inhibitor therapy are provided herein. In one embodiment, the method comprises determining a

bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) and probing a sample from the patient for the levels of at least five hypoxia biomarkers selected from the group consisting of RRAGD, FABP5, UCHL1, GAL, PLOD, DDIT4, VEGF, ADM, ANGPTL4, NDRG1, NP, SLC16A3, and C14ORF58 (see **Table A**) at the nucleic acid level. In a further embodiment, the probing step comprises mixing the sample with five or more oligonucleotides that are substantially complementary to portions of nucleic acid molecules of the at least five biomarkers under conditions suitable for hybridization of the five or more oligonucleotides to their complements or substantial complements, detecting whether hybridization occurs between the five or more oligonucleotides to their complements or substantial complements; and obtaining hybridization values of the sample based on the detecting steps. The hybridization values of the sample are then compared to reference hybridization value(s) from at least one sample training set, wherein the at least one sample training set comprises (i) hybridization value(s) of the at least five biomarkers from a sample that overexpresses the at least five biomarkers, or overexpresses a subset of the at least five biomarkers, (ii) hybridization values of the at least five biomarkers from a reference bladder cancer (e.g., MIBC) sample, or (iii) hybridization values of the at least five biomarkers from a non-bladder cancer (e.g., MIBC) sample. A determination of whether the patient is likely to respond to angiogenesis inhibitor therapy, or a selection of the patient for angiogenesis inhibitor is then made based upon (i) the patient's bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) and (ii) the results of comparison.

Table A. Biomarkers for hypoxia profile		
Name	Abbreviation	GenBank Accession No.
RRAGD	Ras-related GTP binding D	BC003088
FABP5	fatty acid binding protein 5	M94856
UCHL1	ubiquitin carboxyl-terminal esterase L1	NM_004181
GAL	Galanin	BC030241
PLOD	procollagen-lysine, 2-oxoglutarate 5-dioxygenase lysine hydroxylase	M98252

Table A. Biomarkers for hypoxia profile		
Name	Abbreviation	GenBank Accession No.
DDIT4	DNA-damage-inducible transcript 4	NM__019058
VEGF	vascular endothelial growth factor	M32977
ADM	Adrenomedullin	NM__001124
ANGPTL4	angiopoietin-like 4	AF202636
NDRG1	N-myc downstream regulated gene 1	NM__006096
NP	nucleoside phosphorylase	NM 000270
SLC16A3	solute carrier family 16 monocarboxylic acid transporters, member 3	NM__004207
C14ORF58	chromosome 14 open reading frame 58	AK000378

[00169] The aforementioned set of thirteen biomarkers, or a subset thereof, is also referred to herein as a “hypoxia profile”.

[00170] In one embodiment, the method provided herein includes determining the levels of at least five biomarkers, at least six biomarkers, at least seven biomarkers, at least eight biomarkers, at least nine biomarkers, or at least ten biomarkers, or five to thirteen, six to thirteen, seven to thirteen, eight to thirteen, nine to thirteen or ten to thirteen biomarkers selected from RRAGD, FABP5, UCHL1, GAL, PLOD, DDIT4, VEGF, ADM, ANGPTL4, NDRG1, NP, SLC16A3, and C14ORF58 in a sample obtained from a subject. Biomarker expression in some instances may be normalized against the expression levels of all RNA transcripts or their expression products in the sample or against a reference set of RNA transcripts or their expression products. The reference set as explained throughout, may be an actual sample that is tested in parallel with the sample, or may be a reference set of values from a database or stored dataset. Levels of expression, in one embodiment, are reported in number of copies, relative fluorescence value or detected fluorescence value. The level of expression of the biomarkers of the hypoxia profile together with the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) as determined using the methods provided herein can be used in the methods described herein to determine whether a patient is likely to respond to angiogenesis inhibitor therapy.

[00171] In one embodiment, the levels of expression of the thirteen biomarkers (or subsets thereof, as described above, *e.g.*, five or more, from about five to about 13), are normalized against the expression levels of all RNA transcripts or their non-natural cDNA expression products, or protein products in the sample, or of a reference set of RNA transcripts or a reference set of their non-natural cDNA expression products, or a reference set of their protein products in the sample.

[00172] In one embodiment, angiogenesis inhibitor treatments include, but are not limited to an integrin antagonist, a selectin antagonist, an adhesion molecule antagonist, an antagonist of intercellular adhesion molecule (ICAM)-1, ICAM-2, ICAM-3, platelet endothelial adhesion molecule (PCAM), vascular cell adhesion molecule (VCAM)), lymphocyte function-associated antigen 1 (LFA-1), a basic fibroblast growth factor antagonist, a vascular endothelial growth factor (VEGF) modulator, a platelet derived growth factor (PDGF) modulator (*e.g.*, a PDGF antagonist).

[00173] In one embodiment of determining whether a subject is likely to respond to an integrin antagonist, the integrin antagonist is a small molecule integrin antagonist, for example, an antagonist described by Paolillo *et al.* (Mini Rev Med Chem, 2009, volume 12, pp. 1439-1446, incorporated by reference in its entirety), or a leukocyte adhesion-inducing cytokine or growth factor antagonist (*e.g.*, tumor necrosis factor- α (TNF- α), interleukin-1 β (IL-1 β), monocyte chemotactic protein-1 (MCP-1) and a vascular endothelial growth factor (VEGF)), as described in U.S. Patent No. 6,524,581, incorporated by reference in its entirety herein.

[00174] The methods provided herein are also useful for determining whether a subject is likely to respond to one or more of the following angiogenesis inhibitors: interferon gamma 1 β , interferon gamma 1 β (Actimmune®) with pirfenidone, ACUHTR028, α V β 5, aminobenzoate potassium, amyloid P, ANG1122, ANG1170, ANG3062, ANG3281, ANG3298, ANG4011, anti-CTGF RNAi, Aplidin, astragalus membranaceus extract with salvia and schisandra chinensis, atherosclerotic plaque blocker, Azol, AZX100, BB3, connective tissue growth factor antibody, CT140, danazol, Esbriet, EXC001, EXC002, EXC003, EXC004, EXC005, F647, FG3019, Fibrocorin, Follistatin, FT011, a galectin-3 inhibitor, GKT137831, GMCT01, GMCT02, GRMD01, GRMD02, GRN510, Heberon Alfa R, interferon α -2 β , ITMN520, JKB119, JKB121, JKB122, KRX168, LPA1 receptor antagonist, MGN4220, MIA2, microRNA 29a oligonucleotide, MMI0100, noscapine, PBI4050, PBI4419, PDGFR inhibitor, PF-06473871, PGN0052, Pirespa, Pirfenex,

pirfenidone, plitidepsin, PRM151, Px102, PYN17, PYN22 with PYN17, Relivergen, rhPTX2 fusion protein, RXI109, secretin, STX100, TGF- β Inhibitor, transforming growth factor, β -receptor 2 oligonucleotide, VA999260, XV615 or a combination thereof.

[00175] In another embodiment, a method is provided for determining whether a subject is likely to respond to one or more endogenous angiogenesis inhibitors. In a further embodiment, the endogenous angiogenesis inhibitor is endostatin, a 20 kDa C-terminal fragment derived from type XVIII collagen, angiostatin (a 38 kDa fragment of plasmin), a member of the thrombospondin (TSP) family of proteins. In a further embodiment, the angiogenesis inhibitor is a TSP-1, TSP-2, TSP-3, TSP-4 and TSP-5. Methods for determining the likelihood of response to one or more of the following angiogenesis inhibitors are also provided a soluble VEGF receptor, *e.g.*, soluble VEGFR-1 and neuropilin 1 (NPR1), angiopoietin-1, angiopoietin-2, vasostatin, calreticulin, platelet factor-4, a tissue inhibitor of metalloproteinase (TIMP) (*e.g.*, TIMP1, TIMP2, TIMP3, TIMP4), cartilage-derived angiogenesis inhibitor (*e.g.*, peptide troponin I and chondromodulin I), a disintegrin and metalloproteinase with thrombospondin motif 1, an interferon (IFN), (*e.g.*, IFN- α , IFN- β , IFN- γ), a chemokine, *e.g.*, a chemokine having the C-X-C motif (*e.g.*, CXCL10, also known as interferon gamma-induced protein 10 or small inducible cytokine B10), an interleukin cytokine (*e.g.*, IL-4, IL-12, IL-18), prothrombin, antithrombin III fragment, prolactin, the protein encoded by the *TNFSF15* gene, osteopontin, maspin, canstatin, proliferin-related protein.

[00176] In one embodiment, a method for determining the likelihood of response to one or more of the following angiogenesis inhibitors is provided is angiopoietin-1, angiopoietin-2, angiostatin, endostatin, vasostatin, thrombospondin, calreticulin, platelet factor-4, TIMP, CDAI, interferon α , interferon β , vascular endothelial growth factor inhibitor (VEGI) meth-1, meth-2, prolactin, VEGI, SPARC, osteopontin, maspin, canstatin, proliferin-related protein (PRP), restin, TSP-1, TSP-2, interferon gamma 1 β , ACUHTR028, α V β 5, aminobenzoate potassium, amyloid P, ANG1122, ANG1170, ANG3062, ANG3281, ANG3298, ANG4011, anti-CTGF RNAi, Aplidin, astragalus membranaceus extract with salvia and schisandra chinensis, atherosclerotic plaque blocker, Azol, AZX100, BB3, connective tissue growth factor antibody, CT140, danazol, Esbriet, EXC001, EXC002, EXC003, EXC004, EXC005, F647, FG3019, Fibrocorin, Follistatin, FT011, a galectin-3 inhibitor, GKT137831, GMCT01, GMCT02, GRMD01, GRMD02, GRN510, Heberon Alfa R, interferon α -2 β , ITMN520, JKB119, JKB121, JKB122, KRX168, LPA1 receptor antagonist, MGN4220, MIA2, microRNA 29a oligonucleotide, MMI0100, noscapine, PBI4050, PBI4419, PDGFR inhibitor,

PF-06473871, PGN0052, Pirespa, Pirfenex, pirfenidone, plitidepsin, PRM151, Px102, PYN17, PYN22 with PYN17, Relivergen, rhPTX2 fusion protein, RXI109, secretin, STX100, TGF- β Inhibitor, transforming growth factor, β -receptor 2 oligonucleotide, VA999260, XV615 or a combination thereof.

[00177] In yet another embodiment, the angiogenesis inhibitor can include pazopanib (Votrient), sunitinib (Sutent), sorafenib (Nexavar), axitinib (Inlyta), ponatinib (Iclusig), vandetanib (Caprelsa), cabozantinib (Cometrig), ramucirumab (Cyramza), regorafenib (Stivarga), ziv-aflibercept (Zaltrap), motesanib, or a combination thereof. In another embodiment, the angiogenesis inhibitor is a VEGF inhibitor. In a further embodiment, the VEGF inhibitor is axitinib, cabozantinib, aflibercept, brivanib, tivozanib, ramucirumab or motesanib. In yet a further embodiment, the angiogenesis inhibitor is motesanib.

[00178] In one embodiment, the methods provided herein relate to determining a subject's likelihood of response to an antagonist of a member of the platelet derived growth factor (PDGF) family, for example, a drug that inhibits, reduces or modulates the signaling and/or activity of PDGF-receptors (PDGFR). For example, the PDGF antagonist, in one embodiment, is an anti-PDGF aptamer, an anti-PDGF antibody or fragment thereof, an anti-PDGFR antibody or fragment thereof, or a small molecule antagonist. In one embodiment, the PDGF antagonist is an antagonist of the PDGFR- α or PDGFR- β . In one embodiment, the PDGF antagonist is the anti-PDGF- β aptamer E10030, sunitinib, axitinib, sorafenib, imatinib, imatinib mesylate, nintedanib, pazopanib HCl, ponatinib, MK-2461, dovitinib, pazopanib, crenolanib, PP-121, telatinib, imatinib, KRN 633, CP 673451, TSU-68, Ki8751, amuvatinib, tivozanib, masitinib, motesanib diphosphate, dovitinib dilactic acid, linifanib (ABT-869).

[00179] Upon making a determination of whether a patient is likely to respond to angiogenesis inhibitor therapy, or selecting a patient for angiogenesis inhibitor therapy, in one embodiment, the patient is administered the angiogenesis inhibitor. The angiogenesis inhibitor can be any of the angiogenesis inhibitors described herein.

Radiotherapy

[00180] In one embodiment, provided herein is a method for determining whether a patient is likely to respond to radiotherapy by determining the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) of a sample obtained from the patient and, based on the

bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.), assessing whether the patient is likely to respond to or benefit from radiotherapy. In another embodiment, provided herein is a method of selecting a patient suffering from cancer for radiotherapy by determining a bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) of a sample from the patient and, based on the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.), selecting the patient for radiotherapy.

[00181] In some embodiments, the radiotherapy can include but are not limited to proton therapy and external-beam radiation therapy. In some embodiments, the radiotherapy can include any types or forms of treatment that is suitable for patients with specific types of cancer.

[00182] In some embodiments, a patient with a specific type of cancer can have or display resistance to radiotherapy. Radiotherapy resistance in any cancer or subtype thereof can be determined by measuring or detecting the expression levels of one or more genes known in the art and/or provided herein associated with or related to the presence of radiotherapy resistance. Genes associated with radiotherapy resistance can include NFE2L2, KEAP1 and CUL3. In some embodiments, radiotherapy resistance can be associated with the alterations of KEAP1 (Kelch-like ECH-associated protein 1)/NRF2 (nuclear factor E2-related factor 2) pathway. Association of a particular gene to radiotherapy resistance can be determined by examining expression of said gene in one or more patients known to be radiotherapy non-responders and comparing expression of said gene in one or more patients known to be radiotherapy responders.

Surgical Intervention

[00183] In one embodiment, provided herein is a method for determining whether a bladder cancer (e.g., MIBC) cancer patient is likely to respond to surgical intervention by determining the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype,

proliferation signature or score, immune subtype and/or TMB status, etc.) of a sample obtained from the patient and, based on the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.), assessing whether the patient is likely to respond to or benefit from surgery. In another embodiment, provided herein is a method of selecting a patient suffering from cancer for surgery by determining a bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.) of a sample from the patient and, based on the bladder cancer (e.g., MIBC) subtype alone or in combination with other characterization methods as described herein (e.g., determining cell of origin cancer subtype, proliferation signature or score, immune subtype and/or TMB status, etc.), selecting the patient for surgery. In some embodiments, the surgery can include laser technology, excision, dissection, and reconstructive surgery.

Prediction of Overall Survival Rate and Metastasis for Bladder Cancer Patients

[00184] The present disclosure provides methods for predicting overall survival rate for a bladder cancer patient. In some embodiments, the prediction of overall survival rate can involve obtaining a bladder tissue sample for a bladder cancer patient. In some embodiments, the bladder cancer patients can have various stages of cancers. In some embodiments, the overall survival rate can be determined by detecting the expression level of at least one subtype classifier of a publically available bladder cancer database or dataset. In some embodiments, an overall survival rate can be determined by detecting the expression level (e.g., protein and/or nucleic acid) of any subtype classifiers that are relevant to bladder cancer. In one embodiment, the subtype classifiers can be all or a subset of classifiers from **Table 1** or **Table 2**.

[00185] In some embodiments, the present disclosure further provides methods of predicting overall survival in bladder cancer from specific areas of the bladder. In some embodiments, the prediction includes detecting an expression level of at least one gene from a bladder cancer dataset (e.g., Table 1 or Table 2) in a bladder tissue sample obtained from a patient. In some embodiments, the detection of the expression level of a subtype classifier from a bladder cancer dataset (e.g., Table 1 or Table 2) using the methods provided herein specifically identifies a luminal, luminal infiltrated, basal, or basal infiltrated/neuronal

subtype. In some embodiments, the identification of the subtype is indicative of the overall survival in the patient. For example, a basal-infiltrated/neuronal subtype or a luminal-infiltrated subtype of bladder cancer (e.g., MIBC) is indicative of a poor overall survival.

[00186] The present disclosure provides methods for predicting nodal metastasis for a bladder cancer patient. In some embodiments, the prediction of nodal metastasis can involve obtaining a bladder tissue sample for a patient. In some embodiments, the patients can have various stages of cancers. In some embodiments, the nodal metastasis can be determined by detecting the expression level of at least one subtype classifier from a bladder gene set. The bladder gene set can be a publically available bladder database or a bladder gene set provided herein (e.g. Table 1 or Table 2) or a combination thereof. The publically available bladder gene set can be a TCGA bladder cancer gene set. In one embodiment, nodal metastasis of bladder cancer can be determined by detecting the expression level of all the subtype classifiers or subsets thereof of the classifiers found in Table 1 or Table 2.

[00187] In some embodiments, the luminal, luminal infiltrated, basal, or basal infiltrated/neuronal subtype of bladder cancer can be more likely to be associated with nodal metastasis compared with other subtypes. In some embodiments, the luminal, luminal infiltrated, basal, or basal infiltrated/neuronal subtype can be most likely associated with positive lymph node metastasis compared with other subtypes. In some embodiments, the luminal, luminal infiltrated, basal, or basal infiltrated/neuronal subtype can be at least about 0.1 times, at least about 0.2 times, at least about 0.3 times, at least about 0.4 times, at least about 0.5 times, at least about 0.6 times, at least about 0.7 times, at least about 0.8 times, at least about 0.9 times, at least about 1 time, at least about 1.2 times, at least about 1.5 times, at least about 1.7 times, at least about 2.0 times, at least about 2.2 times, at least about 2.5 times, at least about 2.7 times, at least about 3.0 times, at least about 3.2 times, at least about 3.5 times, at least about 3.7 times, at least about 4.0 times, at least about 4.2 times, at least about 4.5 times, at least about 4.7 times, at least about 5.0 times, inclusive of all ranges and subranges therebetween more likely to have occult nodal metastasis compared to other bladder cancer subtypes.

Detection Methods

[00188] In one embodiment, the methods and compositions provided herein allow for the detection of at least one biomarker in a bladder cancer sample obtained from a subject. The at least one biomarker can be a classifier biomarker provided herein. The detection can be at the

nucleic acid level or protein level. In one embodiment, the detection is at the nucleic acid level and the detection can be by using any amplification, hybridization and/or sequencing assay disclosed herein. In one embodiment, the at least one biomarker detected using the methods and compositions provided herein is selected from **Table 1** or **Table 2**. Further to the above embodiment, the detection of the at least one biomarker selected from **Table 1** or **Table 2** is at the nucleic acid level. In one embodiment, the methods of detecting the nucleic acid(s) (e.g., classifier biomarkers) in the bladder cancer sample obtained from the subject comprises, consists essentially of, or consists of measuring the expression level of at least one or a plurality of biomarkers using any of the methods provided herein. The biomarkers can be selected from **Table 1** or **Table 2**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 18 biomarker nucleic acids, at least 36 biomarker nucleic acids, at least 54 biomarker nucleic acids, or all 60 biomarkers nucleic acids of **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 55 biomarker nucleic acids, or all 60 biomarkers nucleic acids of **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 18 biomarker nucleic acids, at least 36 biomarker nucleic acids, at least 54 biomarker nucleic acids, or all 56 biomarkers nucleic acids of **Table 2**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 55 biomarker nucleic acids, or all 56 biomarkers nucleic acids of **Table 2**.

[00189] In another embodiment, the methods and compositions provided herein allow for the detection of at least one or a plurality of biomarkers in a bladder cancer sample obtained from a subject such that the at least one or the plurality of biomarkers are selected from the biomarkers listed in **Table 1** or **Table 2** in combination with the detection of at least one or a plurality of biomarkers from one or more additional sets of biomarkers in a bladder cancer sample obtained from the subject. The detection can be at the nucleic acid level or protein level. In one embodiment, the detection is at the nucleic acid level and the detection can be by using any amplification, hybridization and/or sequencing assay disclosed herein. The one or more additional sets of biomarkers can be selected from a set of biomarkers whose

presence, absence and/or level of expression is indicative of immune activation, proliferation or a combination thereof. The set of biomarkers for indicating immune activation can be gene expression signatures of and/or Adaptive Immune Cells (AIC) and/or Innate Immune Cells (IIC), individual immune biomarkers, interferon genes, major histocompatibility complex, class II (MHC II) genes or a combination thereof. (See, **FIG. 8**). The gene expression signatures for IICs, AICs, IFN and MHC class II can be any known gene signatures for said cell types or genes known in the art. For example, the immune gene signatures can be those from Bindea et al. (*Immunity* 2013; 39(4): 782-795), Faruki H. et al., *JTO*, 12(6): 943-953 (2017), Charoentong P. et al., *Cell reports*, 18, 248-262 (2017) and/or WO2017/201165 and WO2017/201164. The individual immunomarkers can be those found in Faruki H. et al., *JTO*, 12(6): 943-953 (2017), Charoentong P. et al., *Cell reports*, 18, 248-262 (2017), WO2017/201165 or WO2017/201164. In one embodiment, the individual immunomarkers can be CTLA4 (e.g., NM_005214 (SEQ ID NO. 80), PDCD1 (e.g., NM_005018; SEQ ID NO. 81), PDL2 (e.g., AY254343; SEQ ID NO. 82) and CD274 (PD-L1; e.g., NM_014143; SEQ ID NO. 83). In one embodiment, immune subtyping or immune cell activation can be determined using the gene signatures found in Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A. and Ziv, E., 2018, The immune landscape of cancer. *Immunity*, 48(4), pp.812-830. The additional set of biomarkers for indicating proliferation can be gene expression signatures that include the 11 gene signature comprising BIRC5, CCNB1, CDC20, CDCA1, CEP55, KNTC2, MKI67, PTTG1, RRM2, TYMS, and UBE2C found in Martin M. et al., *Breast Cancer Res Treat*, 138: 457-466 (2013), the 18 gene signature found in US 20160115551 and/or the 26 gene signature found in 62/789,668 filed January 8, 2019.

[00190] Further to any of the above embodiments, the methods and compositions provided herein further comprise determining tumor mutation burden (TMB) and/or TMB rate of the tumor sample. The TMB and/or TMB rate can be determined or calculated using any method known in the art. In one embodiment, the TMB and/or TMB rate is determined from RNA as described in 62/743,257 filed on October 9, 2018 and 62/771,702 filed on November 27, 2018.

[00191] Further to any of the above embodiments, the methods and compositions provided herein further comprise determining a cell of origin subtype of the bladder cancer sample. In one embodiment, cell of origin subtype is determined using any method known in the art such as, for example, as provided in Hoadley et al, *Cell*. 2018 Apr 5;173(2):291-304, which is herein incorporated by reference herein.

Kits

[00192] Kits for practicing the methods provided herein can be further provided. By "kit" can encompass any manufacture (e.g., a package or a container) comprising at least one reagent, e.g., an antibody, a nucleic acid probe or primer, etc., for specifically detecting the expression of a biomarker as provided herein. The kit may be promoted, distributed, or sold as a unit for performing the methods provided herein. Additionally, the kits may contain a package insert describing the kit and methods for its use.

[00193] In one embodiment, kits for practicing the methods as provided herein are provided. Such kits are compatible with both manual and automated immunocytochemistry techniques (e.g., cell staining). These kits comprise at least one antibody directed to a biomarker of interest, chemicals for the detection of antibody binding to the biomarker, a counterstain, and, optionally, a bluing agent to facilitate identification of positive staining cells. Any chemicals that detect antigen- antibody binding may be used in the practice of the methods provided herein. The kits may comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more antibodies for use in the methods provided herein.

[00194] In one embodiment, the kits for practicing the methods provided herein comprise at least one primer pair directed to a biomarker of interest, chemicals for the detection of amplification of the biomarker of interest, and, optionally, any agent necessary for quantifying the detection level of the biomarker of interest. Any chemicals that detect amplification products may be used in the practice of the methods provided herein. The kits may comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more primer pairs for use in the methods provided herein.

[00195] In one embodiment, the kits for practicing the methods provided herein comprise at least one probe directed to a biomarker of interest, chemicals for the detection of hybridization of the probe to the biomarker of interest, and, optionally, any agent necessary for quantifying the level of the biomarker of interest. Any chemicals that detect hybridization products may be used in the practice of the methods provided herein. The kits may comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more probes for use in the methods provided herein.

EXAMPLES

[00196] The present invention is further illustrated by reference to the following Examples. However, it should be noted that these Examples, like the embodiments described above, are illustrative and are not to be construed as restricting the scope of the invention in any way.

Example 1- Development and Validation of the 60-Gene Bladder Cancer Subtyping

Signature

Background

[00197] Recent advances in cancer genomics have led to an increased understanding of mutational and gene expression profiles in bladder cancer. Bladder cancer subtypes, as defined by underlying genomic features, have shown varied cell of origin, tumor drivers, proliferation, immune responses, and prognosis.

[00198] Provided herein is a novel 60-gene expression subtype classification method which provides valuable insight into tumor biology and potential therapeutic response. Differential expression of immune infiltrating cells and biomarkers, tumor mutation burden, therapeutic targets, and overall survival were examined in Muscle Invasive Bladder Cancer (MIBC) subtypes (luminal, luminal infiltrated, basal, and basal infiltrated/neuronal).

Objective

[00199] This example was initiated to address the need for an efficient method for improved tumor classification that could inform prognosis, drug response and patient management based on underlying genomic and biologic tumor characteristics. Using multiple available public datasets, including the dataset described in Robertson, AG, et al., Cell, 171(3): 540-556 (2017) (dataset referred to herein as “Gold standard,” “GS,” or “TCGA-2017”); the dataset described in Cancer Genome Atlas Research Network, Nature, 507(492): 315-22 (2014) (dataset described herein as TCGA-2014); the Seiler dataset (Seiler, R., et al., Eur Urol 72(4):544-554 (2017)); and the Sjodahl dataset (Sjodahl, G., et al., Clin Cancer Res, 18(12):3377-86 (2012)), a bladder cancer subtyping method and algorithm was developed. (All datasets and dataset references are incorporated herein by reference in their entireties.) The diagnostic method developed in this example includes evaluation of gene expression subtypes and application of an algorithm for categorization of bladder cancer into one of 4 subtypes (Luminal (Type I), Luminal Infiltrated (Type II), Basal (Type III), and Basal infiltrated/neuronal (Type IV)).

Methods

[00200] The TCGA-2017 bladder cancer gene expression data was examined using the 2708-gene classifier previously developed by Robertson et al. (2017) to differentiate the luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtypes. (**FIG. 1**) To develop a reduced and clinically applicable gene signature for evaluation of bladder cancer, a 5-fold cross-validation (CV) on the entire TCGA bladder cancer dataset (n=408) was conducted to find the number of genes that would be required to provide differentiation of the subtypes with sufficient agreement with the previously developed gold standard (i.e., aforementioned 2708-gene classifier) as shown in **FIG. 2**. Prototype samples were then chosen based on the gold standard silhouette (**FIG. 3**). A minimal gene set that optimally classified the luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtypes was identified using a modification of the software packages ClaNC. The ClaNC t-statistics were calculated for all 2708 gold standard subtyping genes using the prototype samples and 60 genes were selected based on the ranks of the strongest t-statistics (i.e., select an equal number of negatively and positively correlated genes for each bladder cancer subtype) (**FIG. 4**). A nearest centroid classifier (**FIG. 5, FIG. 6**) was fit using the 60 genes and the prototypes only followed by an evaluation of the full TCGA dataset. Validation of the reduced gene signature was compared to the gold standard gene signature in TCGA datasets (**FIG. 5**) and in several other publicly available datasets (**FIG. 16**) including the Seiler, et. al. (2017), and Sjodahl, et. al. (2012), references.

[00201] Signatures of multiple immune cells (**FIG. 8**), single immune-biomarkers (**FIG. 7**), drug target genes (**FIG. 12**), proliferation (**FIG. 12**), mutation burden (**FIG. 11**), tumor driver/suppressor mutation frequencies (**FIG. 14, FIG. 15**) and tumor purity (**FIG. 13**) were examined for differential expression. Differences in gene mutation distributions were evaluated using Fisher's exact test. Survival differences (**FIG. 9, FIG. 10A-B**) were assessed using stratified cox models and Kaplan Meier plots. **FIG. 20** presents a comparison of the subtyping of the TCGA-2017 bladder cancer dataset (n=408) samples based on the 5-type classification system described in Robertson, et al. (2017), and the 4-type classification system based on the 60 gene subtype described herein.

Results

Development of the 60 gene signature

[00202] The 60 gene signature gene list developed in this study is shown in **Table 1** and **FIG. 6**. Agreement of subtype calls using the 60 gene signature with the published 2708 gold standard gene signature subtype call in several different test datasets is shown in **FIG. 1** and

FIG. 16. The newly developed 60 gene signature demonstrated agreement of 0.87 in the TCGA-2014 dataset (**FIG. 1**) and 0.72 or 0.81 in the other 2 test datasets (**FIG. 16**).

Evaluation of assignment to luminal, luminal infiltrated, basal, or basal infiltrated/neuronal subtypes

[00203] Immune cell expression was significantly different across the subtypes in multiple datasets (T cells $p < 1e-05$ and CD274(PD-L1) $p < 0.002$) (**FIG. 17A-C, FIG. 18**). The luminal subtype showed lower immune expression for most markers. Drug target genes were differentially expressed (FGFR2, FGFR3, and ERBB2 $p < 1e-05$ in TCGA), as was proliferation ($p < 1e-20$ in TCGA), and patterns were reproducible across datasets, with luminal and luminal-infiltrated subtypes showing higher expression of FGFR2, FGFR3, and ERBB2 and lower proliferation (**FIG. 19**). Mutation frequencies of FGFR3 and RB1 varied across subtypes ($p = 1e-05$ and $p = 0.0005$ in TCGA), whereas mutation burden did not ($p = 0.16$), despite marked differences in immune infiltration. Significant differences in survival were observed ($p = 0.0385$ adjusting for stage in TCGA), with luminal and luminal-infiltrated showing better survival and basal worse survival (**FIG. 9, FIG. 10A-B**).

Conclusions

[00204] Development and validation of a 60-gene signature for bladder cancer subtyping was described. The resulting 60 gene signature maintains high concordance rates when applied to several independent test sets. Biologic gene expression subtypes of MIBC using a reduced 60-gene signature reveal key differences in prognosis, immune cell expression, and drug targets. Subtypes provide potential biomarkers for targeted and immunotherapy response. The data demonstrate that differences in prognosis that may be meaningful to therapeutic management.

Example 2- Development and Validation of the 56-Gene Bladder Cancer Subtyping Signature

[00205] Using a similar method as described above, a second, 56-gene subtyping signature was identified (**FIG. 4, FIG. 21**). A nearest centroid classifier (**FIG. 22, Table 2**) was fit using the 56 genes. Validation of the reduced gene signature was compared to the gold standard gene signature in TCGA datasets (**FIG. 21**).

Incorporation by reference

[00206] The following references are incorporated by reference in their entireties for all purposes.

[00207] Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421 (2012).

[00208] Dabney AR. ClaNC: Point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics.* 2006;22: 122–123. doi:10.1093/bioinformatics/bti756

[00209] Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A.* 2002;99: 6567–72. doi:10.1073/pnas.082099299

[00210] Martín M. et al., PAM50 proliferation score as a predictor of weekly paclitaxel benefit in breast cancer. *Breast Cancer Res Treat.* 138: 457-466 (2013).

[00211] Seiler R. et al., Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *Eur Urol.* 72(4):544-554 (2017).

[00212] Sjö Dahl G. et al., A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res.* 18(12):3377-86 (2012).

[00213] The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* volume 507, pages 315–322 (2014).

[00214] Robertson, AG, et al., *Cell*, 171(3): 540-556 (2017) Bindea G. et al., Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*, 39(4): 782-95 (2013).

[00215] Faruki H. et al., Lung Adenocarcinoma and Squamous Cell Carcinoma Gene Expression Subtypes Demonstrate Significant Differences in Tumor Immune Landscape. *JTO*, 12(6): 943-953 (2017).

[00216] Charoentong P. et al., Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell reports*, 18, 248-262 (2017).

[00217] Further Numbered Embodiments of the Disclosure

[00218] Other subject matter contemplated by the present disclosure is set out in the following numbered embodiments:

1. A method for determining a bladder cancer subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype.
2. The method of embodiment 1, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set(s) comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.
3. The method of embodiment 2, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.
4. The method of any one of embodiments 1-3, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.
5. The method of embodiment 4, wherein the nucleic acid level is RNA or cDNA.

6. The method embodiment 4 or 5, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
7. The method of embodiment 6, wherein the expression level is detected by performing RNAseq.
8. The method of embodiment 7, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.
9. The method of any one of embodiments 1-8, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, a fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
10. The method of embodiment 9, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
11. The method of any one embodiments 1-10, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.
12. The method of embodiment 11, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1.
13. The method of embodiment 11, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

14. The method of any one of embodiments 1-13, wherein the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.

15. A method for determining a muscle invasive bladder cancer (MIBC) subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype.

16. The method of embodiment 15, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.

17. The method of embodiment 16, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.

18. The method of any one of embodiments 15-17, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.

19. The method of embodiment 18, wherein the nucleic acid level is RNA or cDNA.

20. The method of embodiment 18 or 19, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

21. The method of embodiment 20, wherein the expression level is detected by performing RNAseq.

22. The method of embodiment 21, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.

23. The method of any one of embodiments 15-22, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.

24. The method of embodiment 23, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.

25. The method of any one of embodiments 15-24, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.

26. The method of embodiment 25, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarkers of Table 1.

27. The method of embodiment 25, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30

classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

28. The method of any one of embodiments 15-27, wherein the at least one classifier biomarker comprises, consists essentially of or consist of all the classifier biomarkers of Table 1 or Table 2.

29. A method of detecting a biomarker in a bladder cancer sample obtained from a patient, the method comprising measuring the expression level of a plurality of classifier biomarker nucleic acids selected from Table 1 or Table 2 using an amplification, hybridization and/or sequencing assay.

30. The method of embodiment 29, wherein the bladder cancer sample was previously diagnosed as being muscle invasive bladder cancer (MIBC).

31. The method of embodiment 30, wherein the previous diagnosis was by histological examination.

32. The method of any one of embodiments 29-31, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

33. The method of embodiment 32, wherein the expression level is detected by performing RNAseq.

34. The method of embodiment 31, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from Table 1 or Table 2.

35. The method of any one of embodiments 29-34, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.

36. The method of embodiment 35, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.

37. The method of any one of embodiments 29-36, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1.

38. The method of any one of embodiments 29-36, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

39. The method of any one of embodiments 29-36, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2.

40. A method of treating muscle invasive bladder cancer (MIBC) in a subject, the method comprising:

measuring the expression level of at least one biomarker nucleic acid in a MIBC sample obtained from the subject, wherein the at least one biomarker nucleic acid is selected from a set of biomarkers listed in Table 1 or Table 2, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the MIBC; and administering a therapeutic agent based on the subtype of the MIBC.

41. The method of embodiment 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2 classifier

biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1.

42. The method of embodiment 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

43. The method of embodiment 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2.

44. The method of any one of embodiments 40-43, further comprising measuring the expression of at least one biomarker from an additional set of biomarkers.

45. The method of embodiment 44, wherein the additional set of biomarkers comprises BIRC5, CCNB1, CDC20, CDCA1, CEP55, KNTC2, MKI67, PTTG1, RRM2, TYMS, UBE2C, TP53, RB1, FGFR2, FGFR3, and ERBB2.

46. The method of embodiment 44, wherein the additional set of biomarkers comprise genes selected from Gene Expression Omnibus Dataset GSE87304, Gene Expression Omnibus Dataset GSE32894 or a combination thereof.

47. The method of embodiment 44, wherein the additional set of biomarkers comprises at least an immune cell signature, a cell proliferation signature, or drug target genes.

48. The method of any one of embodiments 40-47, wherein the measuring the expression level is conducted using an amplification, hybridization and/or sequencing assay.

49. The method of embodiment 48, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene

Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

50. The method of embodiment 49, wherein the expression level is detected by performing qRT-PCR.

51. The method of any one of embodiments 40-50, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.

52. The method of embodiment 51, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.

53. The method of any one of embodiments 40-52, wherein the subject's MIBC subtype is selected from luminal, luminal infiltrated, basal, and basal infiltrated/neuronal.

54. The method of embodiment 40, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises at least one biomarker nucleic acid listed in Table 1 or Table 2 in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the MIBC.

55. The method of embodiment 40, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises all of the biomarker nucleic acids listed in Table 1 or Table 2, all of the biomarker nucleic acids listed in embodiment 45 or embodiment 46, in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the bladder cancer.

56. The method of embodiment 54 or 55, wherein the publically available bladder cancer dataset is TCGA bladder cancer RNAseq dataset.

57. A method of predicting overall survival in a muscularly invasive bladder cancer (MIBC) patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2 in a MIBC sample obtained from a patient, wherein the detection of the expression level of the at least one classifier biomarker specifically identifies a luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtype, and wherein identification of the subtype is predictive of the overall survival in the patient.

58. The method of embodiment 57, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarkers of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.

59. The method of embodiment 58, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.

60. The method of any one of embodiments 57-59, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.

61. The method of embodiment 60, wherein the nucleic acid level is RNA or cDNA.

62. The method of any one of embodiments 57-61, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction

(qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

63. The method of embodiment 62, wherein the expression level is detected by performing qRT-PCR.

64. The method of embodiment 57, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.

65. The method of any one of embodiments 57-64, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.

66. The method of embodiment 65, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.

67. The method of any one of embodiments 57-66, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.

68. The method of embodiment 67, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1.

69. The method of embodiment 67, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

70. The method of any of embodiments 57-69, wherein the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.

* * * * *

[00219] The various embodiments described above can be combined to provide further embodiments. All of the U.S. patents, U.S. patent application publications, U.S. patent application, foreign patents, foreign patent application and non-patent publications referred to in this specification and/or listed in the Application Data Sheet are incorporated herein by reference, in their entirety. Aspects of the embodiments can be modified, if necessary to employ concepts of the various patents, application and publications to provide yet further embodiments.

[00220] These and other changes can be made to the embodiments in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

What is claimed is:

1. A method for determining a bladder cancer subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype.
2. The method of claim 1, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set(s) comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.
3. The method of claim 2, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.
4. The method of any one of claims 1-3, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.
5. The method of claim 4, wherein the nucleic acid level is RNA or cDNA.

6. The method claim 4, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
7. The method of claim 6, wherein the expression level is detected by performing RNAseq.
8. The method of claim 7, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.
9. The method of claim 1, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, a fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
10. The method of claim 9, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
11. The method of claim 1, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.
12. The method of claim 11, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1.
13. The method of claim 11, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier

biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.

14. The method of claim 1, wherein the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.
15. A method for determining a muscle invasive bladder cancer (MIBC) subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype.
16. The method of claim 15, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.
17. The method of claim 16, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.
18. The method of any one of claims 15-17, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.

19. The method of claim 18, wherein the nucleic acid level is RNA or cDNA.
20. The method of claim 18, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
21. The method of claim 20, wherein the expression level is detected by performing RNAseq.
22. The method of claim 21, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.
23. The method of claim 15, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
24. The method of claim 23, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
25. The method of claim 15, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.
26. The method of claim 25, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarkers of Table 1.

27. The method of claim 25, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.
28. The method of claim 15, wherein the at least one classifier biomarker comprises, consists essentially of or consist of all the classifier biomarkers of Table 1 or Table 2.
29. A method of detecting a biomarker in a bladder cancer sample obtained from a patient, the method comprising measuring the expression level of a plurality of classifier biomarker nucleic acids selected from Table 1 or Table 2 using an amplification, hybridization and/or sequencing assay.
30. The method of claim 29, wherein the bladder cancer sample was previously diagnosed as being muscle invasive bladder cancer (MIBC).
31. The method of claim 30, wherein the previous diagnosis was by histological examination.
32. The method of any one of claims 29-31, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
33. The method of claim 32, wherein the expression level is detected by performing RNAseq.
34. The method of claim 31, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from Table 1 or Table 2.

35. The method of claim 29, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
36. The method of claim 35, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
37. The method of claim 29, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1.
38. The method of claim 29, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.
39. The method of claim 29, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2.
40. A method of treating muscle invasive bladder cancer (MIBC) in a subject, the method comprising:
measuring the expression level of at least one biomarker nucleic acid in a MIBC sample obtained from the subject, wherein the at least one biomarker nucleic acid is selected from a set of biomarkers listed in Table 1 or Table 2, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the MIBC;
and administering a therapeutic agent based on the subtype of the MIBC.
41. The method of claim 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2

classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers or at least 60 classifier biomarker nucleic acids of Table 1.

42. The method of claim 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.
43. The method of claim 40, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of Table 1 or Table 2.
44. The method of any one of claims 40-43, further comprising measuring the expression of at least one biomarker from an additional set of biomarkers.
45. The method of claim 44, wherein the additional set of biomarkers comprises BIRC5, CCNB1, CDC20, CDCA1, CEP55, KNTC2, MKI67, PTTG1, RRM2, TYMS, UBE2C, TP53, RB1, FGFR2, FGFR3, and ERBB2.
46. The method of claim 44, wherein the additional set of biomarkers comprise genes selected from Gene Expression Omnibus Dataset GSE87304, Gene Expression Omnibus Dataset GSE32894 or a combination thereof.
47. The method of claim 44, wherein the additional set of biomarkers comprises at least an immune cell signature, a cell proliferation signature, or drug target genes.
48. The method of claim 40, wherein the measuring the expression level is conducted using an amplification, hybridization and/or sequencing assay.

49. The method of claim 48, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
50. The method of claim 49, wherein the expression level is detected by performing qRT-PCR.
51. The method of claim 40, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
52. The method of claim 51, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
53. The method of claim 40, wherein the subject's MIBC subtype is selected from luminal, luminal infiltrated, basal, and basal infiltrated/neuronal.
54. The method of claim 40, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises at least one biomarker nucleic acid listed in Table 1 or Table 2 in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the MIBC.
55. The method of claim 40, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises all of the biomarker nucleic acids listed in Table 1 or Table 2, all of the biomarker nucleic acids listed in claim 45 or claim 46, in combination with one or more biomarker nucleic acids from a publically available bladder cancer dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the bladder cancer.

56. The method of claim 54 or 55, wherein the publically available bladder cancer dataset is TCGA bladder cancer RNAseq dataset.
57. A method of predicting overall survival in a muscularly invasive bladder cancer (MIBC) patient, the method comprising detecting an expression level of at least one classifier biomarker of Table 1 or Table 2 in a MIBC sample obtained from a patient, wherein the detection of the expression level of the at least one classifier biomarker specifically identifies a luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtype, and wherein identification of the subtype is predictive of the overall survival in the patient.
58. The method of claim 57, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarker of Table 1 or Table 2 to the expression of the at least one classifier biomarker of Table 1 or Table 2 in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer luminal infiltrated sample, expression data of the at least one classifier biomarkers of Table 1 or Table 2 from a reference bladder cancer basal sample, expression data of the at least one classifier biomarker of Table 1 or Table 2 from a reference bladder cancer basal infiltrated/neuronal sample or a combination thereof; and classifying the sample as luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the comparing step.
59. The method of claim 58, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a luminal, luminal infiltrated, basal or basal infiltrated/neuronal subtype based on the results of the statistical algorithm.
60. The method of any one of claims 57-59, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.

61. The method of claim 60, wherein the nucleic acid level is RNA or cDNA.
62. The method of claim 57, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
63. The method of claim 62, wherein the expression level is detected by performing qRT-PCR.
64. The method of claim 57, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of Table 1 or Table 2.
65. The method of claim 57, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
66. The method of claim 65, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
67. The method of claim 57, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.
68. The method of claim 67, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 60 classifier biomarkers of Table 1.

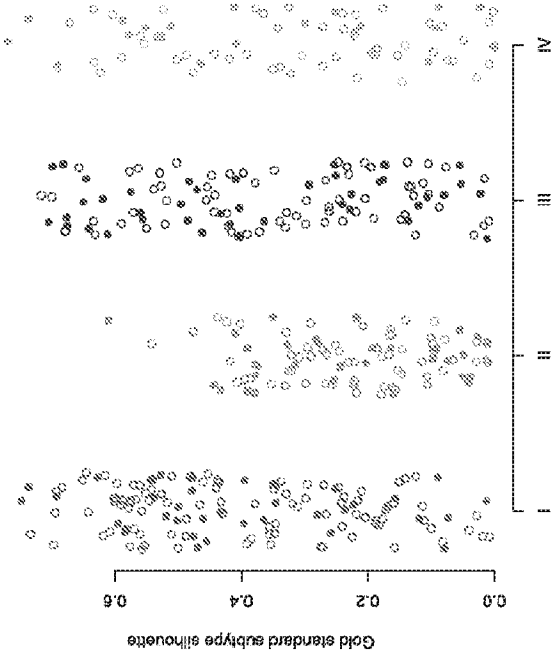
69. The method of claim 67, wherein the plurality of classifier biomarkers comprises, consists essentially of or consists of at least 2 classifier biomarkers, at least 5 classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, or at least 56 classifier biomarkers of Table 2.
70. The method of claim 57, wherein the at least one classifier biomarker comprises, consists essentially of or consists of all the classifier biomarkers of Table 1 or Table 2.

FIG. 1 Gold standard (GS) subtype distribution (n=408)

I	II	III	IV	Sum
132	95	110	71	408

GS sil vs GS subtype (n=408), solid when sample is in 2014 data used to fit gold standard subtyper

TCGA n=408
solid dot Indicates sample used to fit gold standard subtyper

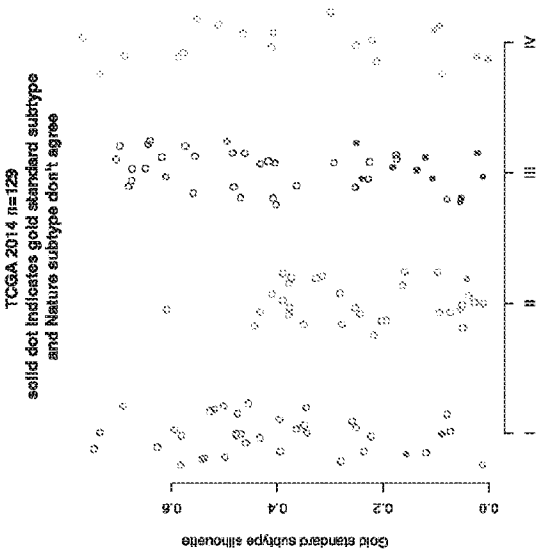


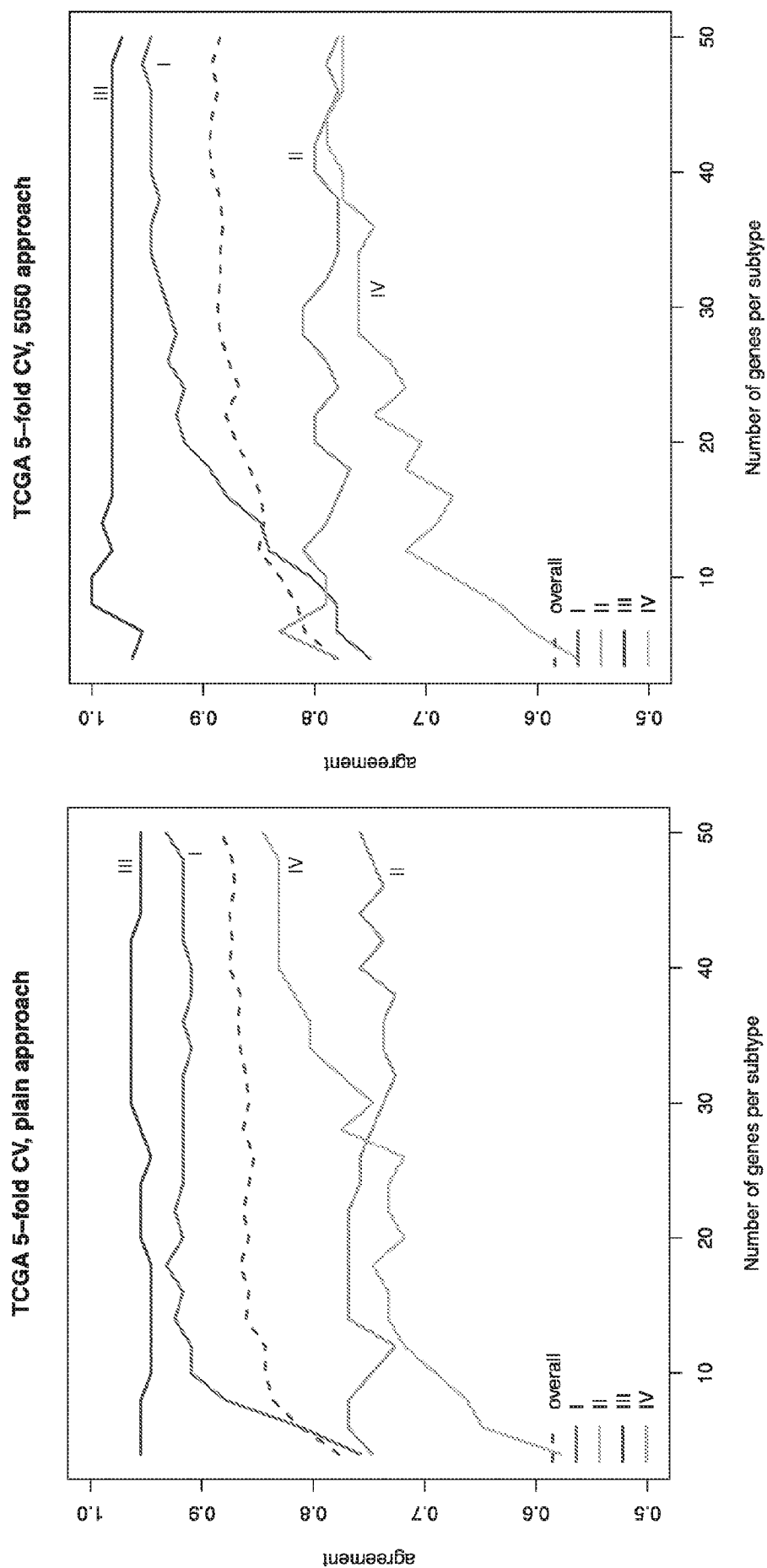
Nature 2014 cluster assignment (rows) vs gold standard subtype (columns), (n=129) agreement:

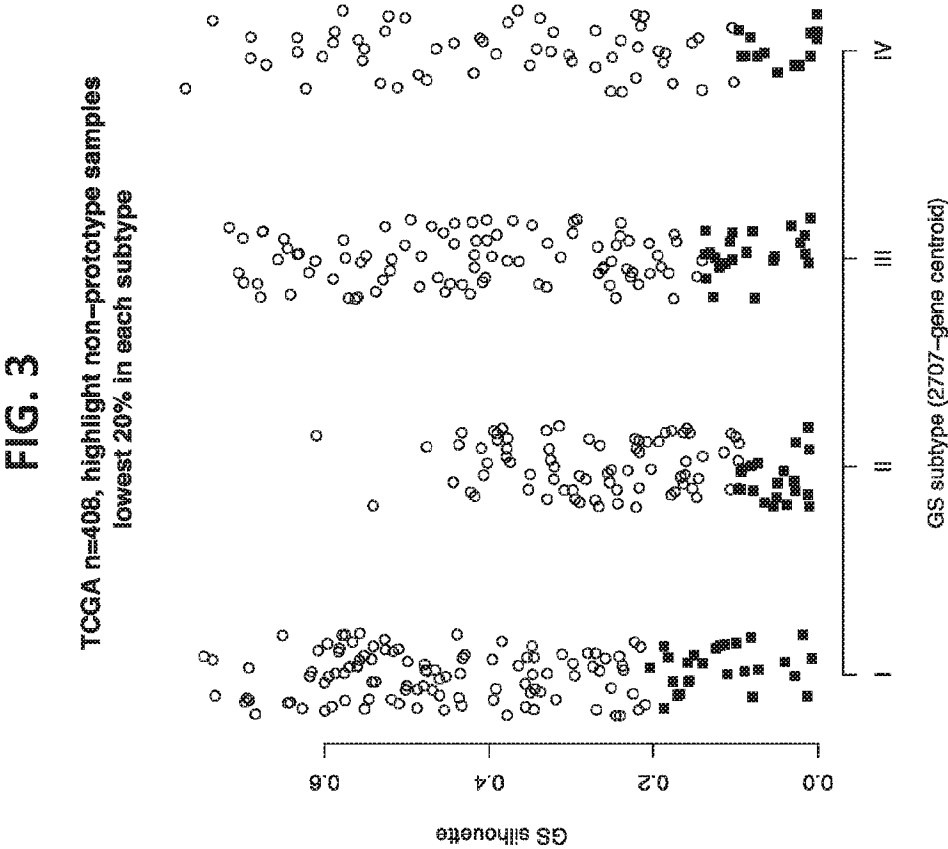
	I	II	III	IV	Sum
I	35	1	5	0	41
II	2	31	5	4	42
III	0	0	31	0	31
IV	0	0	0	15	15
Sum	37	32	41	19	129

0.87

GS sil vs GS subtype (n=129), solid when doesn't match Nature 2014 cluster assignment







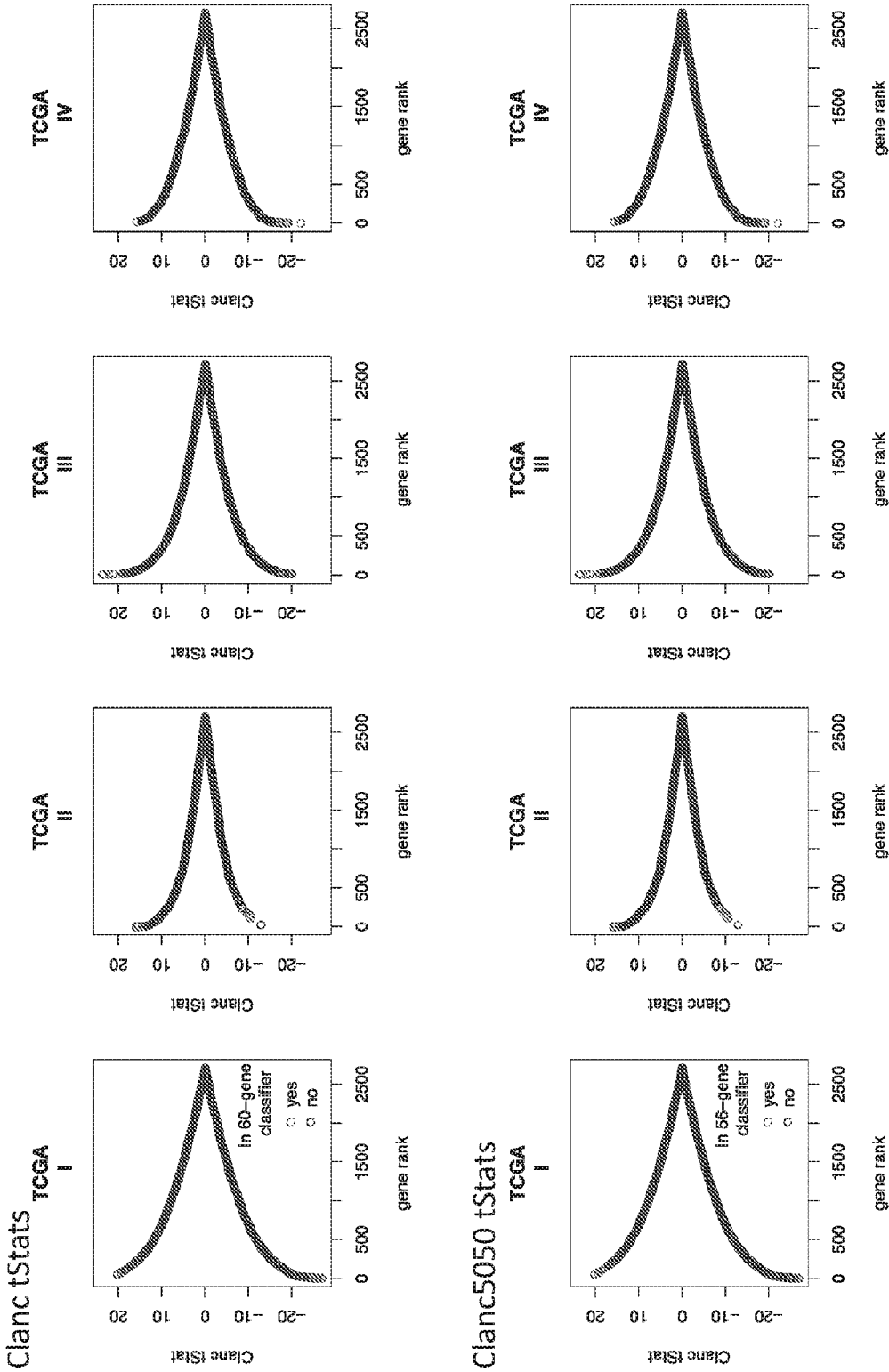


FIG. 4

TCGA 2017 n=408
Solid dot indicates gold standard
subtype and clanc subtype don't agree

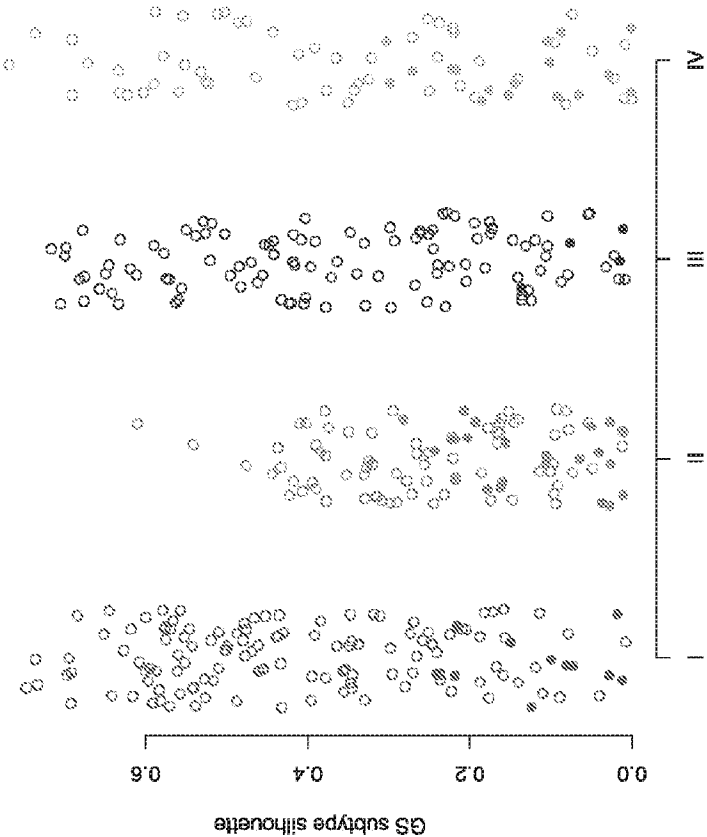


FIG. 5
Gold Standard subtype (rows) vs. clanc (60) subtype (columns), agreement: 0.87

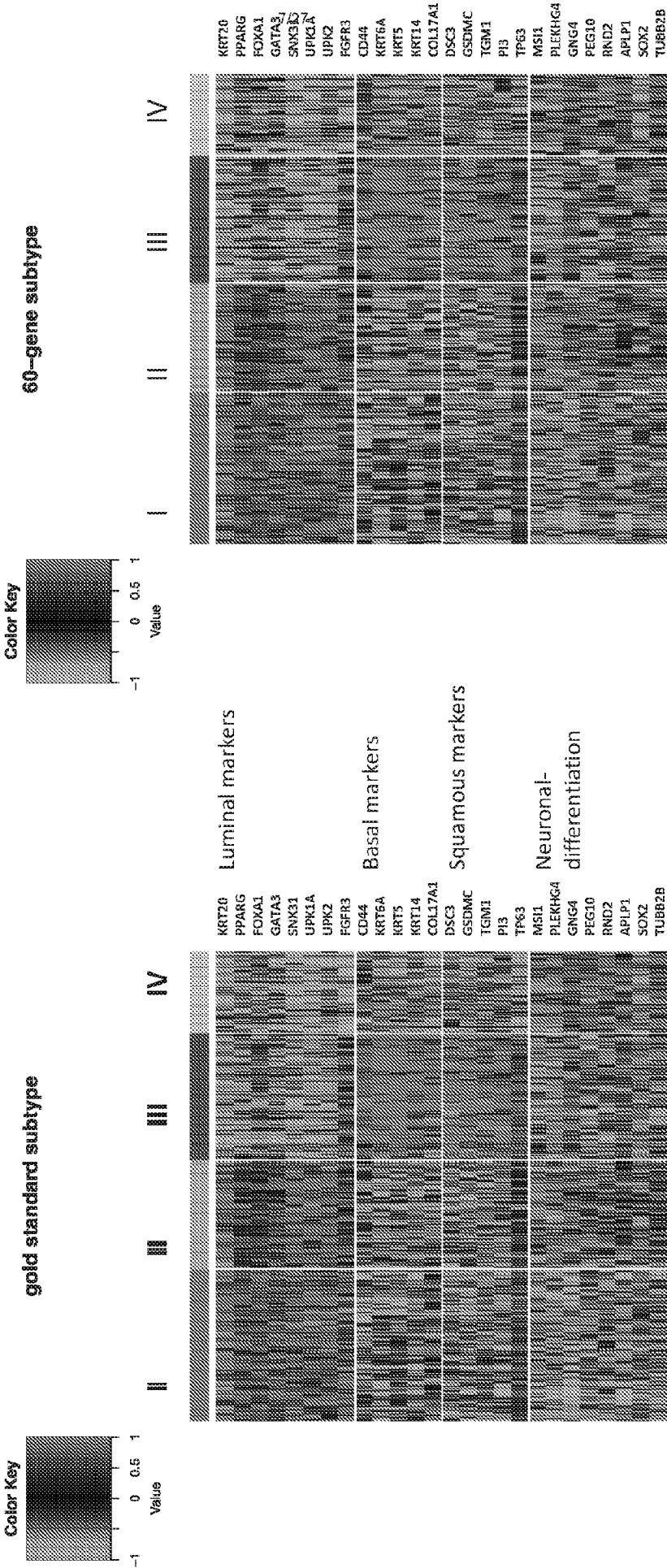
	I	II	III	IV	Sum
I	121	8	3	0	132
II	10	72	5	8	95
III	1	0	106	3	110
IV	0	6	11	54	71
Sum	132	86	125	65	408

6/27

	I	II	III	IV		I	II	III	IV	
ALDH1L2	-2.95	0.51	0.59	1.31		LMOD1	-1.31	2.77	-0.22	0.96
ANXA6	-2.00	0.64	0.09	1.66		LOC100188947	3.20	2.48	-3.88	-4.38
ARSI	-3.67	0.43	1.72	1.33		MRV11	-1.19	1.05	-0.15	0.54
BCAS1	1.88	1.10	-3.83	-5.58		MRP2	-2.34	0.40	-0.62	1.43
BNC1	-1.48	-0.57	7.97	2.24		PDLM3	-2.49	1.95	0.33	1.51
C10orf99	1.69	1.15	-1.26	-7.80		PLAZG4F	1.13	0.24	-0.66	-4.12
C17orf28	0.80	0.63	-3.57	-2.18		PODN	-1.54	1.82	-0.72	1.20
CAPN5	1.17	1.17	-2.49	-1.86		POSTN	-4.38	1.37	0.46	1.42
CCDC80	-3.08	1.14	0.43	1.87		PRRX1	-3.33	0.94	0.50	2.06
COL6A2	-2.25	1.10	0.36	2.02		PVRL4	0.49	0.36	-0.39	-2.97
CPXM2	-2.14	1.78	-0.13	1.70		RAPGEFL1	1.01	0.10	-0.43	-3.66
CTHRC1	-3.30	0.61	0.64	1.15		RHOU	0.59	1.22	-2.78	-1.70
DSG3	-2.16	-1.59	7.87	-1.24		RHOV	-0.17	0.20	0.89	-2.93
EMILIN1	-1.96	1.86	-0.20	1.84		SCUBE2	1.29	3.00	-2.51	-1.66
EPN3	0.61	0.45	-0.73	-2.60		SDC1	0.39	0.13	-0.01	-2.41
EVPL	0.51	0.48	-0.65	-2.53		SERPINB13	-0.64	-2.10	5.68	-2.94
FAP	-3.87	0.94	0.84	1.46		SFRP2	-6.85	2.50	0.88	1.68
FBN1	-2.47	1.02	0.22	1.82		SFRP4	-5.67	3.09	0.39	2.19
FGF7	-2.23	1.98	0.12	1.44		SLC30A2	2.17	3.26	-4.60	-3.18
FMO9P	2.37	3.22	-3.99	-5.26		SMOC2	-1.41	2.14	-0.44	1.20
FNDC1	-4.02	2.05	0.42	2.44		SNX31	1.40	2.06	-6.48	-7.24
GABBR2	0.63	5.25	-2.87	-0.72		SPRR2A	-1.49	-0.32	5.64	-1.95
GFPT2	-4.37	0.83	0.74	1.95		SSC5D	-2.38	1.82	0.09	1.78
GGT6	1.18	0.33	-2.23	-5.61		TBX3	1.48	0.55	-3.24	-2.08
GREM1	-5.82	1.63	0.66	0.80		TLE2	0.96	0.69	-2.92	-1.06
GRHL3	1.02	1.46	-1.39	-6.73		TOX3	2.44	2.08	-6.52	-5.74
IL2ORB	-0.89	-0.97	4.28	-0.59		UPK1A	1.94	2.82	-5.92	-6.97
KRT6A	-2.15	-2.48	7.61	-0.39		UPK2	1.85	2.45	-5.94	-6.04
KRT6B	-1.53	-2.24	7.55	-0.39		UPK3A	1.98	3.61	-5.63	-4.89
KRT6C	-1.76	-2.57	7.25	-1.05		ZN7F50	-0.07	0.52	0.82	-3.97

FIG. 6

FIG. 7



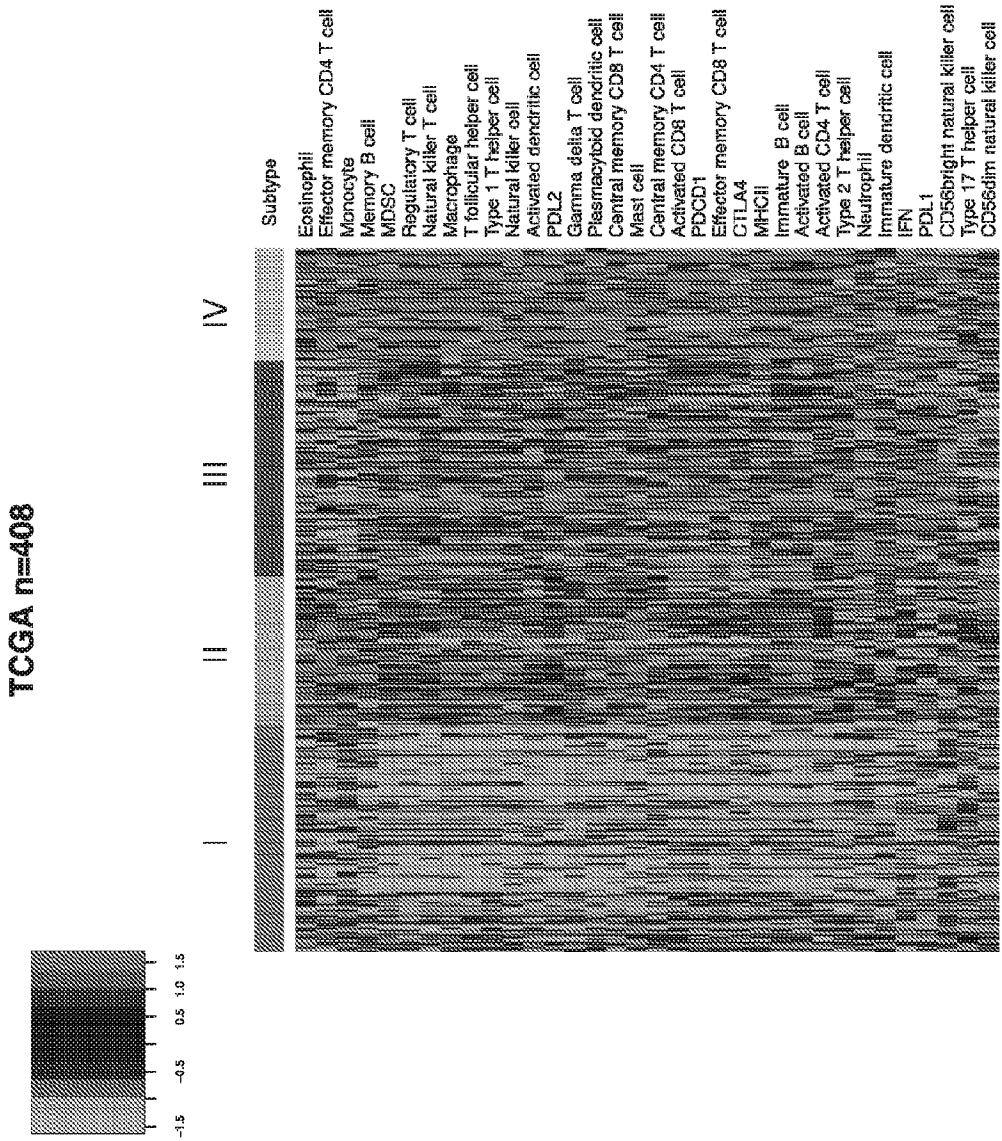


FIG. 8

9/27

FIG. 9

OS by stage n = 402

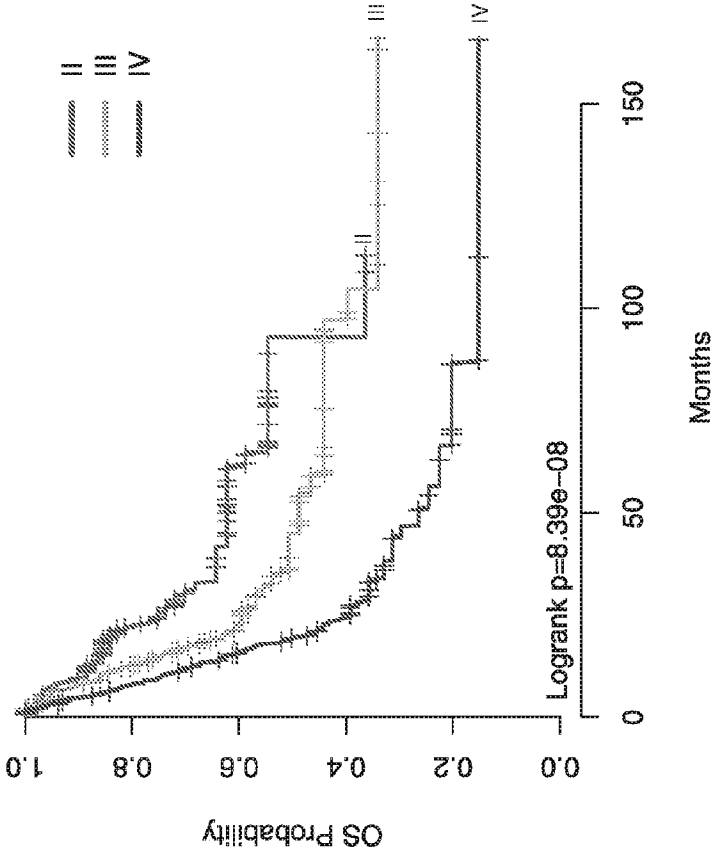
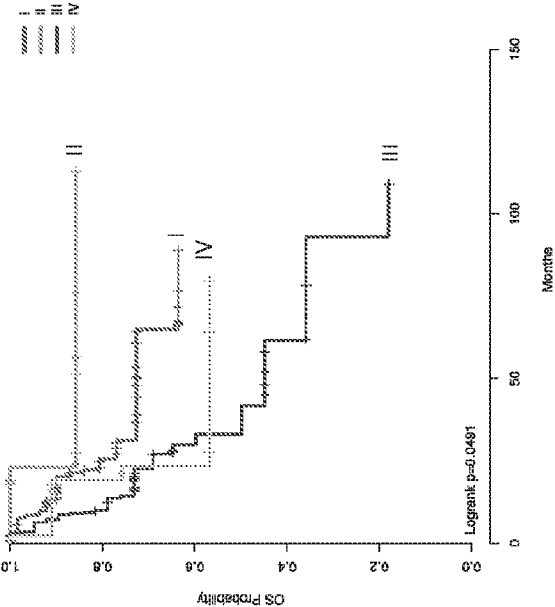
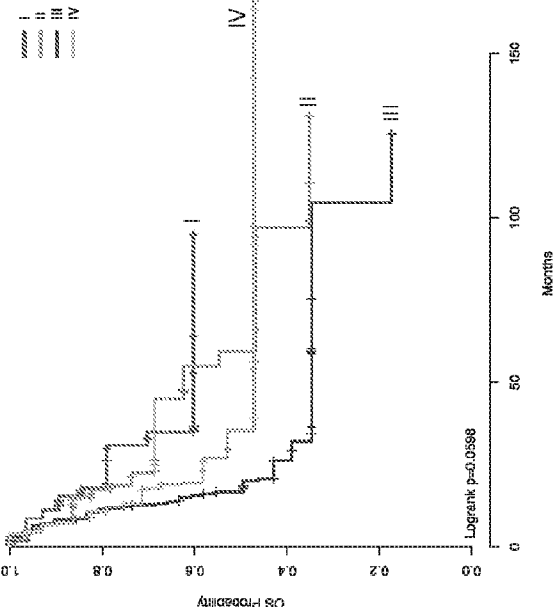


FIG. 10A

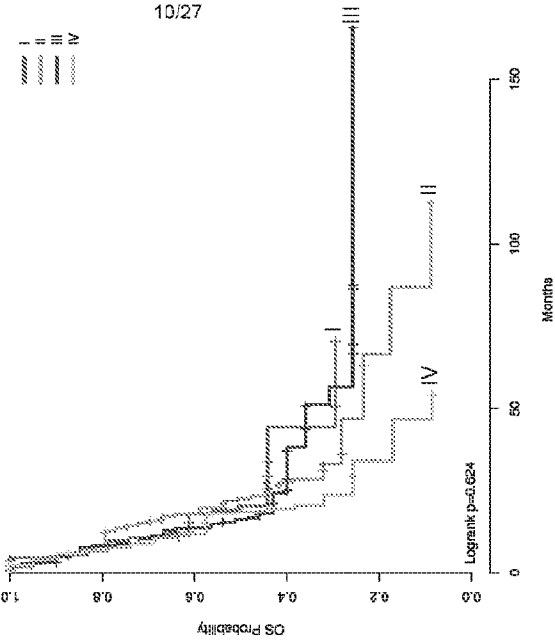
OS by subtype in stage II n = 128



OS by subtype in stage III n = 140



OS by subtype in stage IV n = 133



10/27

FIG. 10B

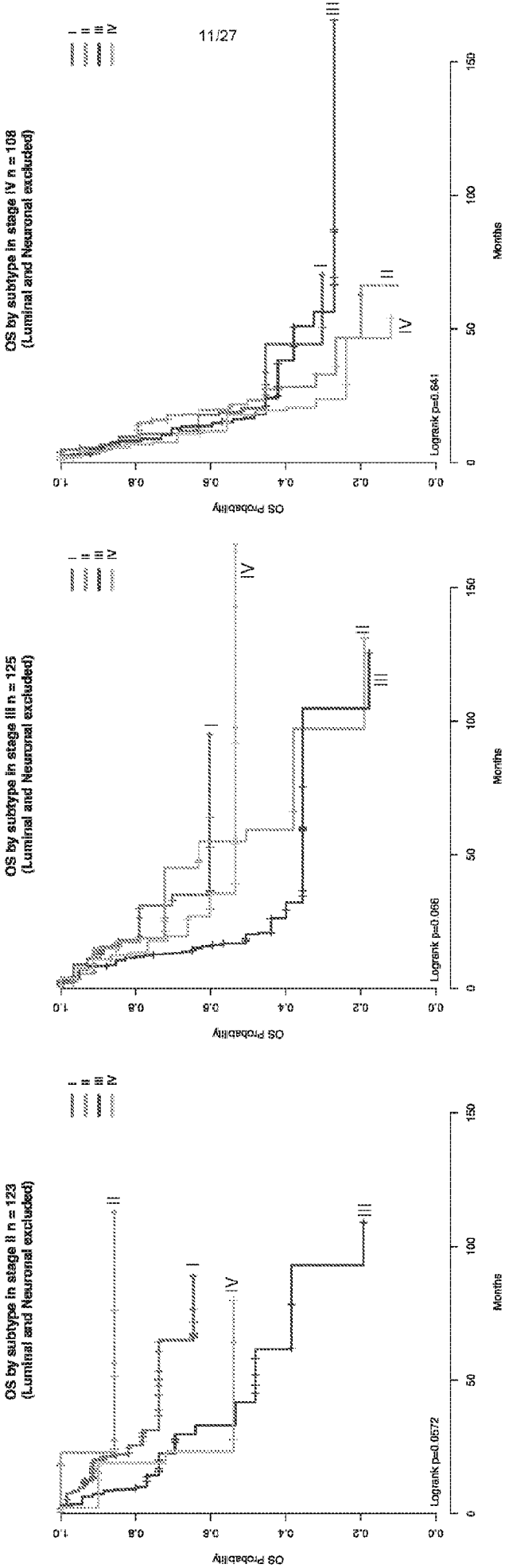
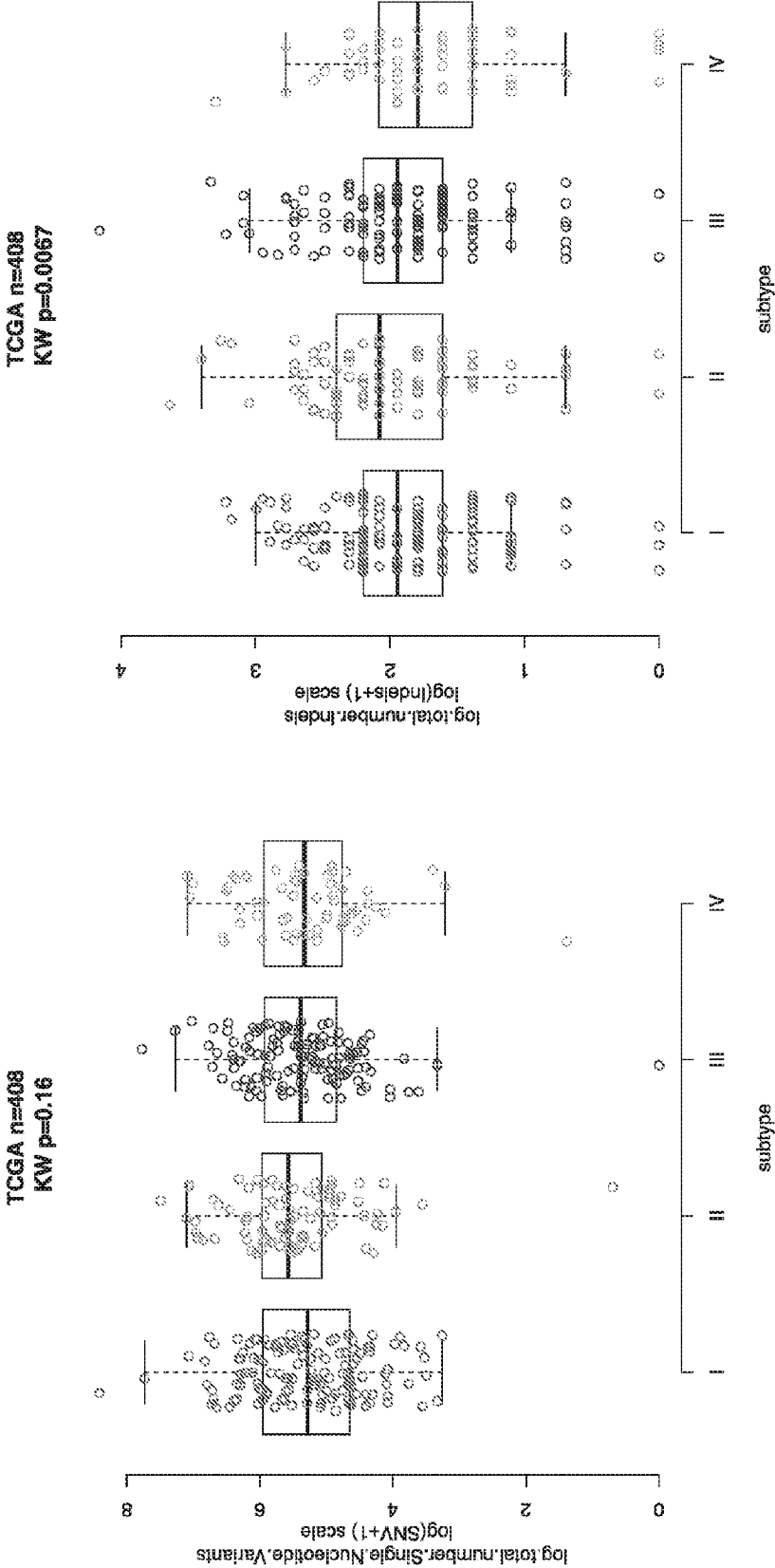


FIG. 11



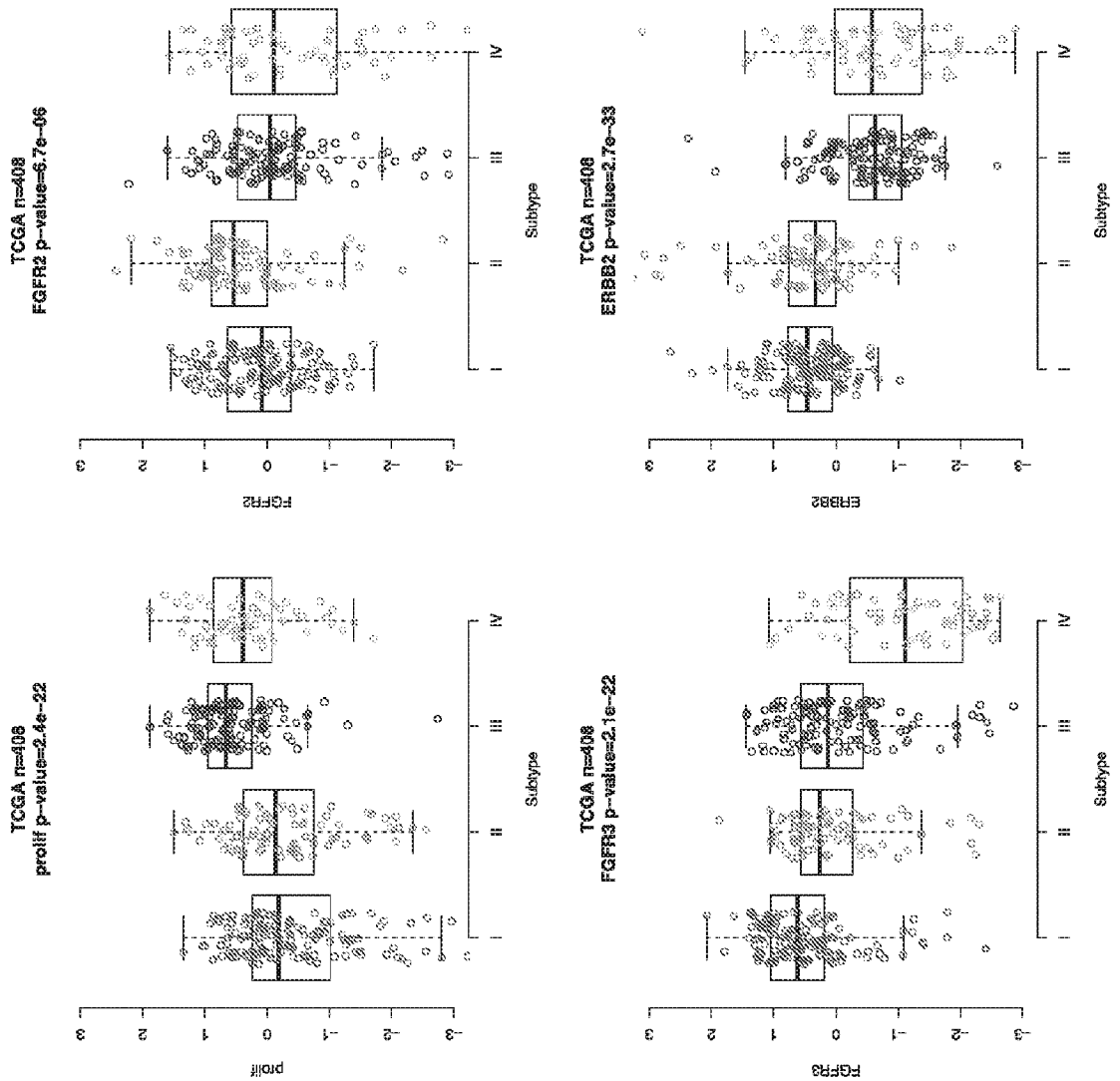


FIG. 12

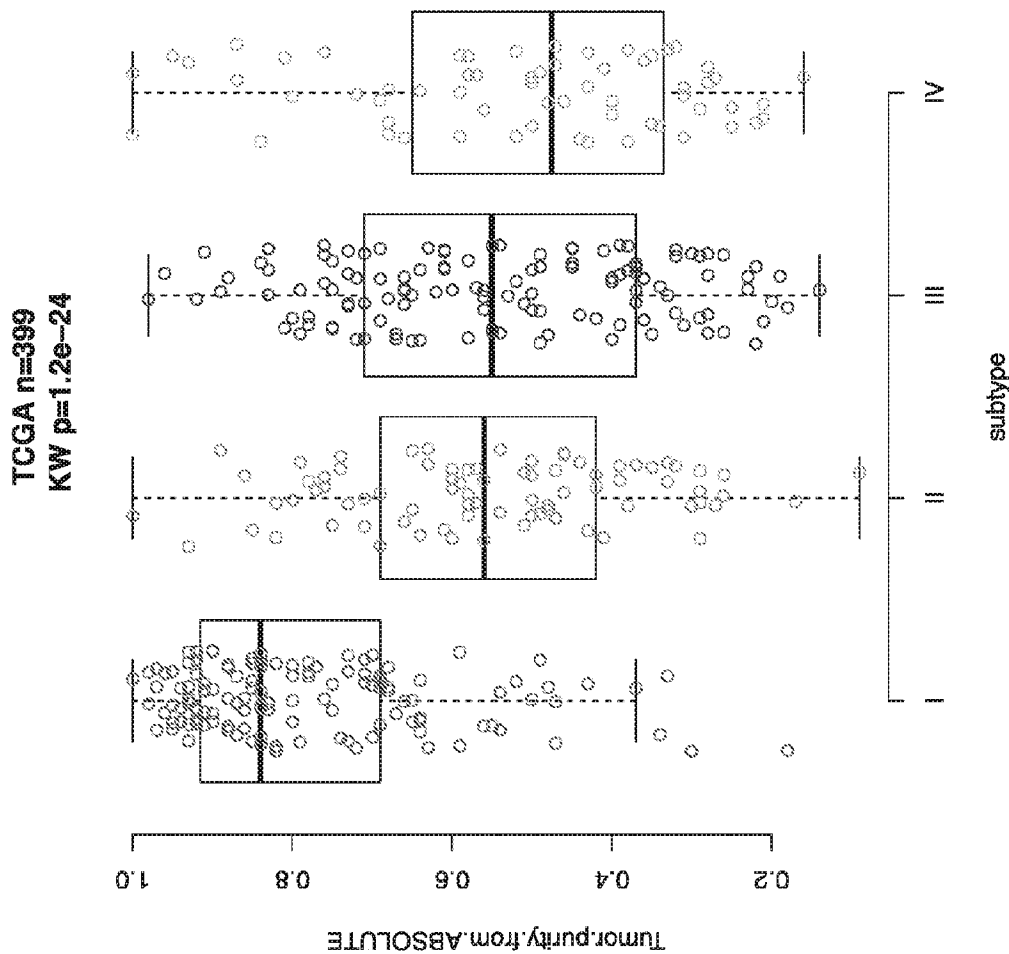


FIG. 13

FIG. 14

subtype {rows} vs mutation in RB1 (columns)			
	no	yes	Sum
I	122(0.92)	10(0.08)	132(1)
II	72(0.84)	14(0.16)	86(1)
III	98(0.78)	27(0.22)	125(1)
IV	46(0.71)	19(0.29)	65(1)
Sum	338(0.83)	70(0.17)	408(1)
Fisher test p-value: 0.0005			

subtype {rows} vs mutation in FGFR3 (columns)			
	no	yes	Sum
I	93(0.7)	39(0.3)	132(1)
II	81(0.94)	5(0.06)	86(1)
III	114(0.91)	11(0.09)	125(1)
IV	63(0.97)	2(0.03)	65(1)
Sum	351(0.86)	57(0.14)	408(1)
Fisher test p-value: 1e-05			

subtype {rows} vs mutation in ERBB2 (columns)			
	no	yes	Sum
I	116(0.88)	16(0.12)	132(1)
II	80(0.93)	6(0.07)	86(1)
III	107(0.86)	18(0.14)	125(1)
IV	56(0.86)	9(0.14)	65(1)
Sum	359(0.88)	49(0.12)	408(1)
Fisher test p-value: 0.38			

subtype {rows} vs focal deletion in RB1 (columns)			
	Del	Loss	Sum
I	106(0.81)	17(0.13)	131(1)
II	75(0.88)	7(0.08)	85(1)
III	105(0.85)	6(0.05)	123(1)
IV	54(0.83)	10(0.15)	65(1)
Sum	340(0.84)	40(0.11)	404(1)
Fisher test p-value: 0.6			

subtype {rows} vs focal amplification in FGFR3 (columns)			
	Amp	Gain	Sum
I	121(0.92)	5(0.04)	131(1)
II	85(1)	0(0)	85(1)
III	123(1)	0(0)	123(1)
IV	65(1)	0(0)	65(1)
Sum	394(0.98)	5(0.01)	404(1)
Fisher test p-value: 0.0016			

subtype {rows} vs focal amplification in ERBB2 (columns)			
	Amp	Gain	Sum
I	121(0.92)	4(0.03)	131(1)
II	79(0.93)	6(0.07)	85(1)
III	120(0.98)	3(0.02)	123(1)
IV	63(0.97)	1(0.02)	65(1)
Sum	383(0.95)	15(0.04)	404(1)
Fisher test p-value: 0.18			

FIG. 15

subtype (rows) vs mutation.in.TP53 (columns)			
	no	yes	Sum
I	91(0.69)	41(0.31)	132(1)
II	37(0.43)	49(0.57)	86(1)
III	55(0.44)	70(0.56)	125(1)
IV	29(0.45)	36(0.55)	65(1)
Sum	212(0.52)	196(0.48)	408(1)
Fisher test p-value: 4e-05			

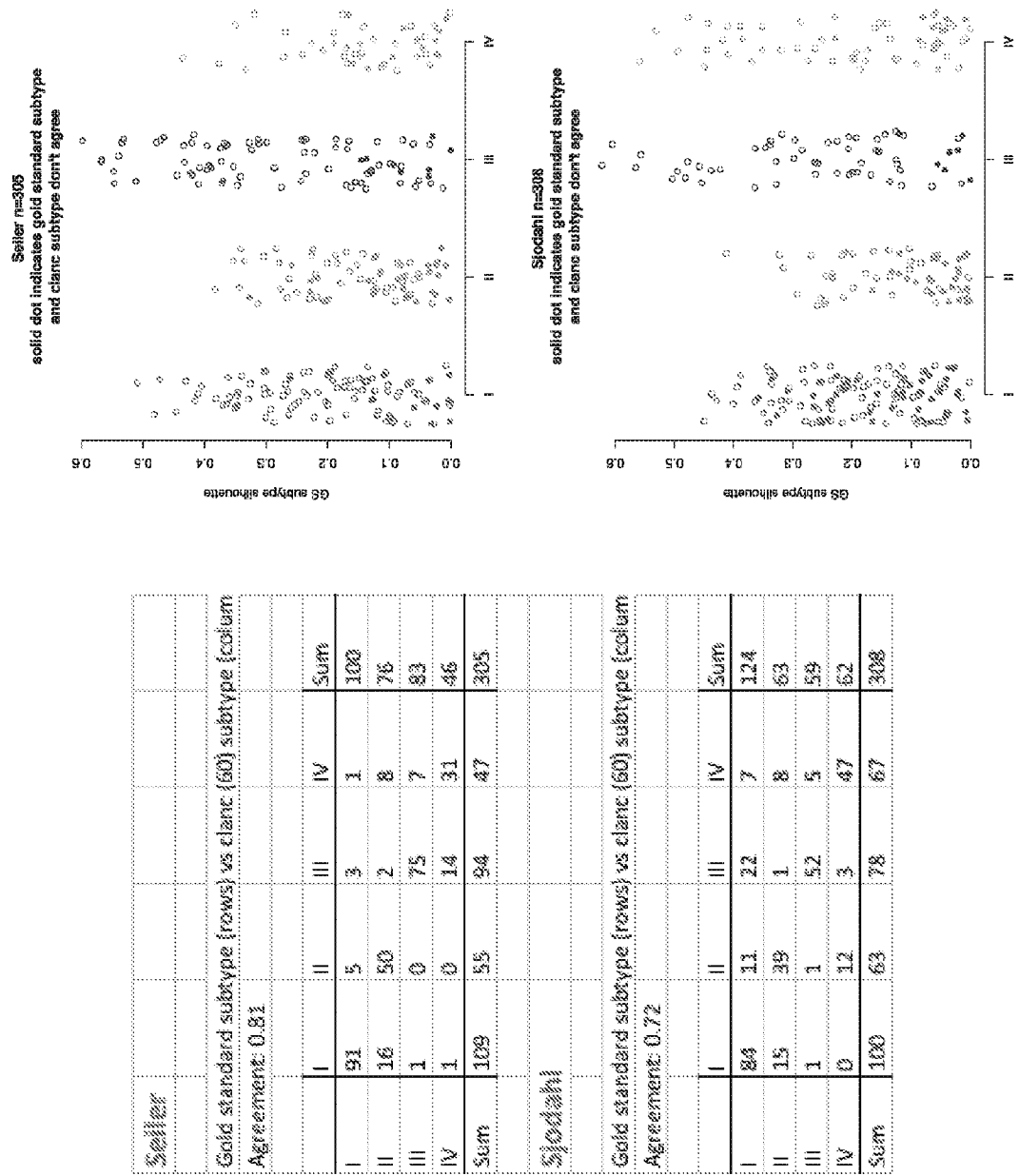


FIG. 16

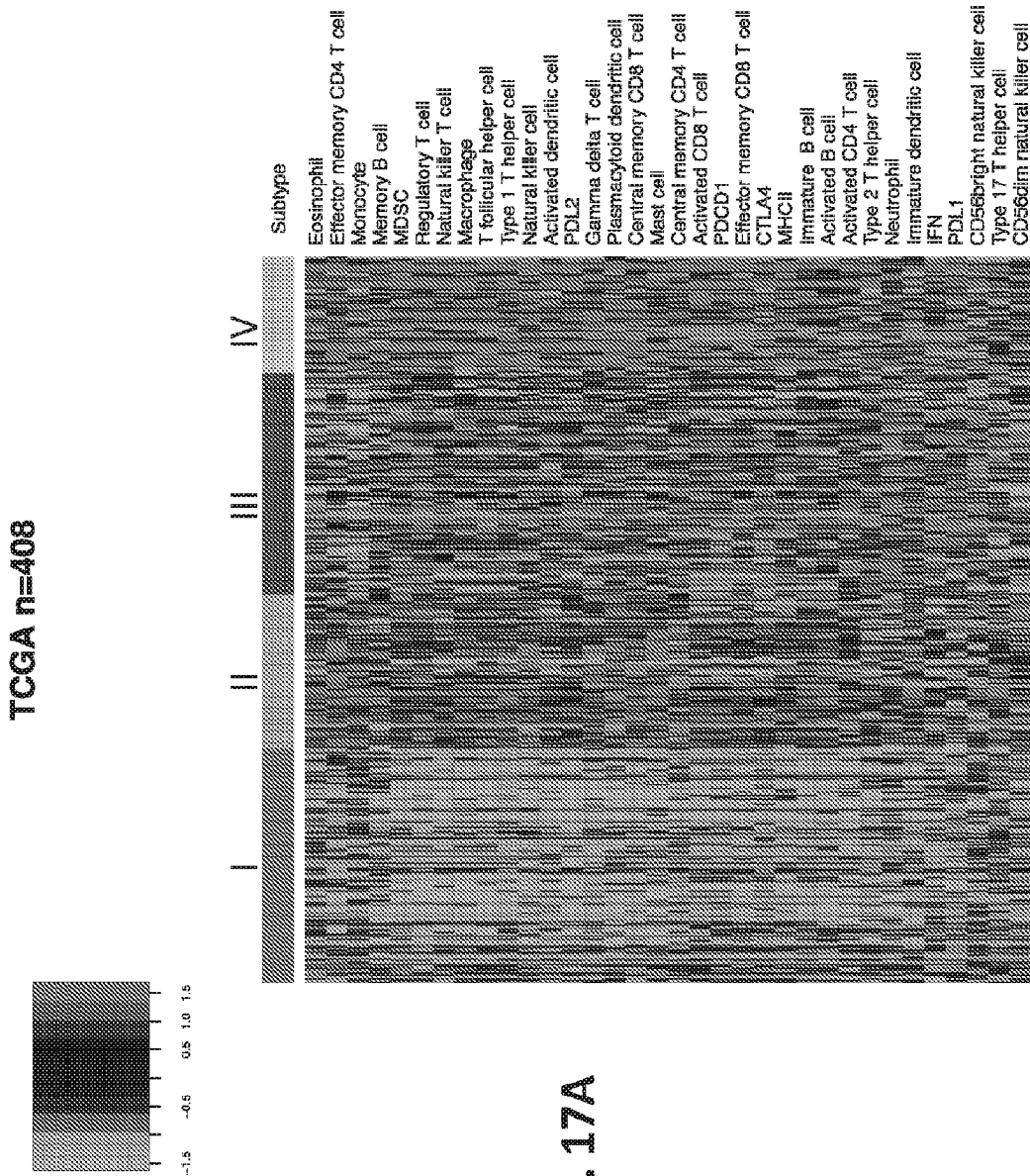


FIG. 17A

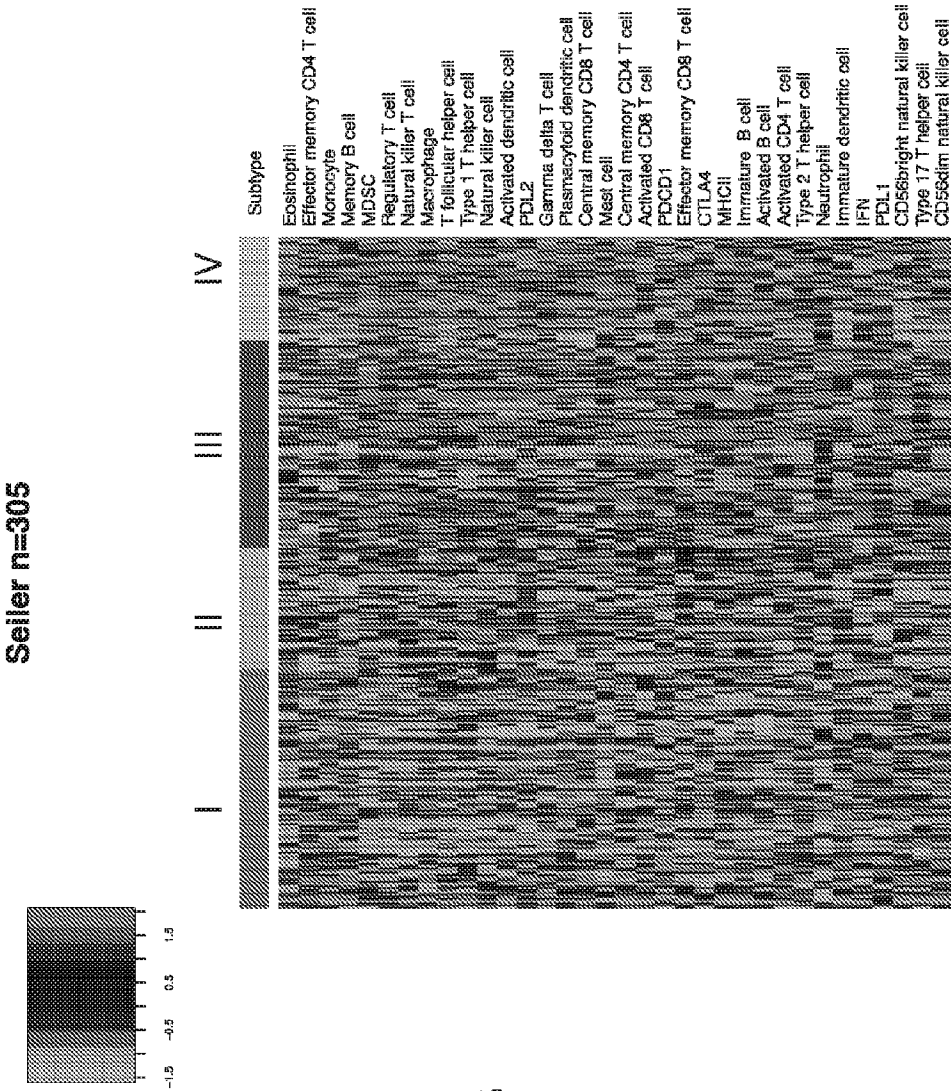


FIG. 17B

Sjodahl n=308

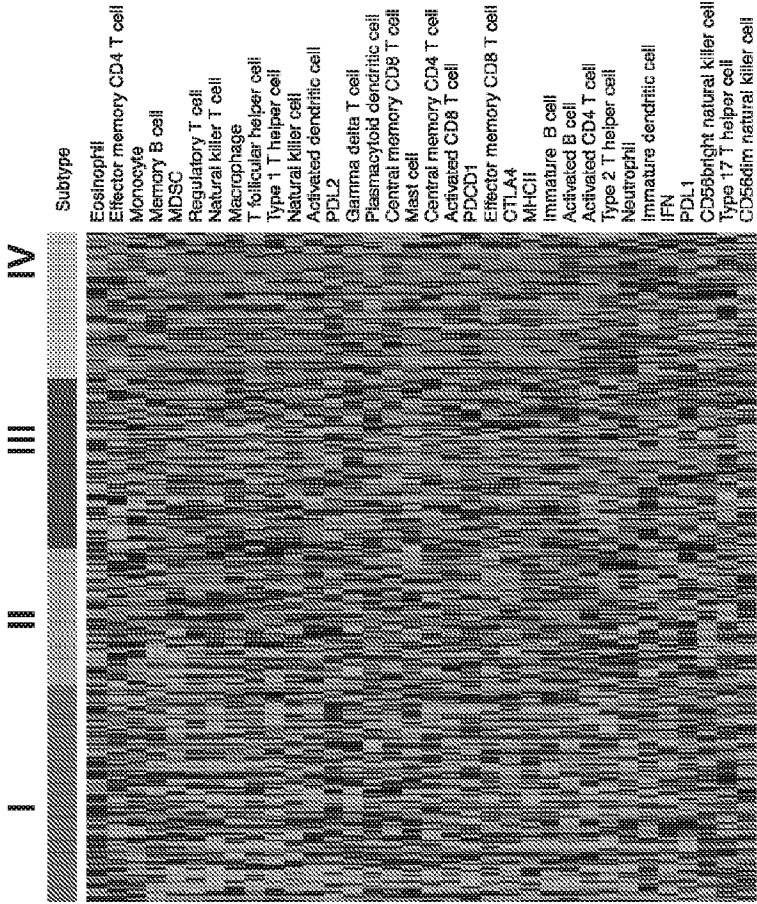
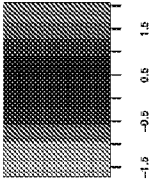


FIG. 17C

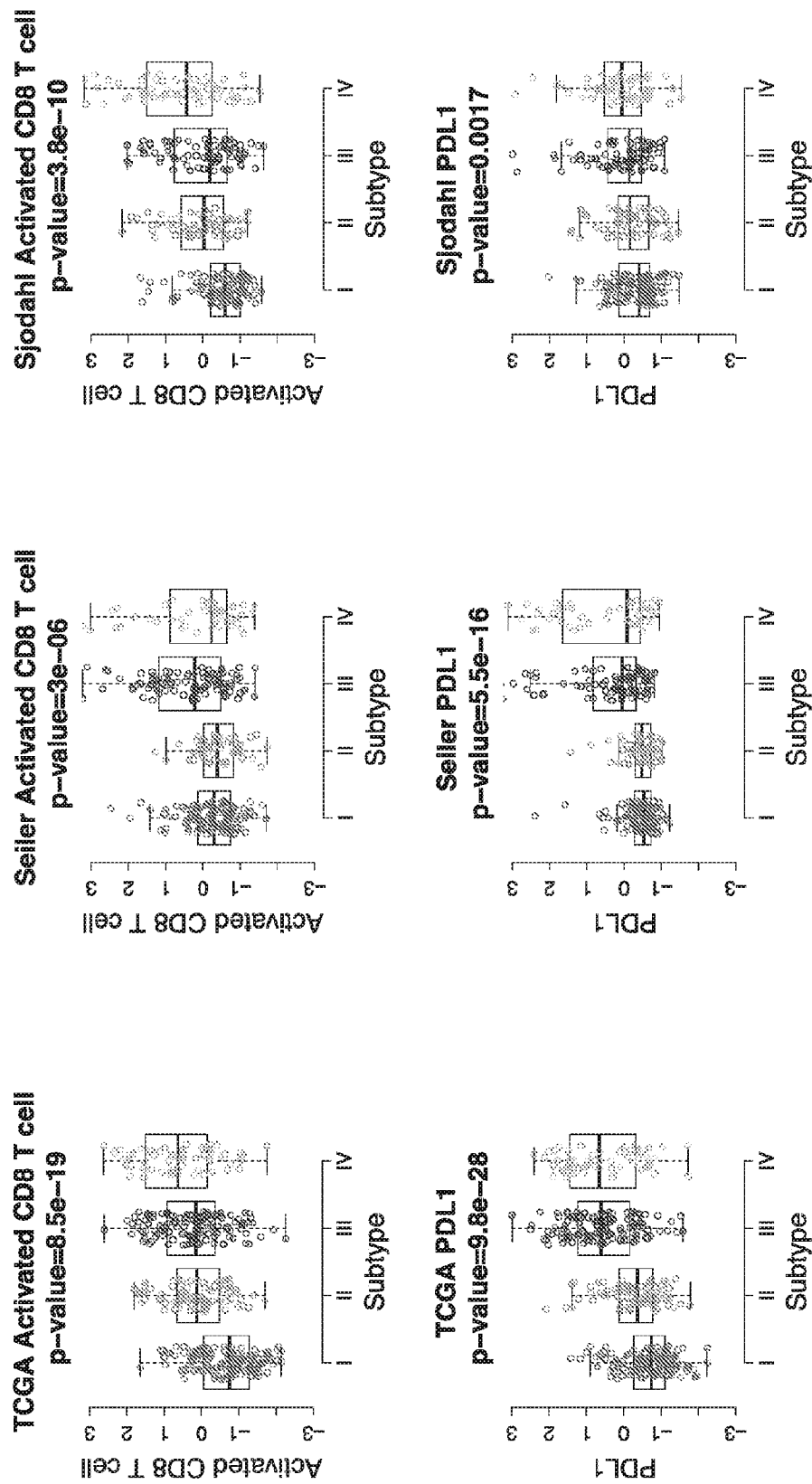


FIG. 18

TCGA RNAseq

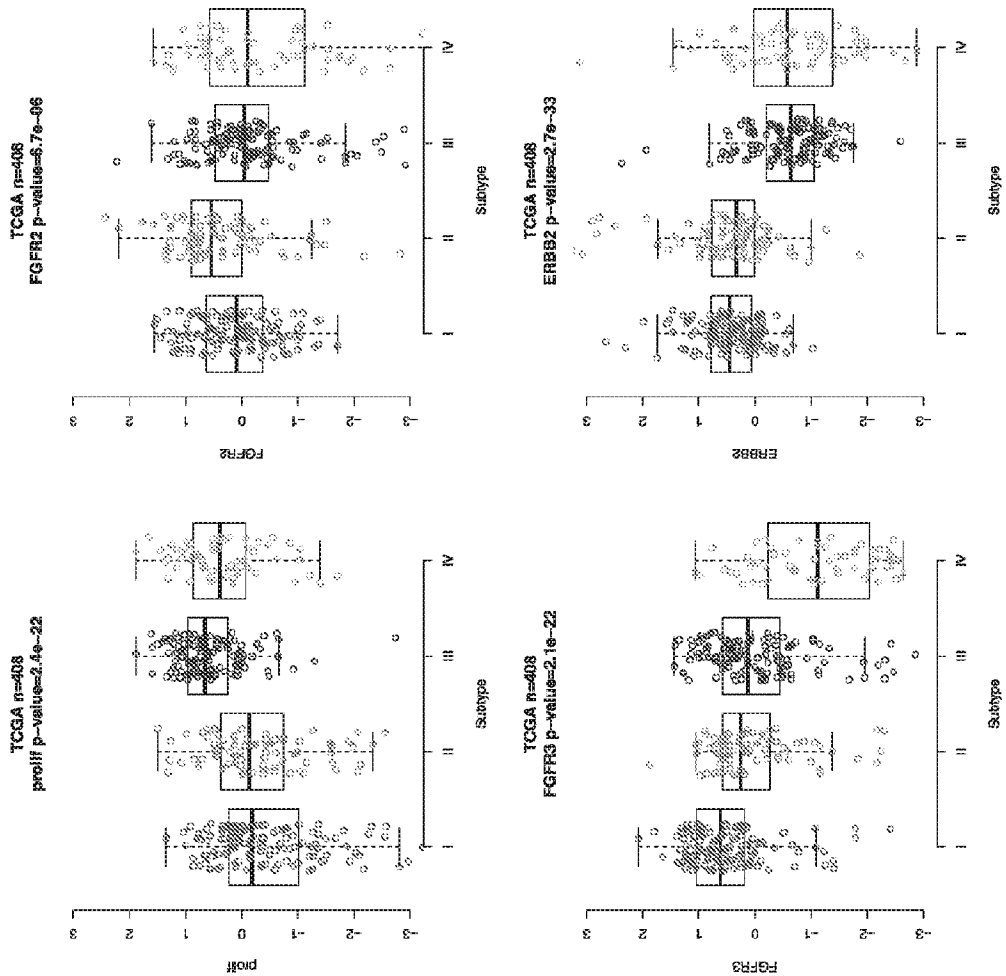


FIG. 19

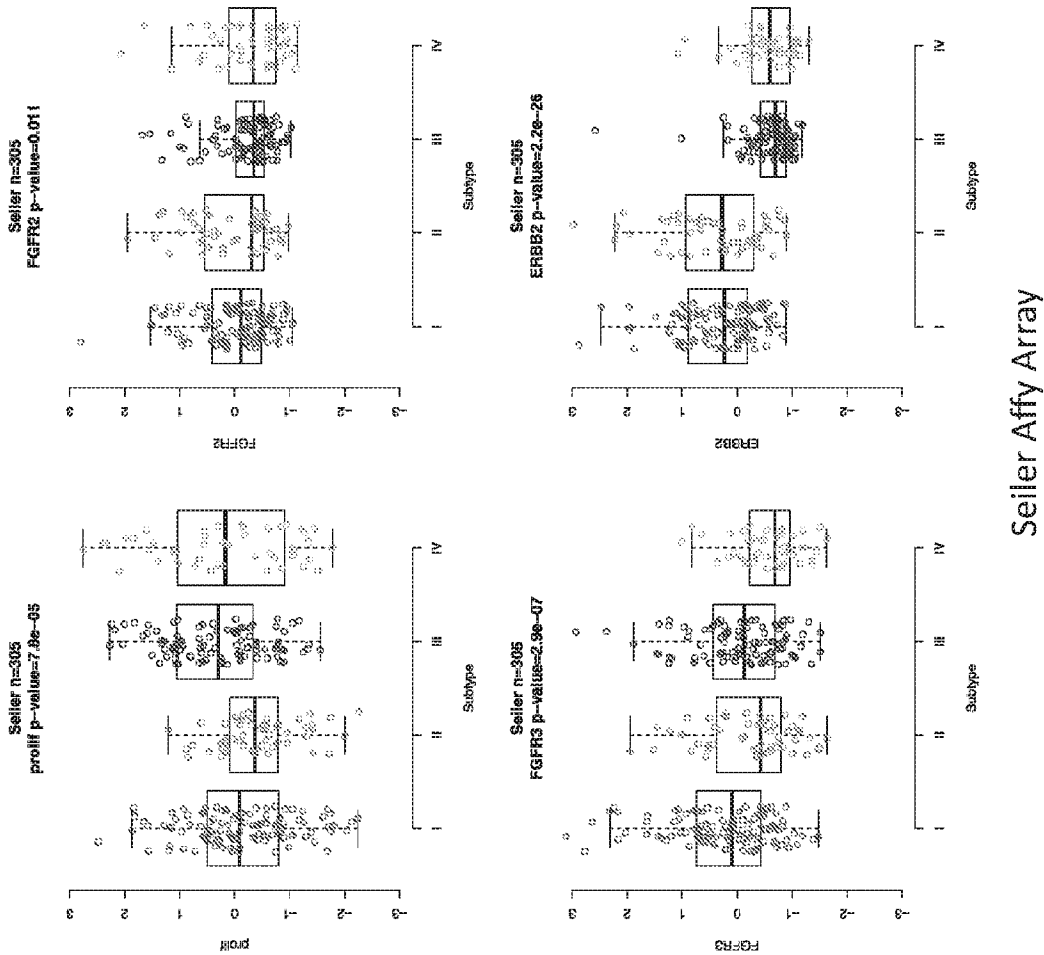


FIG. 19 (continued)

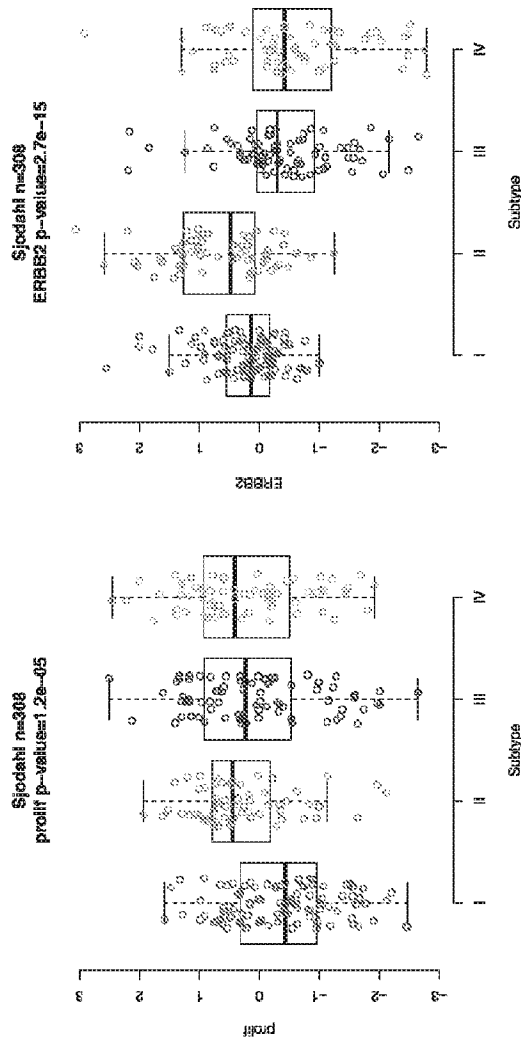
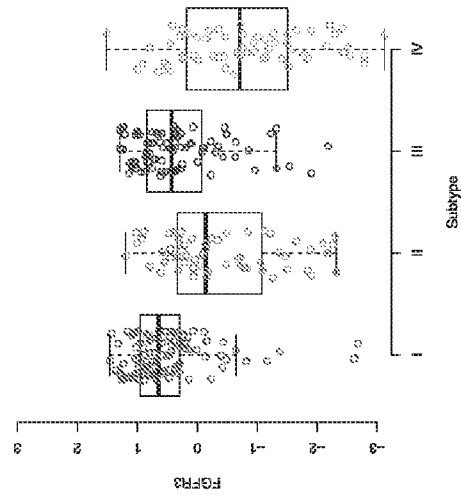


FIG. 19 (continued)

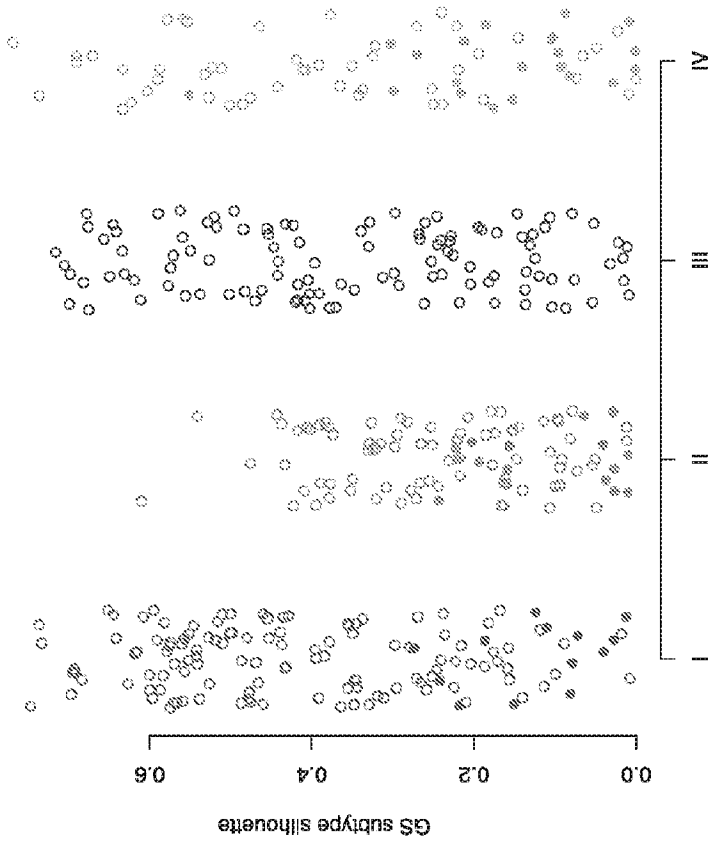


Sjodhal Illumina Beadchip

FIG. 20

5-typers (rows) vs 60-gene subtype (columns)					
	I	II	III	IV	Sum
Basal_squamous	3	1	99	39	142
Luminal	6	19	0	1	26
Luminal_infiltrated	1	52	10	15	78
Luminal_papillary	122	10	10	0	142
Neuronal	0	4	6	10	20
Sum	132	86	125	65	408

FIG. 21
TCGA 2017 n=408
Solid dot indicates gold standard
subtype and clanc subtype don't agree



Gold standard subtype (rows) vs clanc5050 (56) subtype (columns), agreement: 0.88

	I	II	III	IV	Sum
I	119	10	3	0	132
II	4	80	7	4	95
III	0	0	110	0	110
IV	0	6	15	50	71
Sum	123	96	135	54	408

	I	II	III	IV		I	II	III	IV
LOC100188947	3.200446	2.47801	-3.8801	-4.37765	TSSK2		0	0	0
UPK3A	1.882588	3.610797	-5.65419	-4.8912	ADORA2B	-0.04243	-1.99523	1.175231	-0.55327
KRT6B	-1.5279	-2.2406	7.549432	-0.39285	RHOV	0.587884	1.217374	-2.77856	-1.69619
ANXA6	-1.99651	0.642773	0.093877	1.662346	GRHL3	1.022277	1.45992	-1.39229	-6.231
FER1L4	1.710085	0.684431	-3.29832	-3.09831	GFPT2	-4.37252	0.832095	0.740754	1.949319
UPK1A	1.938255	2.816538	-5.91898	-6.9682	KRT6C	-1.75746	-2.5669	7.254879	-1.05224
DSG3	-2.16488	-1.58563	7.867258	-1.23773	C17orf28	0.801032	0.605037	-3.52447	-2.18042
EMP3	-1.88941	0.343301	0.700897	2.157498	SDC1	0.390035	0.130128	-0.01364	-2.41266
TBX3	1.480683	0.549164	-3.24421	-2.07565	ALDH1L2	-2.94616	0.513865	0.585891	1.307414
SNX31	1.396561	2.04992	-6.47583	-7.24402	KRT6A	-2.14698	-2.4806	7.610361	-0.58897
BNC1	-1.48301	-0.5707	7.967862	2.237664	CAPN5	1.168725	1.167339	-2.49187	-1.86401
CMKLR1	-2.01308	0.542859	0.300845	2.383412	GGT6	1.184901	0.334371	-2.22912	-5.6065
PPFIBP2	1.217493	0.728831	-1.84142	-2.19674	SERP2	-6.64573	2.503393	0.675616	1.67583
SFRP4	-5.6671	3.091055	0.381761	2.186904	CA9	0.703749	-4.00206	1.431651	-1.18965
IL20RB	-0.8862	-0.97047	4.280383	-0.59347	TOX3	2.43787	2.076916	-6.52131	-5.74298
V5IG4	-2.65714	0.592793	0.580961	2.608184	RAPGEFL1	1.008142	0.104638	-0.42977	-3.65996
FAM174B	1.460997	0.570248	-3.02958	-1.94295	CCDC80	-3.07843	1.135826	0.432487	1.874276
UPK2	1.848754	2.449998	-5.94129	-6.03892	E2F7	-0.19117	-1.11786	0.991805	0.355714
SERPINB13	-0.63778	-2.09682	5.682929	-2.9445	TLE2	0.958556	0.687707	-2.918	-1.05654
CD163	-2.94862	0.752134	0.398723	2.642102	ZNF750	-0.0658	0.518628	0.823911	-3.96537
C1orf126	1.653253	0.175782	-1.38005	-1.99388	ARSI	-3.66852	0.424043	1.722501	1.228148
MRVI1	-1.19127	1.648828	-0.14507	0.544091	SERPINB4	-1.12424	-3.5076	5.840766	-1.61832
SPRR2A	-1.49003	-0.31559	5.642145	-1.94896	SARM1	0.516351	0.468972	-1.87233	-0.35696
SLC7A7	-2.04585	0.370701	0.465131	2.010601	PVRL4	0.494431	0.363084	-0.39383	-2.9725
HMGCS2	3.624373	2.43858	-6.28311	-6.90542	CTHRC1	-3.30083	0.60929	0.643421	1.148149
EMILIN1	-1.95999	1.856179	-0.19675	1.837401	SERPINB3	-0.98946	-4.25249	5.934203	-2.2102
CD109	-2.26195	-0.3247	2.615978	1.522168	TBX2	1.539893	1.508642	-2.76057	-1.55403
PIK3AP1	-1.83069	0.185984	0.500394	2.183511	RHOV	-0.17371	0.204546	0.887872	-2.9274

FIG. 22

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/017799

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/6876; C12Q 1/6881; C12Q 1/6883; C12Q 1/6886 (2019.01)

CPC - C12Q 2539/00; C12Q 2539/10; C12Q 2539/101; C12Q 2539/103; C12Q 2539/107; C12Q 2539/113; C12Q 2600/112; C12Q 2600/118; C12Q 2600/158; C12Q 2600/178 (2019.05)

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

USPC - 435/18; 435/6.11; 435/6.12; 435/6.14; 702/19; 702/20 (keyword delimited)

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	WO 2015/073949 A1 (THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL) 21 May 2015 (21.05.2015) entire document	1-10, 15-24, 40, 48-53, 57-66
A	US 2011/0143959 A1 (NATIV et al) 16 June 2011 (16.06.2011) entire document	1-10, 15-24, 40, 48-53, 57-66
A	GLASER et al. "APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response," Oncotarget, 16 December 2017 (16.12.2017), Vol. 9, No. 4, Pgs. 4537-4548. entire document	1-10, 15-24, 40, 48-53, 57-66
A	US 2017/0044618 A1 (WASHINGTON UNIVERSITY et al.) 16 February 2017 (16.02.2017) entire document	1-10, 15-24, 40, 48-53, 57-66
A	US 2014/0295416 A1 (ZARBL et al) 02 October 2014 (02.10.2014) entire document	1-10, 15-24, 40, 48-53, 57-66
P, A	MAYHEW et al. "Abstract 4538: Bladder cancer gene expression subtypes (60 gene signature) to define prognosis, differential immune response, and biomarker associations," Journal of Clinical Oncology, 20 May 2018 (20.05.2018), Vol. 36, No. 15, Pgs. 4538-4538. entire document	1-10, 15-24, 40, 48-53, 57-66
P, A	KRUPENKO et al. "Chapter 11: ALDH1L1 and ALDH1L2 Folate Regulatory Enzymes in Cancer," AEMB: Alcohol and Cancer, 26 October 2018 (26.10.2018), Vol. 1032, Pgs. 127-143. entire document	1-10, 15-24, 40, 48-53, 57-66

☐ 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

"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

04 June 2019

Date of mailing of the international search report

03 JUL 2019

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

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/017799

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NO:1 was searched.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/017799

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.:
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:

3. ☐ Claims Nos.:
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.:
1-10, 15-24, 40, 48-53, and 57-66 to the extent that they read on a classifier biomarker of ALDH1L2 of SEQ ID NO:1.

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/US2019/017799

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 need to be paid.

Group I+: claims 1-70 are drawn to biomarkers, and methods comprising the same.

The first invention of Group I+ is restricted to a biomarker, and methods comprising the same, wherein the biomarker is selected to be ALDH1L2 of SEQ ID NO:1. It is believed that claims 1-10, 15-24, 40, 48-53, and 57-66 read on this first named invention and thus these claims will be searched without fee to the extent that they read on ALDH1L2 of SEQ ID NO:1.

Applicant is invited to elect additional biomarkers, each with specified designation and/or SEQ ID NO, to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a biomarker, and methods comprising the same, wherein the biomarker is selected to be ANXA6 of SEQ ID NO:2. Additional biomarkers will be searched upon the payment of an additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ 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 Groups I+ formulas do not share a significant structural element responsible for determining a bladder cancer subtype and/or treating bladder cancer, requiring the selection of alternatives for the bladder cancer biomarkers, where "the at least one biomarker nucleic acid is selected from a set of biomarkers listed in Table 1 or Table 2".

Additionally, even if Groups I+ were considered to share the technical features of a method for determining a bladder cancer subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype; a method for determining a muscle invasive bladder cancer (MIBC) subtype of a bladder cancer sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker, wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype; a method of detecting a biomarker in a bladder cancer sample obtained from a patient, the method comprising measuring the expression level of a plurality of classifier biomarker nucleic acids using an amplification, hybridization and/or sequencing assay; a method of treating muscle invasive bladder cancer (MIBC) in a subject, the method comprising: measuring the expression level of at least one biomarker nucleic acid in a MIBC sample obtained from the subject, wherein the at least one biomarker nucleic acid, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the MIBC; and administering a therapeutic agent based on the subtype of the MIBC; a method of predicting overall survival in a muscularly invasive bladder cancer (MIBC) patient, the method comprising detecting an expression level of at least one classifier biomarker in a MIBC sample obtained from a patient, wherein the detection of the expression level of the at least one classifier biomarker specifically identifies a luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtype, and wherein identification of the subtype is predictive of the overall survival in the patient; these shared technical features do not represent a contribution over the prior art.

Specifically, US 2017/0044618 A1 to Washington University et al. discloses a method for determining a bladder cancer subtype of a bladder cancer sample obtained from a patient (Disclosed are compositions and methods related intrinsic gene sets and methods and compositions related to detecting and classifying cancer, Abstract; disclosed compositions can be used to diagnose or prognose is the following:... bladder cancer, Para. [0088]; Also disclosed is a method of assigning a sample to an intrinsic subtype, Para. [0092]), the method comprising detecting an expression level of at least one classifier biomarker (amplifying and detecting intrinsic genes; and b) diagnosing cancer based on expression levels of the gene within the subject, Para. [0082]; weighting intrinsic classifier genes of multiple individuals identified using the method of step a relative to each other and forming classification clusters, Para. [0091]), wherein the detection of the expression level of the classifier biomarker specifically identifies a luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype (hierarchical clustering of microarray data separates breast tumors into different 'biological' subtypes (Luminal, HER2+/ER —, Basal-like, and Normal-like), Para. [0207]); and administering a therapeutic agent (e term "therapeutically effective" means that the amount of the composition used is of sufficient quantity to ameliorate one or more causes or symptoms of a disease or disorder... administer, Para. [0039]).

WO 2015/073949 A1 The University of North Carolina at Chapel Hill discloses a method for determining a muscle invasive bladder cancer (MIBC) subtype of a bladder cancer sample obtained from a patient (this invention is directed to the discovery of a method of subtyping high-grade bladder cancer and uses of the method for patient selection and treatment decision, Abstract; A predictor consisting of 47 genes was generated that accurately predict the subtypes from the meta-dataset training set, Para. [0024]; The term "high grade bladder cancer" (HGBC) means and includes a tumor that have invaded into the muscularis propria of the bladder: non-muscle invasive bladder cancer (NMIBC, Ta, T1) and muscle invasive bladder cancer (MIBC, >T2) including bladder cancer metastases, Para. [0030]; collecting a patient sample having a cancer cell or tissue, such as a bladder tissue sample, Para. [0032]), the method comprising detecting an expression level of at least one classifier biomarker (measuring levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model; and (c) comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model with levels of a plurality of high-grade bladder cancer classifier biomarkers associated with a control, Para. [0020]), wherein the detection of the expression level of the classifier biomarker specifically identifies a

luminal, luminal infiltrated, basal or basal infiltrated/neuronal bladder cancer subtype ((a) detecting expression levels of a plurality of high-grade bladder cancer classifier biomarkers in the sample; (b) applying a statistical analysis to the detected expression; and (c) identifying the intrinsic high grade bladder cancer subtype based on the statistical analysis of the detected expression of the classifier biomarkers, Para. [0008]); a method of detecting a biomarker in a bladder cancer sample obtained from a patient (this invention is directed to the discovery of a method of subtyping high-grade bladder cancer and uses of the method for patient selection and treatment decision, Abstract; A predictor consisting of 47 genes was generated that accurately predict the subtypes from the meta-dataset training set, Para. [0024]; The term "high grade bladder cancer" (HGBC) means and includes a tumor that have invaded into the muscularis propria of the bladder: non-muscle invasive bladder cancer (NMIBC, Ta, T1) and muscle invasive bladder cancer (MIBC, >T2) including bladder cancer metastases, Para. [0030]; collecting a patient sample having a cancer cell or tissue, such as a bladder tissue sample, Para. [0032]), the method comprising measuring the expression level of a plurality of classifier biomarker nucleic acids using an amplification, hybridization and/or sequencing assay (measuring levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model; and (c) comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model with levels of a plurality of high-grade bladder cancer classifier biomarkers associated with a control, Para. [0020]; Isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays, Para. [0045]); a method of predicting overall survival in a muscularly invasive bladder cancer (MIBC) patient (high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient, Para. [0019]; When examining the disease-specific survival of the MSKCC tumors (>pT2) stratified by BASE47 subtype calls, basal-like tumors had a significantly decreased disease-specific survival ($p=0.0194$, $HR=2.8$) (Figure 4A). Therefore, the intrinsic bladder subtypes not only reflect bladder cancer biology but have prognostic value, Para. [00120]), the method comprising detecting an expression level of at least one classifier biomarker in a MIBC sample obtained from a patient (measuring levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model; and (c) comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model with levels of a plurality of high-grade bladder cancer classifier biomarkers associated with a control, Para. [0020]; Isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays, Para. [0045]), wherein the detection of the expression level of the at least one classifier biomarker specifically identifies a luminal, luminal infiltrated, basal, and basal infiltrated/neuronal subtype, and wherein identification of the subtype is predictive of the overall survival in the patient ((a) detecting expression levels of a plurality of high-grade bladder cancer classifier biomarkers in the sample; (b) applying a statistical analysis to the detected expression; and (c) identifying the intrinsic high grade bladder cancer subtype based on the statistical analysis of the detected expression of the classifier biomarkers, Para. [0008]; high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient, Para. [0019]; When examining the disease-specific survival of the MSKCC tumors (>pT2) stratified by BASE47 subtype calls, basal-like tumors had a significantly decreased disease-specific survival ($p=0.0194$, $HR=2.8$) (Figure 4A). Therefore, the intrinsic bladder subtypes not only reflect bladder cancer biology but have prognostic value, Para. [00120]); a method of treating muscle invasive bladder cancer (MIBC) in a subject (A patient identified with a basal or luminal high-grade bladder cancer subtype may receive differing treatment for their high-grade bladder cancer, Para. [0085]; and muscle invasive bladder cancer (MIBC, Para. [0030]), the method comprising: measuring the expression level of at least one biomarker nucleic acid in a MIBC sample obtained from the subject (measuring levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model; and (c) comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers in the tissue or the animal model with levels of a plurality of high-grade bladder cancer classifier biomarkers associated with a control, Para. [0020];); collecting a patient sample having a cancer cell or tissue, such as a bladder tissue sample, Para. [0032]), wherein the at least one biomarker nucleic acid, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the MIBC ((a) detecting expression levels of a plurality of high-grade bladder cancer classifier biomarkers in the sample; (b) applying a statistical analysis to the detected expression; and (c) identifying the intrinsic high grade bladder cancer subtype based on the statistical analysis of the detected expression of the classifier biomarkers, Para. [0008]; high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient, Para. [0019]; the expression of a biomarker of the invention can be detected on a nucleic acid level, Para. [0036]).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.