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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 May 2015 (21.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/073949 A1

(51) International Patent Classification:
G01N 33/533 (2006.01) *C12Q 1/68* (2006.01)
G01N 33/574 (2006.01)

(21) International Application Number:
PCT/US2014/065910

(22) International Filing Date:
17 November 2014 (17.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/904,488 15 November 2013 (15.11.2013) US

(71) Applicant: **THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL** [US/US]; Office of Technology Development, 100 Europa Drive, Suite 430, Chapel Hill, NC 27517 (US).

(72) Inventors: **KIM, William, Y.**; c/o The University of North Carolina at Chapel Hill, 100 Europa Drive, Suite 430, Chapel Hill, NC 27517 (US). **DAMRAUER, Jeffrey, S.**; c/o The University of North Carolina at Chapel Hill, 100 Europa Drive, Suite 430, Chapel Hill, NC 27517 (US).

(74) Agents: **LETTS, Nathan, P.** et al.; Olive Law Group, PLLC, 125 Edinburgh South Drive, Suite 220, Cary, NC 27511 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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).

Published:

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

(54) Title: METHOD OF SUBTYPING HIGH-GRADE BLADDER CANCER AND USES THEREOF

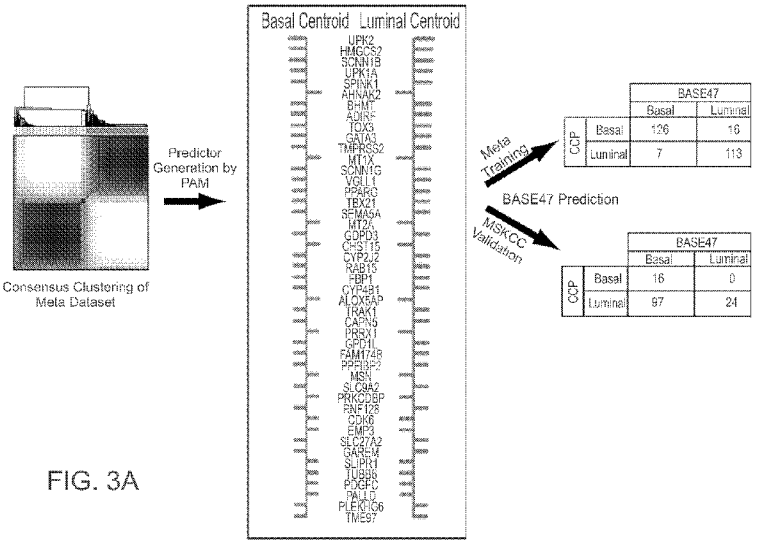


FIG. 3A

(57) Abstract: 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 decisions.

WO 2015/073949 A1

METHOD OF SUBTYPING HIGH-GRADE BLADDER CANCER AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional 61/904,488 filed November 15, 2013, Kim *et al.*, Atty. Dkt. UNC12004USV which is hereby incorporated by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under Grant No. HL069768 awarded by the National Institutes of Health. The United States Government has certain rights in the invention.

1. FIELD OF THE INVENTION

[0003] This invention relates generally to the discovery of a method of subtyping high-grade bladder cancer and uses of the method for patient selection and treatment decisions.

2. BACKGROUND OF THE INVENTION

2.1. Introduction

[0004] In the United States, urothelial carcinoma (UC) of the bladder is the fourth most common malignancy in men and ninth most common in women with 72,570 new cases and 15,210 deaths expected in 2013(1). Bladder cancer is a heterogeneous disease that is histologically divided into low-grade and high-grade disease. While low-grade tumors are almost invariably non-invasive (Ta), high-grade tumors can be classified into two major groups based on whether tumors have invaded into the muscularis propria of the bladder: non-muscle invasive bladder cancer (NMIBC, Tis, Ta, T1) and muscle invasive bladder cancer (MIBC, ≥T2). Low-grade tumors are associated with a high rate of recurrence, yet an excellent overall prognosis with a 5-year survival in the range of 90%. In contrast, high-grade, muscle-invasive bladder cancer has a relatively poor 5-year overall survival: 68% when T2 and decreasing to 15% for non-organ confined disease (pT3 and pT4) (1,2).

[0005] Along with divergent pathologies and prognosis, low-grade and high-grade UCs are associated with distinct genetic alterations. For example, low-grade UC is enriched for activating mutations in *FGFR3*, *PIK3CA* and inactivating *UTX* mutations, whereas high-grade, muscle-invasive tumors are enriched for *TP53* and *RB1* pathway alterations (3-10).

[0006] Several reports have examined the gene expression profiles of primary bladder tumors. From these studies, it is apparent that low-grade, non-invasive and high-grade, muscle-invasive tumors harbor distinct gene expression patterns and that further molecular subsets can be found within low-grade and high-grade tumors (11-15). Moreover, a number of gene signatures have been developed that can predict tumor stage, lymph node metastases, or bladder cancer progression (11,13,14,16-19). Taken together, there are established gene expression patterns that differentiate low-grade and high-grade tumors, however there is little data identifying intrinsic subtypes specifically within high-grade disease.

[0007] The present gold standard strategy for bladder diagnosis is non-invasive voided urine cytology, followed by cystoscopic examination. However, both methods have low sensitivity, especially for low grade tumors. Kaufman *et al* 2009 *Lancet* 374:239-249. To predict which patients will progress from superficial to invasive disease remains a challenge. Patients diagnosed with early-stage bladder undergo frequent monitoring, currently based on cystoscopy and cytology, resulting in bladder becoming one of the most costly cancer diseases to manage. Bischoff *et al.* 2009 *Curr Opin Oncol* 21:272-277. Better, more effective tests for detection of bladder cancer are needed to lower the morbidity and mortality associated with bladder cancer.

3. SUMMARY OF THE INVENTION

[0008] In particular non-limiting embodiments, the present invention provides a method to identify intrinsic high-grade bladder cancer subtypes in a patient sample which comprises:
(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.

[0009] The intrinsic high-grade bladder cancer subtypes may be a basal subtype and a luminal subtype. The classifier biomarkers may be selected from the biomarkers in Table 3.

[0010] The method may involve 47 classifier biomarkers (BASE47). Alternatively, the method may involve 40 classifier biomarkers, 42 classifier biomarkers, 44 classifier biomarkers, 45 classifier biomarkers, 46 classifier biomarkers, 48 classifier biomarkers, 49 classifier biomarkers, 50 classifier biomarkers, or 52 classifier biomarkers.

[0011] In one embodiment, the classifier biomarkers may be nucleic acids and detected by microarray, next generation sequencing, polymerase chain reaction (PCR), or a direct transcript counting method.

[0012] The invention also provides a method of diagnosing high-grade bladder cancer subtypes in a sample from a subject comprising: (a) detecting a plurality of classifier biomarkers in a sample from the subject, by a PCR assay with primers/probes specific for the classifier biomarkers; (b) comparing the detected levels to at least one sample from a training set(s), wherein a sample training set(s) comprises data from the levels from a reference sample, and the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the detected levels in the sample from the subject and the detected levels from at least one training set(s); and (c) diagnosing high-grade bladder cancer subtypes based on the detected levels in the sample from the subject and the results of the statistical algorithm.

[0013] The invention also provides a method of diagnosing high-grade bladder cancer subtypes in a sample from a subject comprising: (a) detecting a plurality of classifier biomarkers in a sample from the subject, by a direct mRNA counting assay specific for the classifier biomarkers; (b) comparing the detected levels to at least one sample from a training set(s), wherein a sample training set(s) comprises data from the levels from a reference sample, and the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the detected levels in the sample from the subject and the detected levels from at least one training set(s); and (c) diagnosing high-grade bladder cancer subtypes based on the detected levels in the sample from the subject and the results of the statistical algorithm.

[0014] The invention also provides a method of diagnosing high-grade bladder cancer subtypes in a sample from a subject comprising: (a) detecting a plurality of classifier biomarkers in a sample from the subject, by antibodies specific for the classifier biomarkers; (b) comparing the detected levels to at least one sample from a training set(s), wherein a sample training set(s) comprises data from the levels from a reference sample, and the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the detected levels in the sample from the subject and the detected levels from at least one training set(s); and (c) diagnosing high-grade bladder cancer subtypes based on the detected levels in the sample from the subject and the results of the statistical algorithm.

[0015] In another embodiment, the classifier biomarkers may be proteins which may be detected by an antibody assay.

[0016] The patient sample may be a bladder tissue sample such as a formalin-fixed paraffin embodded (FFPE) or a fresh frozen bladder tissue sample. Alternatively, the patient sample may be a blood sample containing circulating tumor cells or a voided urine cell sample.

[0017] The patient sample may be a sample of cell-free nucleic acids from blood or urine.

[0018] The invention also provides a method to select patients with a poor prognosis using intrinsic high-grade bladder cancer subtypes in a sample which comprises: (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) selecting patients with the poor prognosis using the intrinsic high-grade bladder cancer subtype based on the statistical analysis of the detected expression of the classifier biomarkers.

[0019] A kit for identification of high-grade bladder cancer subtypes in a patient sample which comprises: (a) a means for measuring levels of a plurality of high-grade bladder cancer classifier biomarkers; and (b) instructions for comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers from the patient sample with levels of a plurality of high-grade bladder cancer classifier biomarkers for a control patient, wherein the levels of a plurality of high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient.

[0020] A method of identifying a compound that slows the progression or treats high-grade bladder cancer, the method comprising the steps of: (a) contacting a tissue or an animal model with a compound; (b) 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; and determining a functional effect of the compound on high-grade bladder cancer, thereby identifying a compound that slows the progression or treats high-grade bladder cancer.

[0021] See also Damrauer *et al.* 2014 "Intrinsic Subtype of High Grade Bladder Cancer Reflect the Hallmarks of Breast Cancer Biology," *PNAS* 111(8) 3110-3115 and supplemental information, the contents of which are hereby incorporated by reference in its entirety.

4. BRIEF DESCRIPTION OF THE FIGURES

[0022] Figure 1A-1G. Consensus Clustering defines two distinct molecular subtypes of invasive bladder cancer. **(1A)** Consensus Clustering was performed on 262 muscle-invasive tumors, curated from four publically available datasets (Meta dataset), yielding two subtypes. **(1B)** Consensus Clustering was independently performed on a dataset of high-grade bladder tumors obtained from MSKCC (n=49). **(1C)** The median gene expression of all common genes between the datasets were compared and the Pearson correlation was plotted (light gray = correlation, dark gray=anti-correlation). **(1D)** Gene expression of epithelial and urothelial markers were visualized by heatmap, supervised by consensus cluster plus calls in the meta-dataset. KRT5 mRNA expression was plotted against **(1E)** UPK2 and **(1F)** KRT20 expression in the meta-dataset. (**1G**) Significantly differentially expressed genes between K1 and K2 and their respective fold change, as determined by 2-class SAM (3,374 genes, FDR=0) were analyzed for predicted pathway enrichment. Selected significant pathways enriched in K1 are represented.

[0023] Figure 2A-2D. Basal, luminal, claudin low and oncogenic breast cancer signatures are associated with intrinsic molecular subtypes of bladder cancer. **(2A)** The meta datasets tumors were run against previously published breast cancer related gene sets and the resulting pathway scores were cluster by hierarchical clustering and heatmaps were generated for visualization. **(2B)** Waterfall plots representing the correlation of basal-like and luminal bladder cancer to the basal and luminal breast cancer centroid as determined by PAM50. **(2C)** Gene groups from Parker. *et al.* that are associated with the claudin low as well as the additional intrinsic molecular subtypes of breast cancer were used to hierarchically cluster the meta dataset. Claudin low samples were identify using a 807 gene signature (717/807) signature. Median gene expression of genes present in the intrinsic gene list were determined and 1-pearson correlation was calculated comparing bladder subtypes (bladder tumors) to breast cancer subtypes (breast tumors) in both the Breast TCGA and UNC337 datasets. **(2D)** The meta-dataset tumors were clustered by genes that defined the intrinsic subtypes of breast cancer. Tracks indicated bladder cancer subtypes as well as subtypes predicted by the breast cancer PAM50 bioclassifier.

[0024] Figure 3A-3B. Generation of the BASE47 subtype predictor. **(3A)** Prediction Analysis of Microarrays (PAM) was performed using the basal-like and luminal subtype calls generated by consensus cluster plus. A predictor consisting of 47 genes was generated that accurately predict the subtypes from the meta-dataset training set ($p<0.001$) as well as a MSKCC validation dataset ($p<0.001$). **(3B)** The BASE47 gene list was used to cluster the MSKCC dataset, showing two distinct expression profiles.

[0025] Figure 4A-4C. Luminal and basal bladder cancer have differential survival and are associated with distinct genomic alterations. **(4A)** A Kaplan-Meier plot for the MSKCC data (≥pT2) was generated for disease specific survival. Basal-like tumors (n=22) had a significantly decreased disease free survival as compared to luminal tumors (n=16) (p=0.0194). **(4B)** Superficial tumors, which were not included in the generation of BASE47, were subjected to BASE47 subtype prediction. **(4C)** Sequencing was performed on common mutations in bladder cancer. *FGFR3* and *TSC1* alterations were significantly enriched in luminal bladder cancer whereas alterations of the RB1 pathway were enriched in basal-like bladder cancer. *TP53* alterations were distributed evenly in both subtypes. Basal-like and luminal tumors from the meta-dataset were annotated for gender (2/4 data). Basal-like bladder cancer was significantly enriched for female patients (chi square p value=0.0203).

[0026] Figure 5A-5E. Consensus Clustering defines two distinct molecular subtypes of invasive bladder cancer **(5A)** Consensus Cumulative Distribution Function (CDF) plot were generated by Consensus Cluster Plus clustering (1) on the Meta-dataset and the **(5B)** MSKCC dataset. **(5C)** Genes representative of epithelial and urothelial differentiation were used to generate a heatmap, supervised by subtype in the MSKCC dataset. KRT5 mRNA expression was plotted against **(5D)** UPK2 and **(5E)** KRT20 expression in the MSKCC dataset.

[0027] Figure 6. Makers of luminal breast cancer are co-expressed with markers of urothelial differentiation Hierarchal clustering was performed on all significantly differentially expressed (2-class SAM) genes. The node containing the urothelial makers of differentiation, UPK2 and UPK1A, was isolated and overlapped with the Parker et. al. breast cancer intrinsic gene list (2). A significant number of genes within the node overlapped the breast cancer gene list (19/65 genes, chi squared pvalue=0.006).

[0028] Figure 7. Basal, luminal, and oncogenic breast cancer signatures are associated with intrinsic molecular subtypes of bladder cancer. The MSKCC datasets tumors were run against previously published breast cancer related gene sets (3) and the resulting pathway scores were cluster by hierarchical clustering and heatmaps were generated for visualization.

[0029] Figure 8. A subset of basal-like bladder tumors are claudin-low. The MSKCC dataset was hierarchically clustered using representative genes known to define claudin-low breast tumors. Claudin low subtype designation was performed using a 807 gene signature.

5. DETAILED DESCRIPTION OF THE INVENTION

[0030] 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 muscleinvasive bladder cancer (MIBC, ≥T2) including bladder cancer metastases. A determination of HGBC may be made by a pathologist.

[0031] In one embodiment, the invention is directed to the use of BASE47 panel of genes in Table 3. For definitions of the gene abbreviations, see GeneCards (www.genecards.org), U.S. National Library of Medicine, National Center for Biotechnology Information (NCBI) Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene), or European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute (WTSI), Ensembl database (http://useast.ensembl.org/index.html) or BLAT on University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/cgi-bin/hgBlat) for additional information such as sequences, single nucleotide polymorphisms (SNPs).

5.1. Samples

[0032] In particular embodiments, the methods for evaluating bladder cancer include collecting a patient sample having a cancer cell or tissue, such as a bladder tissue sample, a primary bladder tumor tissue sample, a lymph node tissue suspected of being from bladder cancer metastasis.

[0033] By "patient sample" is intended any sampling of cells, tissues, or bodily fluids in which expression of a biomarker can be detected. Examples of such samples include, but are not limited to, biopsies and smears. Bodily fluids useful in the present invention include blood, lymph, urine, saliva, nipple aspirates, gynecological fluids, or any other bodily secretion or derivative thereof. Blood can include whole blood, plasma, serum, or any derivative of blood. In some embodiments, the sample includes bladder cells, particularly bladder tissue from a biopsy, such as a bladder tumor tissue sample. Body samples may be obtained from a patient by a variety of techniques including, for example, by scraping or swabbing an area, by using a needle to aspirate cells or bodily fluids, or by removing a tissue sample (i.e., biopsy). Methods for collecting various body samples are well known in the art. In some embodiments, a bladder tissue sample is obtained by, for example, fine needle aspiration biopsy, core needle biopsy, or excisional biopsy. Fixative and staining solutions may be applied to the cells or tissues for preserving the specimen and for facilitating examination. Patient samples, particularly bladder tissue samples, may be transferred to a glass slide for viewing under magnification. In one embodiment, the body sample is a formalin-fixed, paraffin-embedded (FFPE) bladder tissue sample, particularly a primary bladder tumor sample or a metastatic bladder tumor sample.

Preferably, the sample is a high-grade bladder cancer sample. The sample may also be 1) cell free RNA from urine; 2) cell free RNA from blood; 3) RNA from cells from urine; 4) RNA from cells from cytospin (voided urine) cell blocks; or 5) RNA from circulating tumor cells.

[0034] Methods for detecting expression of the biomarkers of the invention, that is, gene expression profiling, include methods based on hybridization analysis of polynucleotides, methods based on sequencing of polynucleotides, immunohistochemistry methods, and proteomics-based methods. These methods are discussed in greater detail below.

[0035] The term "probe" refers to any molecule that is capable of selectively binding to a specifically intended target biomolecule, for example, a nucleotide transcript or a protein encoded by or corresponding to a biomarker. Probes can be synthesized by one of skill in the art, or derived from appropriate biological preparations. Probes may be specifically designed to be labeled. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic molecules.

[0036] Any methods available in the art for detecting expression of biomarkers are encompassed herein. The expression of a biomarker of the invention can be detected on a nucleic acid level (e.g., as an RNA transcript) or a protein level. By "detecting expression" is intended determining the quantity or presence of an RNA transcript or its expression product of a biomarker gene. Thus, "detecting expression" encompasses instances where a biomarker is determined not to be expressed, not to be detectably expressed, expressed at a low level, expressed at a normal level, or overexpressed. In order to determine overexpression, the body sample to be examined can be compared with a corresponding body sample that originates from a healthy person. That is, the "normal" level of expression is the level of expression of the biomarker in, for example, a bladder tissue sample from a human subject or patient not afflicted with bladder cancer. Such a sample can be present in standardized form. In some embodiments, determination of biomarker overexpression requires no comparison between the body sample and a corresponding body sample that originates from a healthy person. For example, detection of overexpression of a biomarker indicative of a poor prognosis in a bladder tumor sample may preclude the need for comparison to a corresponding bladder tissue sample that originates from a healthy person. Moreover, in some aspects of the invention, no expression, under-expression or normal expression (i.e., the absence of overexpression) of a biomarker or combination of biomarkers of interest provides useful information regarding the prognosis of a bladder cancer patient.

[0038] Methods for detecting expression of the biomarkers of the invention, that is, gene expression profiling, include methods based on hybridization analysis of polynucleotides,

methods based on sequencing of polynucleotides, immunohistochemistry methods, and proteomics-based methods.

[0039] The term "probe" refers to any molecule that is capable of selectively binding to a specifically intended target biomolecule, for example, a nucleotide transcript or a protein encoded by or corresponding to a biomarker. Probes can be synthesized by one of skill in the art, or derived from appropriate biological preparations. Probes may be specifically designed to be labeled. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic molecules.

[0040] The term "sensitivity" as used herein refers to the number of true positives divided by the number of true positives plus the number of false negatives, where sensitivity ("sens") may be within the range of $0 < sens < 1$. Ideally, method embodiments herein have the number of false negatives equaling zero or close to equaling zero, so that no subject is wrongly identified as not having HGBC when they indeed have HGBC. Conversely, an assessment often is made of the ability of a prediction algorithm to classify negatives correctly, a complementary measurement to sensitivity. The term "specificity" as used herein refers to the number of true negatives divided by the number of true negatives plus the number of false positives, where specificity ("spec") may be within the range of $0 < spec < 1$. Ideally, the methods described herein have the number of false positives equaling zero or close to equaling zero, so that no subject is wrongly identified as having HGBC when they do not in fact have HGBC. Hence, a method that has both sensitivity and specificity equaling one, or 100%, is preferred.

5.2. Analysis Of Polynucleotides

[0041] In some embodiments, the expression of a biomarker of interest is detected at the nucleic acid level. Nucleic acid-based techniques for assessing expression are well known in the art and include, for example, determining the level of biomarker RNA transcripts (i.e., mRNA) in a body sample. Many expression detection methods use isolated RNA. The starting material is typically total RNA isolated from a body sample, such as a tumor or tumor cell line, and corresponding normal tissue or cell line, respectively. Thus RNA can be isolated from a variety of primary tumors, including bladder, prostate, uterus, urethral tissue, lymph nodes, and the like, or tumor cell lines. If the source of mRNA is a primary tumor, mRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g., formalin-fixed) tissue samples.

[0042] 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).

[0043] Commonly used methods known in the art for the quantification of mRNA expression in a sample include northern blotting and in situ hybridization (Parker and Barnes, *Methods Mol. Biol.* 106:247-83, 1999), RNAse protection assays (Hod, *Biotechniques* 13:852-54, 1992), PCR-based methods, such as reverse transcription PCR(RT-PCR) (Weis *et al.*, *TIG* 8:263-64, 1992), and array-based methods (Schena *et al.*, *Science* 270:467-70, 1995). Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes, or DNA-protein duplexes. Representative methods for sequencing-based gene expression analysis include Serial Analysis of Gene Expression (SAGE), bead-based technologies, single molecule fluorescence *in situ* hybridization (smFISH) studies, and gene expression analysis by massively parallel signature sequencing. Velculescu *et al.* 1995 *Science* 270 484-487; Streefkerk *et al.*, 1976, *Pro Biol Fluid Proc Coll* 24 811-814; Soini U.S. Pat. No. 5,028,545; smFISH, Lyubimova *et al.* 2013 *Nat Protocol* 8(9) 1743-1758.

[0045] 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. One method for the detection of mRNA levels involves contacting the isolated mRNA 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 full-length 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 an mRNA or genomic DNA encoding a biomarker of the present invention. Hybridization of an mRNA with the probe indicates that the biomarker in question is being expressed.

[0046] In one embodiment, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative embodiment, the probes are immobilized on a solid surface and the mRNA is contacted with the probes, for example, in an Agilent gene chip array. A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the biomarkers of the present invention.

5.3. Polynucleotide Sequence Amplification and Determination

[0047] In many instances, it is desirable to amplify a nucleic acid sequence using any of several nucleic acid amplification procedures which are well known in the art. Specifically, nucleic acid amplification is the chemical or enzymatic synthesis of nucleic acid copies which contain a sequence that is complementary to a nucleic acid sequence being amplified (template). The methods and kits of the invention may use any nucleic acid amplification or detection methods known to one skilled in the art, such as those described in U.S. Pat. Nos. 5,525,462 (Takarada *et al.*); 6,114,117 (Hepp *et al.*); 6,127,120 (Graham *et al.*); 6,344,317 (Urnovitz); 6,448,001 (Oku); 6,528,632 (Catanzariti *et al.*); and PCT Pub. No. WO 2005/111209 (Nakajima *et al.*); all of which are incorporated herein by reference in their entirety.

[0048] In some embodiments, the nucleic acids are amplified by PCR amplification using methodologies known to one skilled in the art. One skilled in the art will recognize, however, that amplification can be accomplished by any known method, such as ligase chain reaction (LCR), Qβ-replicase amplification, rolling circle amplification, transcription amplification, self-sustained sequence replication, nucleic acid sequence-based amplification (NASBA), each of which provides sufficient amplification. Branched-DNA technology may also be used to qualitatively demonstrate the presence of a sequence of the technology, which represents a particular methylation pattern, or to quantitatively determine the amount of this particular genomic sequence in a sample. Nolte reviews branched-DNA signal amplification for direct quantitation of nucleic acid sequences in clinical samples (Nolte, 1998, *Adv. Clin. Chem.* 33:201-235).

[0049] The PCR process is well known in the art and is thus not described in detail herein. For a review of PCR methods and protocols, see, e.g., Innis *et al.*, eds., PCR Protocols, A Guide to Methods and Application, Academic Press, Inc., San Diego, Calif. 1990; U.S. Pat. No. 4,683,202 (Mullis); which are incorporated herein by reference in their entirety. PCR reagents and protocols are also available from commercial vendors, such as Roche Molecular Systems. PCR may be carried out as an automated process with a thermostable enzyme. In this process, the temperature of the reaction mixture is cycled through a denaturing region, a primer annealing region, and an

extension reaction region automatically. Machines specifically adapted for this purpose are commercially available.

5.4. High Throughput, Single Molecule Sequencing, and Direct Detection Technologies

[0050] Suitable next generation sequencing technologies are widely available. Examples include the 454 Life Sciences platform (Roche, Branford, CT) (Margulies *et al.* 2005 *Nature*, 437, 376-380); lllumina's Genome Analyzer, GoldenGate Methylation Assay, or Infinium Methylation Assays, i.e., Infinium HumanMethylation 27K BeadArray or VeraCode GoldenGate methylation array (Illumina, San Diego, CA; Bibkova *et al.*, 2006, *Genome Res.* 16, 383-393; U.S. Pat. Nos. 6,306,597 and 7,598,035 (Macevicz); 7,232,656 (Balasubramanian *et al.*)); or DNA Sequencing by Ligation, SOLiD System (Applied Biosystems/Life Technologies; U.S. Pat. Nos. 6,797,470, 7,083,917, 7,166,434, 7,320,865, 7,332,285, 7,364,858, and 7,429,453 (Barany *et al.*); or the Helicos True Single Molecule DNA sequencing technology (Harris *et al.*, 2008 *Science*, 320, 106-109; U.S. Pat. Nos. 7,037,687 and 7,645,596 (Williams *et al.*); 7,169,560 (Lapidus *et al.*); 7,769,400 (Harris)), the single molecule, real-time (SMRT™) technology of Pacific Biosciences, and sequencing (Soni and Meller, 2007, *Clin. Chem.* 53, 1996-2001) which are incorporated herein by reference in their entirety. These systems allow the sequencing of many nucleic acid molecules isolated from a specimen at high orders of multiplexing in a parallel fashion (Dear, 2003, *Brief Funct. Genomic Proteomic*, 1(4), 397-416 and McCaughan and Dear, 2010, *J. Pathol.*, 220, 297-306). Each of these platforms allow sequencing of clonally expanded or non-amplified single molecules of nucleic acid fragments. Certain platforms involve, for example, (i) sequencing by ligation of dye-modified probes (including cyclic ligation and cleavage), (ii) pyrosequencing, and (iii) single-molecule sequencing.

[0051] Pyrosequencing is a nucleic acid sequencing method based on sequencing by synthesis, which relies on detection of a pyrophosphate released on nucleotide incorporation. Generally, sequencing by synthesis involves synthesizing, one nucleotide at a time, a DNA strand complimentary to the strand whose sequence is being sought. Study nucleic acids may be immobilized to a solid support, hybridized with a sequencing primer, incubated with DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5' phosphsulfate and luciferin. Nucleotide solutions are sequentially added and removed. Correct incorporation of a nucleotide releases a pyrophosphate, which interacts with ATP sulfurylase and produces ATP in the presence of adenosine 5' phosphsulfate, fueling the luciferin reaction, which produces a chemiluminescent signal allowing sequence determination. Machines for pyrosequencing and methylation specific reagents are available from Qiagen, Inc. (Valencia, CA). See also Tost and Gut, 2007, *Nat. Prot.*

2 2265-2275. An example of a system that can be used by a person of ordinary skill based on pyrosequencing generally involves the following steps: ligating an adaptor nucleic acid to a study nucleic acid and hybridizing the study nucleic acid to a bead; amplifying a nucleotide sequence in the study nucleic acid in an emulsion; sorting beads using a picoliter multiwell solid support; and sequencing amplified nucleotide sequences by pyrosequencing methodology (e.g., Nakano *et al.*, 2003, *J. Biotech.* 102, 117-124). Such a system can be used to exponentially amplify amplification products generated by a process described herein, e.g., by ligating a heterologous nucleic acid to the first amplification product generated by a process described herein.

[0052] Certain single-molecule sequencing embodiments are based on the principal of sequencing by synthesis, and utilize single-pair Fluorescence Resonance Energy Transfer (single pair FRET) as a mechanism by which photons are emitted as a result of successful nucleotide incorporation. The emitted photons often are detected using intensified or high sensitivity cooled charge-couple-devices in conjunction with total internal reflection microscopy (TIRM). Photons are only emitted when the introduced reaction solution contains the correct nucleotide for incorporation into the growing nucleic acid chain that is synthesized as a result of the sequencing process. In FRET based single-molecule sequencing or detection, energy is transferred between two fluorescent dyes, sometimes polymethine cyanine dyes Cy3 and Cy5, through long-range dipole interactions. The donor is excited at its specific excitation wavelength and the excited state energy is transferred, non-radiatively to the acceptor dye, which in turn becomes excited. The acceptor dye eventually returns to the ground state by radiative emission of a photon. The two dyes used in the energy transfer process represent the "single pair", in single pair FRET. Cy3 often is used as the donor fluorophore and often is incorporated as the first labeled nucleotide. Cy5 often is used as the acceptor fluorophore and is used as the nucleotide label for successive nucleotide additions after incorporation of a first Cy3 labeled nucleotide. The fluorophores generally are within 10 nanometers of each other for energy transfer to occur successfully. Bailey *et al.* recently reported a highly sensitive (15pg methylated DNA) method using quantum dots to detect methylation status using fluorescence resonance energy transfer (MS-qFRET)(Bailey *et al.* 2009, *Genome Res.* 19(8), 1455-1461, which is incorporated herein by reference in its entirety).

[0053] An example of a system that can be used based on single-molecule sequencing generally involves hybridizing a primer to a study nucleic acid to generate a complex; associating the complex with a solid phase; iteratively extending the primer by a nucleotide tagged with a fluorescent molecule; and capturing an image of fluorescence resonance energy transfer signals after each iteration (e.g., Braslavsky *et al.*, PNAS 100(7): 3960-3964 (2003); U.S. Pat. No. 7,297,518 (Quake *et al.*) which are incorporated herein by reference in their entirety). Such a

system can be used to directly sequence amplification products generated by processes described herein. In some embodiments the released linear amplification product can be hybridized to a primer that contains sequences complementary to immobilized capture sequences present on a solid support, a bead or glass slide for example. Hybridization of the primer-released linear amplification product complexes with the immobilized capture sequences, immobilizes released linear amplification products to solid supports for single pair FRET based sequencing by synthesis. The primer often is fluorescent, so that an initial reference image of the surface of the slide with immobilized nucleic acids can be generated. The initial reference image is useful for determining locations at which true nucleotide incorporation is occurring. Fluorescence signals detected in array locations not initially identified in the "primer only" reference image are discarded as non-specific fluorescence. Following immobilization of the primer-released linear amplification product complexes, the bound nucleic acids often are sequenced in parallel by the iterative steps of, a) polymerase extension in the presence of one fluorescently labeled nucleotide, b) detection of fluorescence using appropriate microscopy, TIRM for example, c) removal of fluorescent nucleotide, and d) return to step a with a different fluorescently labeled nucleotide.

[0054] The technology may be practiced with digital PCR. Digital PCR was developed by Kalinina and colleagues (Kalinina *et al.*, 1997, *Nucleic Acids Res.* 25; 1999-2004) and further developed by Vogelstein and Kinzler (1999, *Proc. Natl. Acad. Sci. U.S.A.* 96; 9236-9241). The application of digital PCR is described by Cantor *et al.* (PCT Pub. Nos. WO 2005/023091A2 (Cantor *et al.*); WO 2007/092473 A2, (Quake *et al.*)), which are hereby incorporated by reference in their entirety. Digital PCR takes advantage of nucleic acid (DNA, cDNA or RNA) amplification on a single molecule level, and offers a highly sensitive method for quantifying low copy number nucleic acid. Fluidigm® Corporation offers systems for the digital analysis of nucleic acids.

[0055] In some embodiments, nucleotide sequencing may be by solid phase single nucleotide sequencing methods and processes. Solid phase single nucleotide sequencing methods involve contacting sample nucleic acid and solid support under conditions in which a single molecule of sample nucleic acid hybridizes to a single molecule of a solid support. Such conditions can include providing the solid support molecules and a single molecule of sample nucleic acid in a "microreactor." Such conditions also can include providing a mixture in which the sample nucleic acid molecule can hybridize to solid phase nucleic acid on the solid support. Single nucleotide sequencing methods useful in the embodiments described herein are described in PCT Pub. No. WO 2009/091934 (Cantor).

[0056] In certain embodiments, nanopore sequencing detection methods include (a) contacting a nucleic acid for sequencing ("base nucleic acid," e.g., linked probe molecule) with sequence-specific detectors, under conditions in which the detectors specifically hybridize to substantially complementary subsequences of the base nucleic acid; (b) detecting signals from the detectors and (c) determining the sequence of the base nucleic acid according to the signals detected. In certain embodiments, the detectors hybridized to the base nucleic acid are disassociated from the base nucleic acid (e.g., sequentially dissociated) when the detectors interfere with a nanopore structure as the base nucleic acid passes through a pore, and the detectors disassociated from the base sequence are detected.

[0057] A detector also may include one or more regions of nucleotides that do not hybridize to the base nucleic acid. In some embodiments, a detector is a molecular beacon. A detector often comprises one or more detectable labels independently selected from those described herein. Each detectable label can be detected by any convenient detection process capable of detecting a signal generated by each label (e.g., magnetic, electric, chemical, optical and the like). For example, a CD camera can be used to detect signals from one or more distinguishable quantum dots linked to a detector.

[0058] Next generation sequencing techniques may be applied to measure expression levels or count numbers of transcripts using RNA-seq or whole transcriptome shotgun sequencing. See, *e.g.*, Mortazavi *et al.* 2008 *Nat Meth* 5(7) 621-627 or Wang *et al.* 2009 *Nat Rev Genet* 10(1) 57-63.

[0059] Nucleic acids in the invention may be counted using methods known in the art. In one embodiment, NanoString's n Counter system may be used. Geiss *et al.* 2008 *Nat Biotech* 26(3) 317-325; U.S. Pat. No. 7,473,767 (Dimitrov). Alternatively, Fluidigm's Dynamic Array system may be used. Byrne *et al.* 2009 *PLoS ONE* 4 e7118; Helzer *et al.* 2009 *Can Res* 69 7860-7866. For reviews, see also Zhao *et al.* 2011 *Sci China Chem* 54(8) 1185-1201 and Ozsolak and Milos 2011 *Nat Rev Genet* 12 87-98.

[0060] The invention encompasses any method known in the art for enhancing the sensitivity of the detectable signal in such assays, including, but not limited to, the use of cyclic probe technology (Bakkaoui *et al.*, 1996, *BioTechniques* 20 240-8, which is incorporated herein by reference in its entirety); and the use of branched probes (Urdea *et al.*, 1993, *Clin Chem* 39 725-6; which is incorporated herein by reference in its entirety). The hybridization complexes are detected according to well-known techniques in the art.

[0061] Reverse transcribed or amplified nucleic acids may be modified nucleic acids. Modified nucleic acids can include nucleotide analogs, and in certain embodiments include a

detectable label and/or a capture agent. Examples of detectable labels include, without limitation, fluorophores, radioisotopes, colorimetric agents, light emitting agents, chemiluminescent agents, light scattering agents, enzymes and the like. Examples of capture agents include, without limitation, an agent from a binding pair selected from antibody/antigen, antibody/antibody, antibody/antibody fragment, antibody/antibody receptor, antibody/protein A or protein G, hapten/anti-hapten, biotin/avidin, biotin/streptavidin, folic acid/folate binding protein, vitamin B12/intrinsic factor, chemical reactive group/complementary chemical reactive group (e.g., sulfhydryl/maleimide, sulfhydryl/haloacetyl derivative, amine/isotriocyanate, amine/succinimidyl ester, and amine/sulfonyl halides) pairs, and the like. Modified nucleic acids having a capture agent can be immobilized to a solid support in certain embodiments.

[0062] The invention may also be used in conjunction with other molecular tests for bladder cancer such as a fluorescence in situ hybridization (FISH) test such as the Abbott Molecular UroVysion bladder cancer (Abbott Laboratories, Abbott Park, IL, USA). See also US Pat. Pub. No. US 2013/017637 (Giafis *et al.*); US Pat. Nos. 7,232,655 (Halling *et al.*), 6,573,042 (Wang), or 5,663,319 (Bittner *et al.*), the contents of which are hereby incorporated in their entireties.

5.5. Mass Spectroscopic Detection Methods

[0063] Another method for analyzing expression is mass spectroscopy which may involve a primer extension assay, including an optimized PCR amplification reaction that produces amplified targets for analysis using mass spectrometry. The assay can also be done in multiplex. Mass spectrometry is a particularly effective method for the detection of polynucleotides. These methods are particularly useful for performing multiplexed amplification reactions and multiplexed primer extension reactions (e g., multiplexed homogeneous primer mass extension (hME) assays) in a single well to further increase the throughput and reduce the cost per reaction for primer extension reactions.

[0064] For a review of mass spectrometry methods using Sequenom® standard iPLEX(TM) assay and MassARRAY® technology, see Jurinke *et al.*, 2004, *Mol. Biotechnol.* 26 147-164. For methods of detecting and quantifying target nucleic acids using cleavable detector probes that are cleaved during the amplification process and detected by mass spectrometry, see PCT Pub. Nos. WO 2006/031745 (Van Der Boom and Boecker); WO 2009/073251 A1(Van Den Boom *et al.*); WO 2009/114543 A2 (Oeth *et al.*); and WO 2010/033639 A2 (Ehrich *et al.*); which are hereby incorporated by reference in their entirety.

5.6. STATISTICAL METHODS

[0065] The data may be ranked for its ability to distinguish biomarkers in both the 1 versus all (i.e., disease versus normal) and the all-pairwise (i.e., normal versus specific disease) cases. One statistic used for the ranking is the area under the receiver operator characteristic (ROC) curve (a plot of sensitivity versus (1-specificity)). Although biomarkers are evaluated for reliability across datasets, the independent sample sets are not combined for the purposes of the ROC ranking. As a result, multiple independent analyses are performed and multiple independent rankings are obtained for each biomarker's ability to distinguish groups of interest.

[0066] It is to be understood that other genes and/or diagnostic criteria may be used in this invention. For example, patient characteristics, standard blood workups, the results of imaging tests, and/or histological evaluation may optionally be combined with biomarkers disclosed herein.

[0067] Such analysis methods may be used to form a predictive model, and then use that model to classify test data. For example, one convenient and particularly effective method of classification employs multivariate statistical analysis modeling, first to form a model (a "predictive mathematical model") using data ("modeling data") from samples of known class (e.g., from subjects known to have, or not have, a particular class, subclass or grade of lung cancer), and second to classify an unknown sample (e.g., "test data"), according to lung cancer status.

[0068] Pattern recognition (PR) methods have been used widely to characterize many different types of problems ranging for example over linguistics, fingerprinting, chemistry and psychology. In the context of the methods described herein, pattern recognition is the use of multivariate statistics, both parametric and non-parametric, to analyze spectroscopic data, and hence to classify samples and to predict the value of some dependent variable based on a range of observed measurements. There are two main approaches. One set of methods is termed "unsupervised" and these simply reduce data complexity in a rational way and also produce display plots which can be interpreted by the human eye. The other approach is termed "supervised" whereby a training set of samples with known class or outcome is used to produce a mathematical model and is then evaluated with independent validation data sets.

[0069] Unsupervised PR methods are used to analyze data without reference to any other independent knowledge. Examples of unsupervised pattern recognition methods include principal component analysis (PCA), hierarchical cluster analysis (HCA), and non-linear mapping (NLM).

[0070] Alternatively, and in order to develop automatic classification methods, it has proved efficient to use a "supervised" approach to data analysis. Here, a "training set" of biomarker expression data is used to construct a statistical model that predicts correctly the "class" of each

sample. This training set is then tested with independent data (referred to as a test or validation set) to determine the robustness of the computer-based model. These models are sometimes termed "expert systems," but may be based on a range of different mathematical procedures. Supervised methods can use a data set with reduced dimensionality (for example, the first few principal components), but typically use unreduced data, with all dimensionality. In all cases the methods allow the quantitative description of the multivariate boundaries that characterize and separate each class, for example, each class of lung cancer in terms of its biomarker expression profile. It is also possible to obtain confidence limits on any predictions, for example, a level of probability to be placed on the goodness of fit (see, for example, Sharaf; Illman; Kowalski, eds. (1986). Chemometrics. New York: Wiley). The robustness of the predictive models can also be checked using cross-validation, by leaving out selected samples from the analysis.

[0071] Examples of supervised pattern recognition methods include the following 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, (1977) Chemometrics: theory and application 52: 243-282.); partial least squares analysis (PLS) (see, for example, Wold (1966) Multivariate analysis 1: 391-420; Joreskog (1982) Causality, structure, prediction 1: 263-270); linear discriminant analysis (LDA) (see, for example, Nillson (1965). Learning machines. New York.); K-nearest neighbor analysis (KNN) (see, for example, Brown and Martin 1996 *J Chem Info Computer Sci* 36(3):572-584); artificial neural networks (ANN) (see, for example, Wasserman (1993). Advanced methods in neural computing. John Wiley & Sons, Inc; O'Hare & Jennings (Eds.). (1996). Foundations of distributed artificial intelligence (Vol. 9). Wiley); probabilistic neural networks (PNNs) (see, for example, Bishop & Nasrabadi (2006). Pattern recognition and machine learning (Vol. 1, p. 740). New York: Springer; Specht, (1990). Probabilistic neural networks. Neural networks, 3(1), 109-118); rule induction (RI) (see, for example, Quinlan (1986) Machine learning, 1(1), 81-106); and, Bayesian methods (see, for example, Bretthorst (1990). An introduction to parameter estimation using Bayesian probability theory. In Maximum entropy and Bayesian methods (pp. 53-79). Springer Netherlands; Bretthorst, G. L. (1988). Bayesian spectrum analysis and parameter estimation (Vol. 48). New York: Springer-Verlag); unsupervised hierarchical clustering (see for example Herrero 2001 *Bioinformatics* 17(2) 126-136). In one embodiment, the classifier is the centroid based method described in Mullins *et al.* 2007 *Clin Chem* 53(7):1273-9, which is herein incorporated by reference in its entirety for its teachings regarding disease classification.

[0072] It is often useful to pre-process data, for example, by addressing missing data, translation, scaling, weighting, etc. Multivariate projection methods, such as principal component

analysis (PCA) and partial least squares analysis (PLS), are so-called scaling sensitive methods. By using prior knowledge and experience about the type of data studied, the quality of the data prior to multivariate modeling can be enhanced by scaling and/or weighting. Adequate scaling and/or weighting can reveal important and interesting variation hidden within the data, and therefore make subsequent multivariate modeling more efficient. Scaling and weighting may be used to place the data in the correct metric, based on knowledge and experience of the studied system, and therefore reveal patterns already inherently present in the data.

[0073] If possible, missing data, for example gaps in column values, should be avoided. However, if necessary, such missing data may replaced or "filled" with, for example, the mean value of a column ("mean fill"); a random value ("random fill"); or a value based on a principal component analysis ("principal component fill"). Each of these different approaches will have a different effect on subsequent PR analysis.

[0074] "Translation" of the descriptor coordinate axes can be useful. Examples of such translation include normalization and mean centering. "Normalization" may be used to remove sample-to-sample variation. Many normalization approaches are possible, and they can often be applied at any of several points in the analysis. "Mean centering" may be used to simplify interpretation. Usually, for each descriptor, the average value of that descriptor for all samples is subtracted. In this way, the mean of a descriptor coincides with the origin, and all descriptors are "centered" at zero. In "unit variance scaling," data can be scaled to equal variance. Usually, the value of each descriptor is scaled by 1/StDev, where StDev is the standard deviation for that descriptor for all samples. "Pareto scaling" is, in some sense, intermediate between mean centering and unit variance scaling. In pareto scaling, the value of each descriptor is scaled by 1/sqrt(StDev), where StDev is the standard deviation for that descriptor for all samples. In this way, each descriptor has a variance numerically equal to its initial standard deviation. The pareto scaling may be performed, for example, on raw data or mean centered data.

[0075] "Logarithmic scaling" may be used to assist interpretation when data have a positive skew and/or when data spans a large range, e.g., several orders of magnitude. Usually, for each descriptor, the value is replaced by the logarithm of that value. In "equal range scaling," each descriptor is divided by the range of that descriptor for all samples. In this way, all descriptors have the same range, that is, 1. However, this method is sensitive to presence of outlier points. In "autoscaling," each data vector is mean centred and unit variance scaled. This technique is a very useful because each descriptor is then weighted equally and large and small values are treated with equal emphasis. This can be important for analytes present at very low, but still detectable, levels.

[0076] Several supervised methods of scaling data are also known. Some of these can provide a measure of the ability of a parameter (e.g., a descriptor) to discriminate between classes, and can be used to improve classification by stretching a separation. For example, in "variance weighting," the variance weight of a single parameter (e.g., a descriptor) is calculated as the ratio of the inter-class variances to the sum of the intra-class variances. A large value means that this variable is discriminating between the classes. For example, if the samples are known to fall into two classes (e.g., a training set), it is possible to examine the mean and variance of each descriptor. If a descriptor has very different mean values and a small variance, then it will be good at separating the classes. "Feature weighting" is a more general description of variance weighting, where not only the mean and standard deviation of each descriptor is calculated, but other well-known weighting factors, such as the Fisher weight, are used.

[0077] The methods described herein may be implemented and/or the results recorded using any device capable of implementing the methods and/or recording the results. Examples of devices that may be used include but are not limited to electronic computational devices, including computers of all types. When the methods described herein are implemented and/or recorded in a computer, the computer program that may be used to configure the computer to carry out the steps of the methods may be contained in any computer readable medium capable of containing the computer program. Examples of computer readable medium that may be used include but are not limited to diskettes, CD-ROMs, DVDs, ROM, RAM, and other memory and computer storage devices. The computer program that may be used to configure the computer to carry out the steps of the methods and/or record the results may also be provided over an electronic network, for example, over the internet, an intranet, or other network.

[0078] The process of comparing a measured value and a reference value can be carried out in any convenient manner appropriate to the type of measured value and reference value for the discriminative gene at issue. "Measuring" can be performed using quantitative or qualitative measurement techniques, and the mode of comparing a measured value and a reference value can vary depending on the measurement technology employed. For example, when a qualitative colorimetric assay is used to measure expression levels, the levels may be compared by visually comparing the intensity of the colored reaction product, or by comparing data from densitometric or spectrometric measurements of the colored reaction product (e.g., comparing numerical data or graphical data, such as bar charts, derived from the measuring device). However, it is expected that the measured values used in the methods of the invention will most commonly be quantitative values. In other examples, measured values are qualitative. As with qualitative measurements, the

comparison can be made by inspecting the numerical data, or by inspecting representations of the data (e.g., inspecting graphical representations such as bar or line graphs).

[0079] The process of comparing may be manual (such as visual inspection by the practitioner of the method) or it may be automated. For example, an assay device (such as a luminometer for measuring chemiluminescent signals) may include circuitry and software enabling it to compare a measured value with a reference value for a biomarker protein. Alternately, a separate device (e.g., a digital computer) may be used to compare the measured value(s) and the reference value(s). Automated devices for comparison may include stored reference values for the biomarker protein(s) being measured, or they may compare the measured value(s) with reference values that are derived from contemporaneously measured reference samples (e.g., samples from control subjects).

[0080] As will be apparent to those of skill in the art, when replicate measurements are taken, the measured value that is compared with the reference value is a value that takes into account the replicate measurements. The replicate measurements may be taken into account by using either the mean or median of the measured values as the "measured value."

[0081] The invention also includes methods of identifying patients for particular treatments or selecting patients for which a particular treatment would be desirable or contraindicated.

[0082] The methods above be performed by a reference laboratory, a hospital pathology laboratory or a doctor. The methods may be performed as a Laboratory Developed Test (LDT) in a Clinical Laboratory Improvement Amendments (CLIA) approved lab, or an FDA-cleared test as a 510(K). The methods may be performed in a centralized testing labororatory or on a point-of-care (POC) device. The methods above may further comprise an algorithm and/or statistical analysis.

5.7. Compositions and Kits

[0083] The invention provides compositions and kits for identification of high-grade bladder cancer subtypes in a patient sample which comprises: (a) a means for measuring levels of a plurality of high-grade bladder cancer classifier biomarkers; and (b) instructions for comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers from the patient sample with levels of a plurality of high-grade bladder cancer classifier biomarkers for a control patient, wherein the levels of a plurality of high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient.

[0084] The kits may further comprise positive and negative controls, as well as reference markers and instructions for the use of kit components contained therein, in accordance with the methods of the present invention.

5.8. Methods of Treatment

[0085] A patient identified with a basal or luminal high-grade bladder cancer subtype may receive differing treatment for their high-grade bladder cancer. For details regarding bladder cancer patient management in general see the Nat'l Comp. Cancer Network Guidelines ver. 1.2013, Bladder Cancer.

[0086] Similar basal or luminal bladder cancer might receive a treatment similar to that for breast cancer. Thus a patient with a basal subtype would be a candidate for preoperative radiation or neoadjuvant chemotherapy. The chemotherapy regimen may be dense dose methotrexate, vinblastine, doxorubicin, cisplatin (DDMVAC). Grossman *et al.* 2003 *NEJM* 349 859-866 or Sternberg *et al.* 2001 *J Clin Onc* 19 2634-2646. Alternatively, the chemotherapy may gemcitabine or cisplatin. Dash *et al.* 2008 *Cancer* 113 2471-2477 or Von der Maase *et al.* 2000 *J Clin Onc* 18 3068-3077. The chemotherapy regimen might be cisplatin, methotrexate and vinblastine (CMV). Griffiths *et al.* 2011 *J Clin Onc* 29 2171-2177. For patients with non-muscle invasive bladder cancer, given the relatively poor prognosis of the basal-subtype, patients or their physician may choose to perform cystectomy earlier. Based on the BASE47 diagnosis a patient might be determined not to be a good candidate for chemotherapy or radiation.

5.9. Methods to Identify Compounds

[0087] A variety of methods may be used to identify compounds that prevent, slows the progression, or treats bladder cancer. Typically, an assay that provides a readily measured parameter is adapted to be performed in the wells of multi-well plates in order to facilitate the screening of members of a library of test compounds as described herein. Thus, in one embodiment, an appropriate number of cells can be plated into the cells of a multi-well plate, and the effect of a test compound on bladder cancer can be determined. The compounds to be tested can be any small chemical compound, or a macromolecule, such as a protein, sugar, nucleic acid or lipid. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a test compound in this aspect of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays). It will be

appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma- Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

[0088] The term "test compound" or "drug candidate" or "modulator" or grammatical equivalents as used herein describes any molecule, either naturally occurring or synthetic, *e.g.*, protein, oligopeptide, small organic molecule, polysaccharide, peptide, circular peptide, lipid, fatty acid, shRNA, siRNA, polynucleotide, oligonucleotide, etc., to be tested for the capacity to directly or indirectly modulate a genotype or phenotype associated with HGBC. The test compound can be in the form of a library of test compounds, such as a combinatorial or randomized library that provides a sufficient range of diversity. Test compounds are optionally linked to a fusion partner, *e.g.*, targeting compounds, rescue compounds, dimerization compounds, stabilizing compounds, addressable compounds, and other functional moieties. Conventionally, new chemical entities with useful properties are generated by identifying a test compound (called a "lead compound") with some desirable property or activity, *e.g.*, inhibiting activity, creating variants of the lead compound, and evaluating the property and activity of those variant compounds. Often, high throughput screening ("HTS") methods are employed for such an analysis. The compound may be "small organic molecule" that is an organic molecule, either naturally occurring or synthetic, that has a molecular weight of more than about 50 daltons and less than about 2500 daltons, preferably less than about 2000 daltons, preferably between about 100 to about 1000 daltons, more preferably between about 200 to about 500 daltons.

[0089] The term "functional effects" in the context of assays for testing means compounds that modulate a phenotype or a gene associated with HGBC either *in vitro*, in cell culture, in tissue samples, or *in vivo*. This may also be a chemical or phenotypic effect such as altered HGBC profiles *in vivo*, *e.g.*, changing from a high risk of HGBC profile to a low risk profile; altered expression of genes associated with HGBC; altered transcriptional activity of a gene hyper- or hypo-methylated in HGBC; or altered the activities and the downstream effects of proteins encoded by these genes. A functional effect may include transcriptional activation or repression of a gene resulting in changing the ability of cells to proliferate, expression in cells during HGBC progression, and other cellular characteristics. "Functional effects" include *in vitro*, *in vivo*, and *ex vivo* activities. By "determining the functional effect" is meant assaying for a compound that increases or decreases the transcription of genes or the translation of proteins that are indirectly or directly associated with HGBC. Such functional effects can be measured by any means known to those skilled in the art, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbance, refractive index); hydrodynamic (*e.g.*, shape), chromatographic (*e.g.*, elution time); or

solubility properties for the protein; ligand binding assays, *e.g.*, binding to antibodies; measuring inducible markers or transcriptional activation of the marker; measuring changes in enzymatic activity; the ability to increase or decrease cellular proliferation, apoptosis, cell cycle arrest, measuring changes in cell surface markers. Validation of the functional effect of a compound on HGBC occurrence or progression can also be performed using assays known to those of skill in the art such as studies using mouse models. The functional effects can be evaluated by many means known to those skilled in the art, *e.g.*, microscopy for quantitative or qualitative measures of alterations in morphological features, measurement of changes in RNA or protein levels for other genes associated with HGBC, measurement of RNA stability, identification of downstream or reporter gene expression (CAT, luciferase, β-gal, GFP, and the like), *e.g.*, via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, proteomics, metabolomics, etc.

[0090] "Inhibitors," "activators," and "modulators" of the markers are used to refer to activating, inhibitory, or modulating molecules identified using *in vitro* and *in vivo* assays of the expression of genes hyper- or hypo-methylated in HGBC, mutations associated with HGBC, or the translation of proteins encoded thereby. Inhibitors, activators, or modulators also include naturally occurring and synthetic ligands, antagonists, agonists, antibodies, peptides, cyclic peptides, nucleic acids, antisense molecules, ribozymes, asRNAs, shRNAs, RNAi molecules, small organic molecules and the like. Such assays for inhibitors and activators include, *e.g.*, (1)(a) the mRNA expression, or (b) proteins expressed by genes hyper- or hypo-methylated in HGBC *in vitro*, in cells, or cell extracts; (2) applying putative modulator compounds; and (3) determining the functional effects on activity, as described above.

[0091] Assays comprising *in vivo* measurement of HGBC; or genes hyper- or hypo-methylated in HGBC are treated with a potential activator, inhibitor, or modulator are compared to control assays without the inhibitor, activator, or modulator to examine the extent of inhibition. Controls (untreated) are assigned a relative activity value of 100%. Inhibition of gene expression, protein expression associated with HGBC is achieved when the activity value relative to the control is about 80%, preferably 50%, more preferably 25-0%. Activation of gene expression, or proteins associated with HGBC is achieved when the activity value relative to the control (untreated with activators) is 110%, more preferably 150%, more preferably 200- 500% (*i.e.*, two to five fold higher relative to the control), more preferably 1000-3000% higher.

[0092] In one preferred embodiment, high throughput screening methods are used which involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds. Such "combinatorial chemical libraries" or "ligand libraries" are

then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. In this instance, such compounds are screened for their ability to modulate the ApoE isoforms associated with AD. A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[0093] Preparation and screening of combinatorial chemical libraries are well-known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, *e.g.*, U.S. Pat. No. 5,010,175 (Rutter and Santi), Furka, *Int. J. Pept. Prot. Res.*, 37:487-493 (1991); and Houghton *et al.*, *Nature*, 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: U.S. Pat. Nos. 6,075,121 (Bartlett *et al.*) peptoids; 6,060,596 (Lerner *et al.*) encoded peptides; 5,858,670 (Lam *et al.*) random bio-oligomers; 5,288,514 (Ellman) benzodiazepines; 5,539,083 (Cook *et al.*) peptide nucleic acid libraries; 5,593,853 (Chen and Radmer) carbohydrate libraries; 5,569,588 (Ashby and Rine) isoprenoids; 5,549,974 (Holmes) thiazolidinones and metathiazanones; 5,525,735 (Takarada *et al.*) and 5,519,134 (Acevado and Hebert) pyrrolidines; 5,506,337 (Summerton and Weller) morpholino compounds; 5,288,514 (Ellman) benzodiazepines; diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, 1993, *Proc. Nat. Acad. Sci. USA*, 90, 6909-6913), vinylogous polypeptides (Hagihara *et al.*, 1992, *J. Amer. Chem. Soc.*, 114, 6568), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, 1992, *J. Amer. Chem. Soc.*, 114, 9217-9218), analogous organic syntheses of small compound libraries (Chen *et al.*, 1994, *J. Amer. Chem. Soc.*, 116:2661 (1994)), oligocarbamates (Cho *et al.*, 1993, *Science*, 261, 1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, 1994, *J. Org. Chem.*, 59:658), nucleic acid libraries (see Ausubel, Berger and Sambrook, all *supra*); antibody libraries (see, *e.g.*, Vaughn *et al.*, 1996, *Nat. Biotech.*, 14(3):309-314, carbohydrate libraries, *e.g.*, Liang *et al.*, 1996, *Science*, 274:1520-1522, small organic molecule libraries (see, *e.g.*, benzodiazepines, Baum, 1993, C&EN, Jan 18, page 33. Devices for the preparation of combinatorial libraries are commercially available (see, *e.g.*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433 A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous

combinatorial libraries are themselves commercially available (see, *e.g.*, ComGenex (Princeton, NJ), Asinex (Moscow, RU), Tripos, Inc. (St. Louis, MO), ChemStar, Ltd., (Moscow, RU), 3D Pharmaceuticals (Exton, PA), Martek Biosciences (Columbia, MD), etc.).

[0094] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The article "a" and "an" are used herein to refer to one or more than one (i.e., to at least one) of the grammatical object(s) of the article. By way of example, "an element" means one or more elements.

[0095] Throughout the specification the word "comprising," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps. The present invention may suitably "comprise", "consist of", or "consist essentially of", the steps, elements, and/or reagents described in the claims.

[0096] It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely", "only" and the like in connection with the recitation of claim elements, or the use of a "negative" limitation.

[0097] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0098] The following Examples further illustrate the invention and are not intended to limit the scope of the invention. In particular, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

6. EXAMPLES

[0099] We have identified two intrinsic, molecular subsets of high-grade bladder cancer: "luminal" and "basal-like" with differences in clinical outcome. In addition we have developed a 47-gene predictor, "BASE47", which can accurately classify high-grade UC into luminal and basal-like tumors. The molecular subtypes appear to reflect different stages of urothelial differentiation and recapitulate many aspects of breast cancer biology. As an appreciation of subtype heterogeneity has revolutionized the care of breast cancer, these results also suggest stratification for therapy is indicated in bladder cancer as well.

[00100] **MATERIALS AND METHODS**

[00101] *Training Dataset Analysis* – A meta-dataset was generated by combining the muscle invasive (≥T2) UC samples from four publically available data sets (GSE13507, GSE31684, GSE32894, GSE5287) with clinical annotation provided by the Michor Lab (Dana-Farber Cancer Institute, Boston MA). The data were normalized, median centered by gene, and merged into a single dataset consisting of n=262 tumors. The Mean Absolute Deviation (MAD) was computed across samples by gene. Genes with a MAD score of >0.10 were selected for clustering analysis (7303 genes). Consensus hierarchical clustering was performed as described previously (20) with 90% resampling and 1000 iterations. Two Class significance analysis of microarrays (SAM; FDR=0) was performed to generate subtype-specific gene lists (21). Gene set enrichment analysis (GSEA) was performed comparing basal and luminal tumors against MSigDBv4.0c2 (22,23).

[00102] *Validation Dataset* – Gene expression data were derived from 49 high-grade tumors from Memorial Sloan-Kettering Cancer Center (MSKCC) using Human HT-12 Expression BeadChip arrays (Illumina) as previously described (8). The MSKCC dataset was normalized, median centered and the MAD was computed across samples by gene. Genes with a MAD score of >0.10 were selected for clustering analysis (3357 genes). Consensus clustering was performed identically to the meta-dataset (20). The resulting subtypes assignments for K=2 were used to validate the training dataset. Centroids were generated for both the Meta and MSKCC datasets using all common genes and correlations were calculated by 1-Pearson correlation. Copy number alterations and hotspot mutation analyses were determined as previously described(8).

[00103] *Subtype Predictor* –Prediction Analysis of Microarrays (PAM) was used to determine the minimal number of genes that could accurately predict subtype classification on the meta-dataset using the consensus clustering calls as the reference (24). The resulting 47-gene predictor (delta=6.3) was then used to classify the MSKCC samples (24). Tumors were then analyzed for enrichment of mutations or copy number alteration (8) by chi square or fisher's exact test when appropriate. Categoric survival analyses were performed using a log-rank test and visualized with

Kaplan-Meier plots. The BASE47 was then applied to superficial tumors, which were excluded from the Meta dataset. The superficial were normalized and median centered as previously described and BASE47 calls were made using PAM.

[00104] RESULTS

[00105] Consensus Cluster reveals two distinct molecular subtypes of high-grade bladder cancer.

[00106] While previous studies have examined the gene expression changes associated with bladder cancer (5,13,14,25,26) (27) few have looked exclusively for intrinsic subtypes of high-grade disease. We therefore asked whether there were definable subtypes of high-grade bladder cancer agnostic to clinical stage or outcome. We first created a meta-dataset of 262 high grade, muscle-invasive tumors, curated from four publically available datasets (GSE13507 (26), GSE31684 (13), GSE32894 (25), GSE5287 (28)) (Table 1). In parallel, we used an independent, unique set of high-grade tumors from MSKCC as a validation set (n=49) (8). In both the meta-dataset as well as the MSKCC tumors, Consensus Cluster identified two groups (K=2) as the optimal number of molecular subtypes as defined by the criterion of subclass stability (Figure 1A, 1B).

[00107] While two subtypes were discovered in both datasets, it is possible that the gene expression patterns that define the subtypes are different. We therefore determined the correlation between the centroid gene expression (using all common genes between datasets) for each subtype (light gray = correlation, dark gray=anti-correlation) (Figure 1C). There appeared to be a high level of correlation between the meta-dataset Cluster 1 (K1) and MSKCC Cluster 2 (K2) as well as the meta-dataset Cluster 2 (K2) and MSKCC Cluster 1 (K1). Therefore, the intrinsic molecular subtypes defined by independent discovery in the two datasets are defined by the highly concordant gene expression patterns.

[00108] The intrinsic molecular subtypes of bladder cancer differentially express markers of urothelial differentiation.

[00109] To understand the gene expression patterns that differentiate the intrinsic subtypes of high-grade bladder cancer we performed 2-class significance analysis of microarrays (SAM) comparing Cluster 1 and Cluster 2 from the meta-dataset. 2,393 genes were found to be differentially expressed using an FDR cut off of 0 (Table 2 shows the first 200). The two intrinsic molecular subtypes were characterized by gene expression patterns representative of urothelial differentiation. Cluster 1 (K1) of the meta-dataset, expressed high levels of the high molecular weight keratins [HMWK] (*KRT14*, *KRT5*, *KRT6B*) and *CD44*, which have been previously described to be expressed in urothelial basal cells (29,30). In contrast, Cluster 2 (K2) expressed

high levels of uroplakins (*UPK1B*, *UPK2*, *UPK3A*) as well as the low molecular weight keratins (LMWK), *KRT20* (Figure 1D), characteristic of urothelial umbrella cells (30). Moreover, the gene expression of *KRT5* was inversely correlated with both *UPK2* and *KRT20* across all tumors (Figure 1E and 1F). Similar findings were seen in the MSKCC dataset (Figure 2D). In aggregate, these findings demonstrate that the two molecular subtypes of muscle-invasive bladder cancer represent different stages of urothelial differentiation. Based on these findings we have named Clusters 1 and 2 "Basal-like" and "Luminal", respectively.

[00110] Bladder Cancer Analysis of Subtypes by Expression of 47 genes (BASE47) can accurately predict basal and luminal bladder cancer.

[00111] We next sought to define a minimal set of genes that could accurately classify bladder tumors into the luminal and basal-like bladder intrinsic subtypes. To this end, we applied prediction analysis of microarrays (PAM) to our meta-dataset and derived a 47-gene signature that could accurately classify basal-like and luminal tumors relative to Consensus Cluster calls (training error rate = 0.11 and, 0.05 respectively) (Figure 3A and Table 3). A pairwise comparison of the subtype classification by Consensus Cluster relative to classification by BASE47, showed a strong correlation in the meta-dataset (chi square p= 1.0e-6). A parallel analysis in the MSKCC validation set of tumors showed similar results with 16/16 basal-like tumors being predicted as basal-like, while 24/33 luminal tumors were predicted as luminal. (chi square p= 0.001). Cluster analysis and the corresponding heatmap illustrate the gene expression patterns that comprise BASE47 (Figure 3B).

[00112] Basal-like bladder cancer is enriched for the signatures of basal-like breast cancer and tumor initiating cells (TIC).

[00113] We noted that a number of genes fundamental for breast development and breast cancer were co-regulated with genes that regulate urothelial development such as the uroplakins (Figure 6, breast cancer related genes: dark gray/black, Urothelial related genes: darker gray). Moreover, when Gene Set Enrichment Analysis (GSEA) was performed on the meta-dataset to identify known gene sets enriched in the basal-like and luminal bladder subtypes multiple breast cancer-derived gene signatures were enriched in the basal-like bladder cancer subtype as well as signatures related to mammary stem cells (data not shown). Conversely, multiple breast cancer derived luminal gene signatures were enriched in the luminal bladder cancer subtype.

[00114] Because of the enrichment of the mammary stem cell signature in the basal-like subtype and the fact that the transcriptional programs that govern maintenance of a stem cell like state in tissue and cancer stem cells are often conserved, we asked whether a previously described bladder TIC signature (29) was enriched in one of the molecular subtypes. The basal-like subtype

was significantly associated with the activated bladder TIC signature by both hierarchical clustering (chi squared p= $2x10^{-16}$) (Figure 2B) as well as by GSEA (Enrichment Score [ES] 0.76 and nominal p=0.006) suggesting that basal-like bladder cancer possesses a more "stem-like" phenotype similar to previous observations described in basal-like breast cancer (31).

[00115] The intrinsic bladder subtypes reflect the attributes of molecular subtypes of breast cancer

[00116] We next asked whether the basal-like and luminal bladder cancer subtypes correlated with any of the previously defined molecular subtypes of breast cancer (32,33). Hierarchical clustering of the bladder tumors using a comprehensive list of 1906 genes (1,426 were present in the meta-dataset) that have been previously shown to define the intrinsic subtypes of breast cancer (34) clustered the bladder tumors along the lines of basal-like and luminal bladder subtypes (chi squared p=2.2e-16) (Figure 2C) showing that the gene expression patterns that distinguish basal-like and luminal bladder cancer are present in the breast cancer intrinsic gene list.

[00117] We next generated breast molecular subtype classifications on two independent sets of breast tumors (TCGA Breast (35) and UNC337 (36)) using the PAM50 nearest centroid classifier (34). To see whether the gene expression patterns of luminal and basal-like bladder cancer were reflected in the intrinsic breast subtypes we correlated the centroid gene expression (using the breast intrinsic gene list) between the bladder (bladder tumors) and breast (breast tumors) subtypes (Figure 2D: light gray = correlation, dark gray = anti-correlation). Basal-like bladder cancer had positive correlations to basal-like breast as well as normal-like breast whereas luminal bladder cancer had positive correlations to both luminal A and luminal B breast intrinsic subtypes. Indeed, when the PAM50 was applied to our meta-dataset of bladder tumors, there were positive correlations between basal-like bladder tumors and the basal centroid (Figure 7) and luminal bladder tumors and the luminal A centroid.

[00118] Finally, we analyzed specific gene signatures previously shown to be representative of basal-like and luminal breast cancer, as well as signatures relating to immune response and oncogenic pathways in breast cancer (Figure 8). The breast signatures faithfully clustered basal-like and luminal bladder tumors, with basal-like tumors associating with an increased immune response and MYC/E2F3 pathways. In aggregate, the above findings demonstrate that the gene expression patterns that define the intrinsic molecular subtypes of bladder cancer reflect the attributes that specify the intrinsic breast subtypes.

[00119] Luminal and basal-like bladder cancer have differential survival and are associated with distinct genomic alterations.

[00120] We next asked whether the intrinsic bladder subtypes, which were defined in an unbiased manner, had prognostic significance. 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. While the BASE47 predictor was developed on muscle-invasive tumors we also noted that when applied to a meta-dataset of superficial tumors, it classified a significant proportion them as basal-like (Figure 4B). Further studies will be needed to determine the clinical implications of this finding.

[00121] The MSKCC tumors have been previously sequenced for a panel of bladder cancer relevant genetic alterations as well as copy number alterations (CNAs) (8). To examine whether the intrinsic subtypes correlate with a specific mutational spectrum or CNAs, we examined the relative enrichment of these molecular events in the bladder subtypes (Figure 4C). Notably, FGFR3 ($p<0.001$) and TSC-1 ($p=0.02$) mutations were significantly enriched in luminal bladder cancer while RB1 pathway alterations (*RB1, CCND1, E2F3, CCNE*) were significantly enriched in basal-like bladder cancer ($p = 0.009$).

[00122] Finally, multiple studies have shown that females have a poorer bladder cancer specific outcome than males (37). We saw that there was a trend towards enrichment of basal-like tumors in female patients in the MSKCC dataset (Figure 4C) and a significantly higher incidence of basal-like bladder cancer in female patients in the meta-dataset with annotated gender.

[00123] DISCUSSION

[00124] Using independent discovery in distinct datasets, we have defined two molecular subsets of high-grade urothelial carcinoma. The subtypes harbor molecular features that reflect different stages of urothelial differentiation. Luminal bladder cancers express markers of terminal urothelial differentiation such as those seen in umbrella cells (UPK1B, UPK2, UPK3A, and KRT20) while basal-like tumors express high levels of genes that typically mark urothelial basal cells (KRT14, KRT5, and KRT6B). The basal cell compartment is a common feature of most organs with stratified or pseudostratified epithelium. It is characterized by its proximity to the basal lamina and is thought to harbor multipotent tissue stem cells important for normal tissue homeostasis and orderly regeneration after injury. Because basal cells are a long-lived population, they are potentially more likely to incur multiple genomic alterations including changes in their chromatin landscape. In this regard it is interesting to note that there appears to be a relatively high prevalence of mutations in histone and chromatin modifying genes in urothelial carcinoma (3).

[00125] The luminal and basal-like subtypes of bladder cancer reflect many of the hallmarks of the intrinsic breast cancer subtypes. For example, we noted that a number of basal-like and luminal breast cancer specific gene signatures were enriched in the corresponding bladder subtype including *bona fide* luminal breast cancer pathways such as GATA3 and estrogen receptor signaling in the luminal bladder subtype. Moreover, the gene expression patterns that define luminal and basal-like bladder cancers corresponded highly with the gene expression patterns that define luminal (Lum A and Lum B) and basal-like breast cancer. These similarities may reflect the presence of urothelial basal cells and their corollary, the basal/myoepithelial cells of the breast. In both tissues, these basal cells represent a multipotent "stem/progenitor cell" population (38,39) and their similar functional roles may explain their similar molecular profile.

[00126] Our study has created a gene signature, the BASE47, which accurately discriminates intrinsic bladder subtypes. Interestingly, even in superficial bladder tumors, there appears to be a significant number of basal-like tumors. While the characteristics of our meta-dataset did not allow us to determine whether the subtypes were prognostic or predicted the development of muscle-invasive disease in superficial bladder tumors, these will be important questions to answer in the future. The ability to accurately classify basal-like and luminal bladder subtypes with only 47 genes (BASE47) should allow the adoption of the BASE47 to formalin-fixed, paraffin embedded (FFPE) tissues allowing its widespread use.

[00127] Female patients with UC have inferior outcomes to males, even when controlled for other known prognostic variables, such as stage and grade (37). Interestingly, we found that females have an increased incidence of basal-like bladder cancer, which is associated with a worse outcome. To what extent this increased prevalence of basal-like bladder tumors in women contributes to their poorer outcome remains unclear. However, it will be of interest to see how epidemiologic variables that have been associated with basal-like breast cancer such as race, parity, and age are related to our intrinsic bladder cancer subtypes.

[00128] Finally, although bladder cancer is a chemo-sensitive disease, with neoadjuvant cisplatin-based chemotherapy for MIBC associated with a 14% reduction in the risk of death and 5% absolute survival benefit at 5 years (40), cisplatin is associated with significant toxicity in this older patient population with no available biomarkers to select those patients who will benefit from chemotherapy in the neoadjuvant or metastatic setting. In basal-like *breast* cancer, although the prognosis is poor, these tumors are more sensitive to neoadjuvant chemotherapy than luminal breast tumors, the so-called "triple negative paradox" (41). It will be interesting to see whether basal-like *bladder* tumors are more responsive to chemotherapy as well and whether the basal-like

molecular subtype may be used as a predictive biomarker for neoadjuvant chemotherapy in MIBC.

[00129] In summary, the basal-like and luminal intrinsic subtypes of bladder cancer reflect many aspects of physiologic urothelial development as well as breast cancer biology. These findings underscore the notion that there are common themes underlying the development and maintenance of solid tumors that extend beyond overlapping mutational spectra. An appreciation of subtype heterogeneity has revolutionized the care of breast cancer, our results suggest that stratification for therapy is indicated in bladder cancer as well.

7. REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013;63:11–30.

2. Goebell PJ, Knowles MA. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. Urol Oncol. 2010;28:409–28.

3. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet. 2011;43:875–8.

4. Lindgren D, Sjödahl G, Lauss M, Staaf J, Chebil G, Lövgren K, et al. Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. PLoS One. 2012;7:e38863.

5. Lindgren D, Frigyesi A, Gudjonsson S, Sjodahl G. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and …. Cancer Res. 2010.

6. López-Knowles E, Hernández S, Malats N, Kogevinas M, Lloreta J, Carrato A, et al. PIK3CA mutations are an early genetic alteration associated with FGFR3 mutations in superficial papillary bladder tumors. Cancer Res. 2006;66:7401–4.

7. Sjödahl GG, Lauss MM, Gudjonsson SS, Liedberg FF, Halldén CC, Chebil GG, et al. A systematic study of gene mutations in urothelial carcinoma; inactivating mutations in TSC2 and PIK3R1. PLoS One. 2011;6:e18583–3.

8. Iyer G, Al-Ahmadie H, Schultz N, Hanrahan AJ, Ostrovnaya I, Balar AV, et al. Prevalence and Co-Occurrence of Actionable Genomic Alterations in High-Grade Bladder Cancer. J Clin Oncol. 2013.

9. Dinney CPN, McConkey DJ, Millikan RE, Wu X, Bar-Eli M, Adam L, et al. Focus on bladder cancer. Cancer Cell. 2004;6:111–6.

10. Wu X-R. Urothelial tumorigenesis: a tale of divergent pathways. Nat. Rev. Cancer. 2005;5:713–25.

11. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. J Clin Oncol. 2006;24:778–89.

12. Lindgren D, Frigyesi A, Gudjonsson S, Sjödahl G, Hallden C, Chebil G, et al. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. Cancer Res. 2010;70:3463–72.

13. Riester M, Taylor JM, Feifer A, Koppie T, Rosenberg JE, Downey RJ, et al. Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. Clin Cancer Res. 2012;18:1323–33.

14. Blaveri E, Simko JP, Korkola JE, Brewer JL, Baehner F, Mehta K, et al. Bladder cancer outcome and subtype classification by gene expression. Clin Cancer Res. 2005;11:4044–55.

15. Dyrskjøt LL, Thykjaer TT, Kruhøffer MM, Jensen JLJ, Marcussen NN, Hamilton-Dutoit SS, et al. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet. 2003;33:90–6.

16. Smith SC, Baras AS, Dancik G, Ru Y, Ding K-F, Moskaluk CA, et al. A 20-gene model for molecular nodal staging of bladder cancer: development and prospective assessment. Lancet Oncol. 2011;12:137–43.

17. Lee J-S, Leem S-H, Lee S-Y, Kim S-C, Park E-S, Kim S-B, et al. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. J Clin Oncol. 2010;28:2660–7.

18. Modlich OO, Prisack H-BH, Pitschke GG, Ramp UU, Ackermann RR, Bojar HH, et al. Identifying superficial, muscle-invasive, and metastasizing transitional cell carcinoma of the bladder: use of cDNA array analysis of gene expression profiles. Clin Cancer Res. 2004;10:3410–21.

19. Kim W-J, Kim S-K, Jeong P, Yun S-J, Cho I-C, Kim IY, et al. A four-gene signature predicts disease progression in muscle invasive bladder cancer. Mol Med. 2011.

20. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.

21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U.S.A. 2001;98:5116–21.

22. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34:267–73.

23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. 2005;102:15545–50.

24. 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.

25. Sjödahl G, Lauss M, Lövgren K, Chebil G, Gudjonsson S, Veerla S, et al. A molecular taxonomy for urothelial carcinoma. Clin Cancer Res. 2012;18:3377–86.

26. Kim W-J, Kim E-J, Kim S-K, Kim Y-J, Ha Y-S, Jeong P, et al. Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. Mol Cancer. 2010;9:3.

27. Sanchez-Carbayo M, Socci N, Lozano J, Li W, Charytonowicz E, Belbin T, et al. Gene discovery in bladder cancer progression using cDNA microarrays. Am J Pathol. 2003;163:505–16.

28. Als AB, Dyrskjøt L, Maase von der H, Koed K, Mansilla F, Toldbod HE, et al. Emmprin and survivin predict response and survival following cisplatin-containing chemotherapy in patients with advanced bladder cancer. Clin Cancer Res. 2007;13:4407–14.

29. Chan KS, Espinosa I, Chao M, Wong D, Ailles L, Diehn M, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. Proceedings of the National Academy of Sciences. 2009;106:14016–21.

30. Castillo-Martin M, Domingo-Domenech J, Karni-Schmidt O, Matos T, Cordon-Cardo C. Molecular pathways of urothelial development and bladder tumorigenesis. Urol Oncol. 2010;28:401–8.

31. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. 2009;15:907–13.

32. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.

33. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U.S.A. 2001;98:10869–74.

34. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160–7.

35. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

36. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12:R68.

37. Donsky H, Coyle S, Scosyrev E, Messing EM. Sex differences in incidence and mortality of bladder and kidney cancers: National estimates from 49 countries. Urol Oncol. 2013.

38. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol [Internet]. 2008;26:2568–81. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=18487574&retmode=ref&cmd=prlinks

39. Kurzrock EA, Lieu DK, Degraffenried LA, Chan CW, Isseroff RR. Label-retaining cells of the bladder: candidate urothelial stem cells. Am. J. Physiol. Renal Physiol. 2008;294:F1415–21.

40. Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. Eur Urol. 2005;48:202–5–discussion205–6.

41. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13:2329–34.

Table 1: Dataset Characteristics

	Training Datasets (Meta)				Validation Dataset
	Als	Kim	Riester	Sjödahl	Iyer
GEO ID	GSE5287	GSE13507	GSE31684	GSE32894	cBioPortal
Clinical Characteristic	No.	No.	No.	No.	No.
Sex					
Male	NA	NA	57	68	35
Female	NA	NA	21	25	14
Stage					
pT0	0	23	5	116	4
pT1	0	80	10	97	6
pT2	0	31	17	85	5
pT3	0	19	42	7	17
pT4	30	11	19	1	16
NA	0	0	0	0	1

Table 2 (The BASE47 classifier genes are shown in bold)

Two Class SAM Basal vs Luminal FDR=0 Basal Positive Genes

Gene ID	Score(d)	Numerator(r)	Denominator(s+s0)	Fold Change	q-value(%)
CHST15	9.281429412	0.823265373	0.088700278	1.769406309	0
EMP3	9.076449488	0.659469062	0.072657162	1.579501231	0
CORO1C	8.979612828	0.488520876	0.054403334	1.403005706	0
AHNAK2	8.871419905	1.100118422	0.124007029	2.143722883	0
CLIC4	8.871284458	0.519597763	0.058570748	1.433555503	0
CDK6	8.85803396	0.674245576	0.076116843	1.595762088	0
MSN	8.844014145	0.695487617	0.078639361	1.619431693	0
SACS	8.73941002	0.44439121	0.050849109	1.360739789	0
PDGFC	8.619555882	0.654870585	0.07597498	1.57447471	0
PALLD	8.605600952	0.643071577	0.074727097	1.561650462	0
GLIPR1	8.567820429	0.666614799	0.077804478	1.587343983	0

MT1X	8.558313751	1.042815151	0.121848203	2.060243916	0
TUBB6	8.465595621	0.673464576	0.079553124	1.594898459	0
OSMR	8.350219756	0.495092521	0.059290957	1.409411139	0
PRKCDBP	8.332713249	0.72462016	0.08696089	1.652465509	0
CD14	8.300941501	0.756055736	0.091080721	1.68886702	0
DEGS1	8.250068549	0.568037535	0.068852462	1.48250558	0
PRRX1	8.24163799	0.812504934	0.09858537	1.756258167	0
FLNA	8.220707206	0.331269159	0.040296917	1.258119673	0
LRIG1	8.205032301	0.552204034	0.067300653	1.466324118	0
DPYD	8.171017138	0.566823556	0.069370012	1.481258627	0
ANXA5	8.10571954	0.564231798	0.069609095	1.478599979	0
PTGS1	8.088083663	0.475088861	0.05873936	1.390003833	0
ALOX5AP	8.084986513	0.897382454	0.110993686	1.862683366	0
NOD2	8.065137085	0.572746392	0.071015085	1.487352281	0
FAP	8.051000136	0.807959709	0.100355197	1.750733764	0
DSE	8.04723211	0.655774396	0.081490677	1.575461388	0
CERK	8.037135638	0.412361251	0.051306992	1.330862247	0
TNC	8.001358244	0.753878755	0.094218848	1.686320496	0
MPP1	7.955402627	0.477533566	0.060026323	1.392361247	0
TLR1	7.950560535	0.586962915	0.073826608	1.502081314	0
MT2A	7.912812393	1.030399646	0.130219143	2.042589998	0
TNFAIP6	7.887040266	0.80553086	0.102133479	1.747788797	0
SNAI2	7.773334472	0.686646707	0.088333611	1.609538077	0
RRAS	7.713193222	0.513763275	0.066608376	1.427769685	0
FAS	7.696921523	0.383114317	0.049775006	1.304154071	0
KCTD12	7.685036877	0.33982247	0.044218717	1.265600847	0
KIAA0922	7.650279639	0.363704951	0.047541393	1.286726069	0
ST3GAL6	7.614594689	0.401516934	0.052729915	1.320896048	0
CD44	7.596907234	0.61294349	0.080683293	1.529376371	0
CD86	7.596481363	0.406327366	0.053488891	1.325307712	0
MT1M	7.588992738	0.651728693	0.085878155	1.571049561	0
C1S	7.570437932	0.582109084	0.076892392	1.497036176	0
SIRPA	7.567005842	0.528134575	0.069794392	1.442063382	0
PTRF	7.561215386	0.615679218	0.081425959	1.532279221	0
ITGA5	7.513903831	0.645714213	0.085935917	1.564513614	0
ZEB2	7.510579707	0.429606133	0.05720013	1.346865822	0
LY96	7.504085324	0.647870723	0.086335735	1.566853965	0
LAMP2	7.503515087	0.394445326	0.052568073	1.314437301	0
HCK	7.500030458	0.565966575	0.075461904	1.480379	0
EFEMP1	7.441958844	0.680372565	0.091423855	1.602553549	0
PRNP	7.441286966	0.651755756	0.087586429	1.571079032	0
CAV1	7.401535397	0.579524152	0.078297829	1.494356279	0
TGFBI	7.341884909	0.75211343	0.102441463	1.684258324	0
SLA	7.328543343	0.408744567	0.055774326	1.327530095	0

ATXN1	7.328390404	0.350960471	0.047890526	1.275409446	0
AXL	7.32051532	0.562649475	0.076859271	1.476979165	0
FPR3	7.319965575	0.575366114	0.078602298	1.490055548	0
CSF1R	7.295936676	0.677598064	0.092873348	1.599474579	0
IL15RA	7.257772176	0.35269666	0.048595719	1.276945241	0
FCGR2A	7.228522872	0.466943359	0.064597341	1.38217794	0
SAMSN1	7.226425267	0.455944797	0.0630941	1.371680802	0
NMT2	7.2258854	0.382055678	0.052873199	1.303197442	0
SPI1	7.22323439	0.434034451	0.060088657	1.351006346	0
IFITM2	7.21635042	0.562541756	0.077953775	1.47686889	0
SERPINA3	7.20588124	1.064301733	0.147699039	2.091157511	0
NCF2	7.187023002	0.485901806	0.067608216	1.400460999	0
RNASE6	7.183406108	0.521197222	0.072555723	1.435145711	0
SYNC	7.173635118	0.481632741	0.067139286	1.396323033	0
RGS2	7.15446335	0.643183673	0.089899639	1.561771806	0
C1QA	7.150504	0.519746106	0.072686639	1.433702914	0
AP1S2	7.144685226	0.477050062	0.06676992	1.391894691	0
NFIL3	7.141085905	0.44299157	0.062034203	1.359420299	0
TMEM45A	7.140582547	0.888447322	0.124422247	1.851182745	0
FCER1G	7.132730967	0.707919783	0.099249472	1.633447157	0
SPHK1	7.126205749	0.628868746	0.088247346	1.546351984	0
TIMP2	7.107453403	0.515955058	0.07259352	1.429940441	0
WIPF1	7.101983055	0.457422078	0.064407655	1.373086085	0
C5AR1	7.0873843	0.55516356	0.078331234	1.469335204	0
CD163	7.084608927	0.65248272	0.092098622	1.571870888	0
STX2	7.077648146	0.538665884	0.076108034	1.452628594	0
PMEPA1	7.076769773	0.535069268	0.075609252	1.449011724	0
SLC7A7	7.071523952	0.614561216	0.086906474	1.531092257	0
MYO5A	7.057334814	0.395942924	0.056103747	1.315802469	0
GPR68	7.053575337	0.558586353	0.079191945	1.472825339	0
MS4A4A	7.039380146	0.41716039	0.059260955	1.335296745	0
ZNF532	7.011202738	0.330339074	0.047115892	1.257308843	0
TYMP	7.003854762	0.554515959	0.079172967	1.468675793	0
GAS1	6.977440558	0.686341609	0.098365812	1.609197731	0
KIAA0226L	6.966468326	0.304229203	0.043670507	1.234758759	0
COL5A2	6.960460377	0.686729519	0.098661508	1.609630468	0
CASP1	6.960163516	0.603794702	0.086750074	1.519708584	0
RAB27A	6.959529673	0.261059977	0.037511152	1.19835884	0
NXN	6.9541317	0.556008631	0.07995371	1.470196132	0
SRGN	6.953905276	0.749260797	0.107746765	1.680931341	0
SLAMF8	6.953662591	0.472081071	0.067889557	1.387108914	0
FYB	6.952772131	0.445596377	0.064089024	1.361876969	0
COL6A2	6.940376246	0.468512537	0.067505351	1.383682113	0
SERPINA1	6.935974416	0.595469529	0.085852325	1.510964253	0

CDA	6.934941405	0.573409922	0.082684177	1.488036507	0
ADCY7	6.907413219	0.267078589	0.0386655	1.203368576	0
NR3C1	6.898860662	0.213111951	0.030890891	1.1591859	0
DACT1	6.888980974	0.465951714	0.067637248	1.381228219	0
ROR2	6.886394344	0.455486661	0.066142983	1.371245286	0
SLC15A3	6.885096732	0.538210192	0.078170317	1.452169837	0
GPM6B	6.879306249	0.332042911	0.048266918	1.258794614	0
CRYAB	6.878741152	0.579755262	0.084282174	1.494595685	0
CALD1	6.877402245	0.499558077	0.072637612	1.41378043	0
IFI30	6.859204507	0.532189306	0.077587613	1.446122038	0
RBP1	6.855555922	0.702221315	0.102430981	1.627007967	0
MT1E	6.850988751	0.484037731	0.070652244	1.398652662	0
CYBB	6.845217852	0.522187165	0.076284959	1.436130812	0
AEBP1	6.840889746	0.732617037	0.107093823	1.66165058	0
RAC2	6.838987859	0.610857688	0.089319897	1.527166846	0
EPB41L2	6.838093246	0.340260596	0.049759572	1.26598525	0
MAF	6.831426318	0.387009555	0.056651355	1.307680009	0
CPVL	6.831097337	0.57399897	0.084027345	1.488644192	0
IL32	6.830844167	0.554133143	0.081122205	1.468286134	0
ARL4C	6.83012375	0.287674065	0.042118427	1.220670707	0
VIM	6.805398348	0.581702227	0.085476587	1.496614053	0
CLEC4A	6.794312525	0.551935252	0.081234893	1.46605096	0
GREM1	6.793089009	0.775662691	0.114184091	1.711976254	0
IL10RA	6.791416087	0.480324879	0.070725291	1.395057783	0
FMOD	6.790788322	0.497246092	0.073223618	1.411516596	0
FBN1	6.786984362	0.332707764	0.049021443	1.259374852	0
COL5A1	6.782222954	0.724569697	0.106833659	1.652407711	0
GLT8D2	6.781125143	0.569468995	0.083978541	1.48397727	0
C1QB	6.778856043	0.756708072	0.111627695	1.689630839	0
SH2B3	6.778484221	0.445985772	0.065794322	1.3622446	0
AK4	6.769571601	0.530297285	0.078335427	1.444226766	0
C3AR1	6.76535249	0.386161883	0.057079344	1.306911893	0
SERPING1	6.764802341	0.571615168	0.084498429	1.486186499	0
ECI2	6.763133616	0.444103522	0.065665348	1.360468471	0
IFI16	6.76045995	0.614687312	0.090923889	1.531226086	0
PLAUR	6.752530447	0.526257603	0.077934873	1.440188451	0
ACTN1	6.743331498	0.50747383	0.075255655	1.421558855	0
LILRB2	6.737832622	0.48702142	0.072281614	1.401548258	0
IFITM3	6.714809175	0.588839443	0.087692655	1.504036358	0
RGS1	6.698183763	0.714161751	0.106620209	1.640529744	0
GFPT2	6.675174651	0.580730283	0.086998515	1.495606123	0
VSNL1	6.674396461	0.615173508	0.092169159	1.531742203	0
OAT	6.672604428	0.606121047	0.090837252	1.52216109	0
EVI2A	6.657500529	0.445682547	0.066944425	1.361958314	0

SULF1	6.647659949	0.833412476	0.1253693	1.781895184	0
MN1	6.647289427	0.412138613	0.062001003	1.330656884	0
COL6A1	6.644039294	0.712231036	0.107198499	1.638335742	0
COL16A1	6.638969619	0.612763415	0.092297969	1.529185489	0
TNFSF4	6.638516529	0.393561399	0.05928454	1.313632203	0
LAPTM5	6.63687404	0.595013841	0.089652725	1.510487077	0
ITGB2	6.635315414	0.68667227	0.10348751	1.609566596	0
SLC2A3	6.627371709	0.634248962	0.095701432	1.552129533	0
MNDA	6.623325717	0.445717488	0.067295118	1.3619913	0
MT1F	6.617048562	0.591369646	0.089370607	1.506676456	0
BIN1	6.6150366	0.505088152	0.076354551	1.419210071	0
TYROBP	6.604121988	0.676582357	0.102448495	1.598348891	0
FGL2	6.566317703	0.600653629	0.091474957	1.516403435	0
IL15	6.543897135	0.30579845	0.046730327	1.23610256	0
NCKAP1L	6.536497175	0.339900907	0.052000467	1.265669657	0
EDNRA	6.532620869	0.434403405	0.066497569	1.351351895	0
NNMT	6.523940472	0.513697755	0.078740411	1.427704844	0
CLIP4	6.522061912	0.352814992	0.054095621	1.277049983	0
PRDM1	6.520608985	0.264543873	0.040570424	1.201256196	0
MAFB	6.51444871	0.504666304	0.077468766	1.41879515	0
CCL8	6.490303907	0.720243071	0.110972164	1.647459582	0
CLEC5A	6.485969	0.414071389	0.063841099	1.332440756	0
AIF1	6.480103245	0.529729733	0.081747113	1.443658722	0
PDLIM3	6.476674459	0.587270982	0.090674772	1.502402097	0
HIF1A	6.475191965	0.374653594	0.057859843	1.296528207	0
ADAM19	6.464774929	0.628882754	0.097278368	1.546366998	0
TRPS1	6.462096027	0.233041525	0.036062838	1.175310157	0
ZYX	6.456699771	0.348016546	0.053900066	1.272809533	0
CFI	6.447923006	0.47570821	0.073776968	1.390600689	0
VSIG4	6.446364685	0.418462709	0.064914526	1.33650266	0
ISG20	6.446029043	0.631004508	0.097890423	1.548642895	0
HPSE	6.441943516	0.363768836	0.056468802	1.286783049	0
EGR2	6.435249503	0.606762109	0.09428727	1.522837613	0
LGALS1	6.427468935	0.61313413	0.09539278	1.529578479	0
MS4A6A	6.402548905	0.589533244	0.092077898	1.504759833	0
GNLY	6.394334726	0.434913589	0.068015456	1.351829862	0
IL1RAP	6.38916049	0.363861883	0.056949874	1.286866043	0
COL1A2	6.385118702	0.651654428	0.102058311	1.57096869	0
POSTN	6.38290921	0.781942113	0.122505598	1.71944398	0
RUNX3	6.379422473	0.595264144	0.093310037	1.510749164	0
IGSF6	6.375140206	0.471022433	0.073884247	1.386091439	0
BNC2	6.372625608	0.321194786	0.050402268	1.249364798	0
SOCS1	6.366748642	0.326254266	0.051243466	1.253753967	0
PSMB9	6.364814297	0.489919526	0.076973106	1.404366537	0

HS3ST3A1	6.358237209	0.618286062	0.097241742	1.535050441	0
CYBRD1	6.355025542	0.528458057	0.08315593	1.442386758	0
HEG1	6.347990296	0.355950035	0.05607287	1.279828089	0
ACOT9	6.346337083	0.353614988	0.055719541	1.277758322	0
ATP8B2	6.342137711	0.219175114	0.034558555	1.16406782	0
NID2	6.334379204	0.405811333	0.064064894	1.324833752	0
GEM	6.333438748	0.373389298	0.058955224	1.295392501	0
CREB5	6.321641758	0.291022972	0.04603598	1.223507523	0
TGFB3	6.316944968	0.485787928	0.076902352	1.400350458	0
CXCL10	6.315344305	0.896457589	0.141949123	1.861489642	0
HBEGF	6.30866107	0.500052718	0.079264477	1.414265241	0
SRPX	6.307093311	0.744022976	0.117966064	1.674839648	0
CCR1	6.302563305	0.327193395	0.051914337	1.25457037	0
CYTIP	6.298433006	0.464148837	0.073692748	1.379503233	0

Basal Negative Genes					
Gene ID	Score(d)	Numerator(r)	Denominator(s+s0)	Fold Change	q-value(%)
PPARG	-10.75957355	-0.733344244	0.068157371	0.601507969	0
RAB15	-9.976460982	-0.734951766	0.073668585	0.600838112	0
UPK2	-9.849181854	-1.46595255	0.148840033	0.361996449	0
GAREM	-9.76118614	-0.603124784	0.061788063	0.658326517	0
TRAK1	-9.73091238	-0.698647946	0.071796756	0.616149374	0
SCNN1B	-9.658341003	-1.225200234	0.126854108	0.427738142	0
TOX3	-9.452715333	-0.935158696	0.098930166	0.522984939	0
GATA3	-9.362973895	-0.945231876	0.100954236	0.519346074	0
SEMA5A	-9.285555431	-0.835775833	0.090008168	0.560281657	0
RNF128	-9.07737643	-0.660699376	0.072785279	0.63257157	0
TMEM97	-9.028752745	-0.596672548	0.066085822	0.661277377	0
PLEKHG6	-8.89845132	-0.609174361	0.06845847	0.655571771	0
ADIRF	-8.866310326	-1.086359021	0.122526618	0.470948425	0
ERBB3	-8.865375224	-0.564801656	0.063708714	0.676048354	0
SLC27A2	-8.748486217	-0.662390966	0.075714924	0.631830302	0
SCNN1G	-8.743191203	-0.978495716	0.111915168	0.507508638	0
ACSF2	-8.640196152	-0.604759039	0.06999367	0.657581201	0
GPD1L	-8.630692503	-0.769615155	0.089171889	0.586573925	0
VGLL1	-8.601691889	-0.953125172	0.110806709	0.516512378	0
TBX2	-8.556340691	-0.931635592	0.10888248	0.524263644	0
TMPRSS2	-8.517023642	-1.0652207	0.125069595	0.477899545	0
ATP8B1	-8.475552629	-0.436118935	0.051456106	0.739120281	0
PLCE1	-8.420499756	-0.583123697	0.069250486	0.667516916	0
EPB41L1	-8.414430775	-0.404350681	0.048054431	0.755576281	0
BCAT2	-8.403524667	-0.65217732	0.077607593	0.636319254	0

HMGCS2	-8.386953719	-1.630730758	0.194436599	0.322924597	0
SPAG4	-8.266662973	-0.529263633	0.064023855	0.692908312	0
FBP1	-8.234664654	-0.902018259	0.109539161	0.535137576	0
SLC9A2	-8.230985661	-0.734715202	0.089262117	0.600936642	0
MYCL1	-8.217912843	-0.389496384	0.047396023	0.763396044	0
TBX3	-8.19303596	-0.534753564	0.065269281	0.690276577	0
CYP2J2	-8.189282792	-0.936792778	0.11439253	0.52239291	0
VIPR1	-8.154732885	-0.697041664	0.085476946	0.616835771	0
PPFIBP2	-8.111713227	-0.763680651	0.09414542	0.588991757	0
SLC29A3	-8.091059205	-0.517564078	0.063967407	0.698550307	0
GDPD3	-8.080070871	-0.985066898	0.121913151	0.505202296	0
EVPL	-8.033606451	-0.745168662	0.092756431	0.596598121	0
TM7SF2	-8.013257836	-0.763756093	0.095311558	0.588960958	0
DHRS11	-8.009321586	-0.504181726	0.062949367	0.705060163	0
PIK3C2B	-7.981083921	-0.628029345	0.078689731	0.647059666	0
ACSL5	-7.968369763	-0.493489043	0.061930992	0.710305203	0
TLE2	-7.963782495	-0.710853508	0.089260789	0.610958585	0
ACOX1	-7.903772604	-0.340154313	0.043036956	0.789956812	0
CAPN5	-7.903118094	-0.868919271	0.109946386	0.547556874	0
FAM174B	-7.892790823	-0.843110272	0.106820299	0.5574405	0
CNGA1	-7.849462647	-0.809189254	0.10308849	0.570702483	0
SNCG	-7.798782813	-0.871648288	0.111767222	0.546522089	0
NR2F6	-7.794981496	-0.409499184	0.052533695	0.752884684	0
CYP4B1	-7.738968705	-0.975208371	0.126012704	0.508666373	0
VSIG10	-7.724199791	-0.478375491	0.061932045	0.717785413	0
EPN3	-7.694187895	-0.530802959	0.068987522	0.692169387	0
UPK1A	-7.6939247	-1.616724074	0.210129958	0.326075042	0
BHMT	-7.668163406	-1.435365689	0.187185068	0.369753142	0
ZBTB7C	-7.62374209	-0.739898612	0.097051894	0.598781431	0
TJP3	-7.610538008	-0.700763043	0.09207799	0.615246716	0
MCCC1	-7.595942898	-0.514533939	0.067737995	0.700019037	0
ZSCAN16	-7.565939179	-0.522469422	0.069055462	0.696179182	0
ACOXL	-7.537052933	-0.424099608	0.056268625	0.745303733	0
CYP4F12	-7.533350243	-0.921921295	0.122378658	0.527805652	0
SPINK1	-7.516496798	-1.57476353	0.209507643	0.335698145	0
FOXA1	-7.43010809	-0.878719605	0.118264714	0.543849885	0
SORL1	-7.423715891	-0.632929312	0.08525775	0.644865721	0
SYT17	-7.405480536	-0.501440314	0.067712056	0.706401194	0
THEM6	-7.38226674	-0.607432846	0.082282701	0.656363607	0
CAB39L	-7.355409754	-0.434009856	0.059005531	0.740201592	0
PSCA	-7.282159271	-1.196509329	0.164306943	0.436329726	0
SLC37A1	-7.265024308	-0.402581051	0.055413586	0.75650365	0
SSH3	-7.262167503	-0.369368641	0.050862038	0.774121197	0
CBLC	-7.256799624	-0.489333041	0.06743097	0.712354343	0

TMEM51	-7.247598873	-0.352282985	0.048606855	0.783343519	0
SUOX	-7.236466271	-0.386064716	0.053349895	0.765214058	0
ZNF443	-7.200435147	-0.352922545	0.049014058	0.782996332	0
TNFRSF21	-7.194314734	-0.535381992	0.074417372	0.689975963	0
PPP1R3C	-7.174910314	-0.855181705	0.119190578	0.552795699	0
P4HTM	-7.165441185	-0.332288374	0.046373749	0.79427562	0
DAPK1	-7.162264597	-0.580660846	0.081072242	0.66865742	0
BCAS1	-7.161920832	-0.823297441	0.114954837	0.565148755	0
SCAP	-7.131970146	-0.423627723	0.059398415	0.745547552	0
PHLPP1	-7.126921075	-0.295808471	0.041505787	0.814615703	0
ACP6	-7.090515885	-0.489862218	0.069086964	0.712093102	0
SLC14A1	-7.086720827	-0.739270471	0.104317708	0.599042194	0
TNNC1	-7.068356216	-0.700600141	0.099117832	0.615316191	0
CEBPA	-7.066234762	-0.635695576	0.089962419	0.643630422	0
DHRS2	-7.059443936	-1.253399285	0.177549294	0.419458712	0
CYB5A	-7.059232545	-0.680106303	0.09634281	0.624119285	0
POF1B	-7.052682768	-0.351950479	0.049903064	0.783524081	0
PPP1R9A	-7.048767493	-0.42581305	0.060409575	0.744419086	0
ALAS1	-7.019858833	-0.31445589	0.044795187	0.804154219	0
C1orf116	-7.017868851	-0.532496973	0.075877305	0.691357118	0
TMC7	-7.009622961	-0.349697211	0.049888163	0.784748781	0
CXXC1	-6.986821804	-0.378081915	0.054113576	0.76945992	0
PTPRR	-6.986666989	-0.585720803	0.083834081	0.666316349	0
ELOVL6	-6.971123549	-0.496094224	0.07116417	0.709023709	0
ELF3	-6.91620046	-0.78581815	0.113619921	0.580022934	0
ST3GAL4	-6.915410053	-0.564844479	0.081679101	0.676028288	0
MSX2	-6.900658921	-0.401470974	0.058178643	0.757085964	0
TOP2B	-6.895840172	-0.460201568	0.066736113	0.726884694	0
SGPL1	-6.888918533	-0.348512301	0.050590278	0.785393574	0
FUT9	-6.877738454	-0.37064412	0.053890406	0.773437103	0
KRT20	-6.852201889	-0.909072581	0.13266868	0.53252731	0
MAOA	-6.845253761	-0.709939162	0.103712614	0.611345919	0
ASCC2	-6.831163574	-0.45420018	0.066489431	0.729914719	0
SCIN	-6.830891174	-0.667784584	0.097759512	0.629472569	0
ELF5	-6.8307154	-0.546528078	0.080010372	0.684665832	0
RALBP1	-6.827806692	-0.514798666	0.07539737	0.699890598	0
ZNF165	-6.788504358	-0.485614284	0.071534797	0.714192909	0
PEBP1	-6.776913567	-0.407026199	0.060060704	0.754176343	0
CLN3	-6.769557995	-0.356777924	0.052703282	0.780906689	0
HNF1B	-6.703070268	-0.352279485	0.052554944	0.783345419	0
PLA2G2F	-6.686518926	-0.514000681	0.076871192	0.70027783	0
ICA1	-6.684578145	-0.304738301	0.045588262	0.809589059	0
CXADR	-6.67611992	-0.623155378	0.09334095	0.649249374	0
PLEKHA6	-6.655833435	-0.518003928	0.077827057	0.698337365	0

NEDD4L	-6.648338963	-0.463495765	0.069716025	0.725226845	0
ZNF552	-6.645459457	-0.265451273	0.039944758	0.831938465	0
LIMCH1	-6.633406132	-0.641617597	0.096725209	0.640993843	0
POGK	-6.622452882	-0.394384433	0.059552622	0.760813931	0
SH2D4A	-6.579979538	-0.43822196	0.066599289	0.738043646	0
ABCD3	-6.575222337	-0.378935815	0.057630875	0.769004628	0
BDH1	-6.550597484	-0.353627252	0.053983969	0.782613959	0
UPK3B	-6.534894344	-0.602444145	0.092188812	0.658637178	0
HSDL2	-6.528940782	-0.400220366	0.061299433	0.757742532	0
FA2H	-6.528283239	-0.285302687	0.04370256	0.82056943	0
GNL3	-6.526433583	-0.341203442	0.052280229	0.789382564	0
SLC25A10	-6.526255698	-0.419491424	0.064277504	0.747688151	0
MYH14	-6.484650617	-0.332686332	0.051303663	0.794056554	0
RAB11A	-6.478309407	-0.346053071	0.053417188	0.786733504	0
PTPRF	-6.473853033	-0.39096177	0.060390894	0.762621035	0
RAP1GAP	-6.469735707	-0.632298888	0.097731796	0.645147574	0
TRAPPC6A	-6.451776338	-0.35391786	0.054855879	0.78245633	0
KIAA0195	-6.445452325	-0.334362316	0.051875695	0.793134632	0
SH3GLB2	-6.424387725	-0.41870326	0.065174033	0.748096735	0
ACSM3	-6.380394666	-0.473869015	0.074269546	0.720031029	0
SLC44A4	-6.353960301	-0.524101227	0.082484183	0.695392192	0
STAP2	-6.343499378	-0.35497957	0.055959581	0.781880715	0
ACOX3	-6.341811347	-0.389232454	0.061375596	0.763535714	0
FASN	-6.331387673	-0.460590125	0.072747105	0.72668895	0
TOB1	-6.331206285	-0.446286296	0.070489931	0.733929659	0
ACAA1	-6.328993969	-0.344438748	0.054422354	0.787614323	0
MYLIP	-6.305394337	-0.486757523	0.077196999	0.713627184	0
GDF15	-6.302973885	-0.764454777	0.121284776	0.588675799	0
MYO6	-6.29850259	-0.345773649	0.054897755	0.786885894	0
KRT7	-6.287893923	-0.765080126	0.121675101	0.588420687	0
C4orf19	-6.275089196	-0.525701188	0.083775891	0.694621424	0
TMEM159	-6.272204513	-0.27902434	0.044485848	0.824148181	0
DNAL4	-6.257691195	-0.275353699	0.044002443	0.826247726	0
RAB25	-6.246523297	-0.603897542	0.096677386	0.657973989	0
PLS1	-6.227716213	-0.499070831	0.080137054	0.707562341	0
ZNF91	-6.21870126	-0.54364391	0.087420811	0.686035953	0
INA	-6.215070501	-0.891134822	0.143382898	0.539189825	0
BAMBI	-6.213452778	-0.803597425	0.129331863	0.572918796	0
FDFT1	-6.203325146	-0.376310948	0.06066278	0.770405043	0
TSPAN6	-6.182836008	-0.468308189	0.075743265	0.722811724	0
ENGASE	-6.181792096	-0.340758468	0.055122926	0.789626073	0
BRAF	-6.178702736	-0.306342766	0.049580434	0.808689191	0
CACNB3	-6.17869709	-0.340371944	0.05508798	0.789837656	0
GATA2	-6.176769455	-0.478581215	0.077480828	0.717683066	0

FAM13A	-6.172160223	-0.347213604	0.056254794	0.786100894	0
TESC	-6.171772218	-0.747830578	0.121169504	0.595498353	0
ZFHX3	-6.167792835	-0.427148057	0.069254605	0.743730552	0
ERBB2	-6.143531677	-0.578176396	0.094111405	0.669809899	0
DCXR	-6.136445721	-0.299483911	0.048804133	0.812543012	0
CXorf57	-6.119417976	-0.670656066	0.109594747	0.628220939	0
PRR15L	-6.11280337	-0.506743023	0.08289863	0.703809541	0
TUFT1	-6.075428661	-0.414911234	0.068293327	0.750065643	0
CYP4F8	-6.074009573	-0.753096911	0.123986784	0.593328542	0
B4GALT4	-6.068974875	-0.341033444	0.056192924	0.789475585	0
ECHS1	-6.065641468	-0.34814424	0.057396113	0.785593969	0
ALDH5A1	-6.063906132	-0.437390231	0.072130113	0.738469258	0
SPINT1	-6.039719436	-0.359350927	0.05949795	0.779515207	0
EEF1A2	-6.038107931	-0.921455713	0.152606698	0.527976011	0
BMP3	-6.030739069	-0.306885179	0.050886828	0.808385204	0
GFOD2	-6.030217285	-0.269425512	0.044679238	0.82964985	0
HADH	-6.022730362	-0.355890743	0.059091263	0.781387053	0
UPK3A	-6.018615844	-0.599461927	0.099601294	0.660000066	0
RBM47	-6.011350976	-0.353610803	0.058823849	0.782622882	0
TMEM135	-5.993299395	-0.294935563	0.049210884	0.815108738	0
HMGCL	-5.982513651	-0.263941493	0.044118828	0.832809544	0
FZD3	-5.981897908	-0.314009075	0.052493219	0.804403311	0
FNBP1L	-5.98059607	-0.292016341	0.048827297	0.81675974	0
SCUBE2	-5.979557938	-0.763991157	0.127767164	0.588865004	0
UPK1B	-5.967959891	-0.749776771	0.125633681	0.594695568	0
ZNF446	-5.959458387	-0.240048067	0.040280182	0.846717101	0
NPAS2	-5.934201474	-0.391916267	0.06604364	0.762116647	0
MFSD5	-5.933047814	-0.265510376	0.044751093	0.831904383	0
TGFBR3	-5.922158853	-0.586094768	0.098966404	0.666143653	0
ADCK2	-5.920425679	-0.298865559	0.050480417	0.81289135	0
ZNF254	-5.909354082	-0.29735326	0.05031908	0.813743906	0
PAFAH1B3	-5.908722786	-0.363866259	0.061581203	0.777079303	0
GORASP1	-5.906464448	-0.254628809	0.043110191	0.838202768	0
NT5C2	-5.906305192	-0.332953659	0.05637258	0.793909432	0
OCEL1	-5.905846775	-0.30035801	0.050857738	0.812050858	0
FAM63A	-5.902127185	-0.336339249	0.056986107	0.792048539	0
TJP2	-5.900111024	-0.282469702	0.04787532	0.822182345	0
S100P	-5.889071598	-0.824563533	0.140015878	0.564653005	0
C9orf116	-5.875697628	-0.346441129	0.058961701	0.786521916	0
IKZF2	-5.857718058	-0.161682824	0.027601674	0.893981682	0
DTX4	-5.852553503	-0.195555096	0.033413637	0.873236842	0
DNM2	-5.840135799	-0.285314222	0.048854039	0.820562869	0
MFSD9	-5.837944567	-0.198531136	0.034007027	0.871437356	0
SLC37A4	-5.835071222	-0.296635781	0.050836703	0.814148697	0

TABLE 3

List of Significant Genes				
Offset Quantile	50		**Offset Value**	0.283592351
	both		**RNG Seed**	420473
		Class	1	2
Prior Distribution (Sample Prior)				
		Prob.	0.541984733	0.458015267
	id	**name**	**1 score**	**2 score**
	UPK2	UPK2	-0.1166	0.138
	SCNN1B	SCNN1B	-0.0955	0.113
	PPARG	PPARG	-0.0815	0.0965
	TOX3	TOX3	-0.0652	0.0771
	GATA3	GATA3	-0.0629	0.0745
	HMGCS2	HMGCS2	-0.0611	0.0723
	RAB15	RAB15	-0.0583	0.069
	AHNAK2	AHNAK2	0.0569	-0.0674
	ADIRF	ADIRF	-0.0558	0.066
	SEMA5A	SEMA5A	-0.0491	0.0581
	CHST15	CHST15	0.0476	-0.0563
	TRAK1	TRAK1	-0.0453	0.0536
	SCNN1G	SCNN1G	-0.0433	0.0512
	MT1X	MT1X	0.0411	-0.0486
	TMPRSS2	TMPRSS2	-0.041	0.0485
	VGLL1	VGLL1	-0.036	0.0426
	TBX2	TBX2	-0.0326	0.0386
	UPK1A	UPK1A	-0.03	0.0355
	GAREM	GAREM	-0.0296	0.035
	BHMT	BHMT	-0.0234	0.0277
	SPINK1	SPINK1	-0.0209	0.0248
	GPD1L	GPD1L	-0.0196	0.0232
	RNF128	RNF128	-0.0196	0.0232
	CYP2J2	CYP2J2	-0.0194	0.023
	EMP3	EMP3	0.0194	-0.0229
	GDPD3	GDPD3	-0.0188	0.0222
	FBP1	FBP1	-0.0184	0.0218
	MSN	MSN	0.0174	-0.0206
	MT2A	MT2A	0.0153	-0.0181
	CDK6	CDK6	0.0149	-0.0176
	ALOX5AP	ALOX5AP	0.0125	-0.0148
	PRRX1	PRRX1	0.0107	-0.0127
	SLC27A2	SLC27A2	-0.0097	0.0115
	TMEM97	TMEM97	-0.0077	0.0091

	CD14	CD14	0.007	-0.0082
	PLEKHG6	PLEKHG6	-0.006	0.0071
	CYP4B1	CYP4B1	-0.005	0.0059
	GLIPR1	GLIPR1	0.0047	-0.0055
	PDGFC	PDGFC	0.0046	-0.0055
	PRKCDBP	PRKCDBP	0.0045	-0.0053
	FAP	FAP	0.0035	-0.0042
	CAPN5	CAPN5	-0.0035	0.0041
	PALLD	PALLD	0.0025	-0.003
	TUBB6	TUBB6	0.0024	-0.0028
	SLC9A2	SLC9A2	-0.0022	0.0026
	PPFIBP2	PPFIBP2	-0.0013	0.0015
	FAM174B	FAM174B	-0.001	0.0012

[00130] It is to be understood that, while the invention has been described in conjunction with the detailed description, thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications of the invention are within the scope of the claims set forth below. All publications, patents, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

CLAIMS

What is claimed is:

1. A method to identify intrinsic high-grade bladder cancer subtypes in a patient sample which comprises:
   (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.
2. The method of claim 1, wherein the intrinsic high-grade bladder cancer subtypes are a basal subtype and a luminal subtype.
3. The method of claim 1, wherein the classifier biomarkers are selected from the biomarkers in Table 3.
4. The method of claim 1, wherein the classifier biomarkers are nucleic acids.
5. The method of claim 4, wherein the nucleic acids are detected by microarray.
6. The method of claim 4, wherein the nucleic acids are detected by next generation sequencing.
7. The method of claim 4, wherein the nucleic acids are detected by polymerase chain reaction (PCR).
8. The method of claim 4, where the nucleic acids are detected by a direct transcript counting method.
9. The method of claim 1, wherein the classifier biomarkers are proteins.
10. The method of claim 9, wherein the proteins are detected by an antibody assay.
11. The method of claim 1, wherein the patient sample is a bladder tissue sample.
12. The method of claim 11, wherein the patient sample is a formalin-fixed paraffin embodded (FFPE) bladder tissue sample.
13. The method of claim 1, wherein the patient sample is a sample containing circulating tumor cells.
14. The method of claim 1, wherein the patient sample is a sample of cell-free nucleic acids from blood.
15. The method of claim 1, wherein the patient sample is a voided urine cell sample.
16. The method of claim 1, wherein the patient sample is a sample of cell-free nucleic acids from urine.

17. A method to select patients with a poor prognosis using intrinsic high-grade bladder cancer subtypes in a sample which comprises:
    (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) selecting patients with the poor prognosis using the intrinsic high-grade bladder cancer subtype based on the statistical analysis of the detected expression of the classifier biomarkers.
18. The method of claim 17, wherein the intrinsic high-grade bladder cancer subtypes are a basal subtype and a luminal subtype.
19. The method of claim 17, wherein the classifier biomarkers are selected from the biomarkers in Table 3.
20. A kit for identification of high-grade bladder cancer subtypes in a patient sample which comprises:
    (a) a means for measuring levels of a plurality of high-grade bladder cancer classifier biomarkers; and
    (b) instructions for comparing the levels of a plurality of high-grade bladder cancer classifier biomarkers from the patient sample with levels of a plurality of high-grade bladder cancer classifier biomarkers for a control patient, wherein the levels of a plurality of high-grade bladder cancer classifier biomarkers are able to identify high-grade bladder cancer subtypes in the patient.
21. The kit of claim 20, wherein the intrinsic high-grade bladder cancer subtypes are a basal subtype and a luminal subtype.
22. The kit of claim 20, wherein the classifier biomarkers are selected from the biomarkers in Table 3.
23. The kit of claim 20, wherein the classifier biomarkers are nucleic acids.
24. The kit of claim 20, wherein the classifier biomarkers are proteins.
25. A method of identifying a compound that slows the progression or treats high-grade bladder cancer, the method comprising the steps of:
    (a) contacting a tissue or an animal model with a compound;
    (b) 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; and determining a functional effect of the compound on high-grade bladder cancer, thereby identifying a compound that slows the progression or treats high-grade bladder cancer.

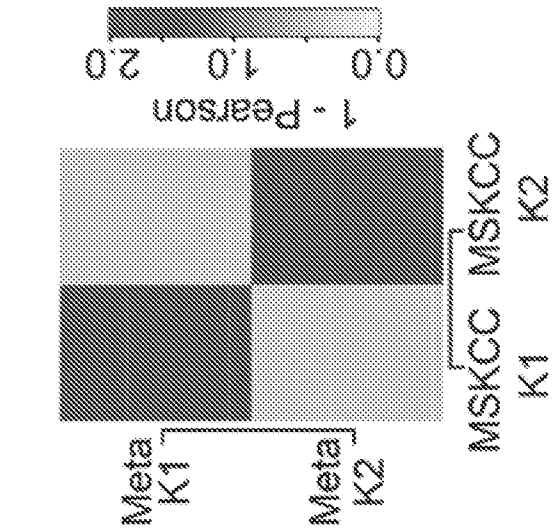


FIG. 1C

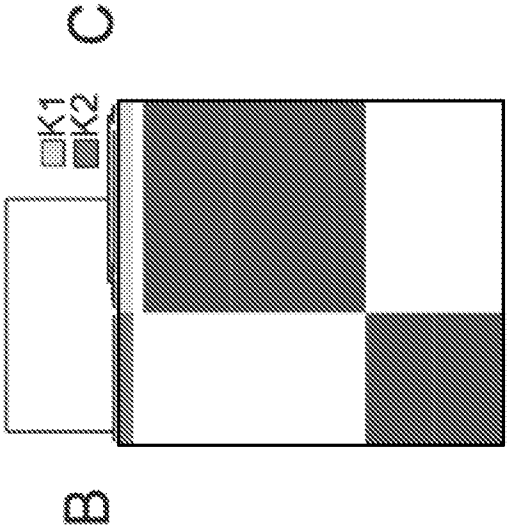


FIG. 1B

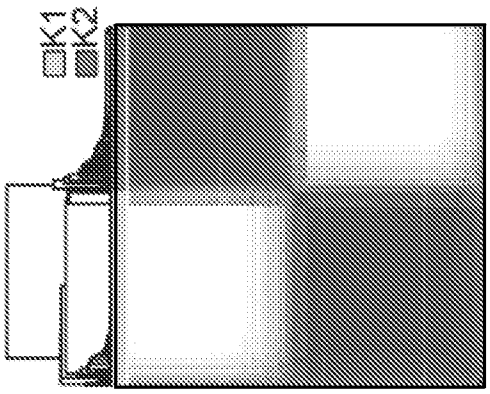


FIG. 1A

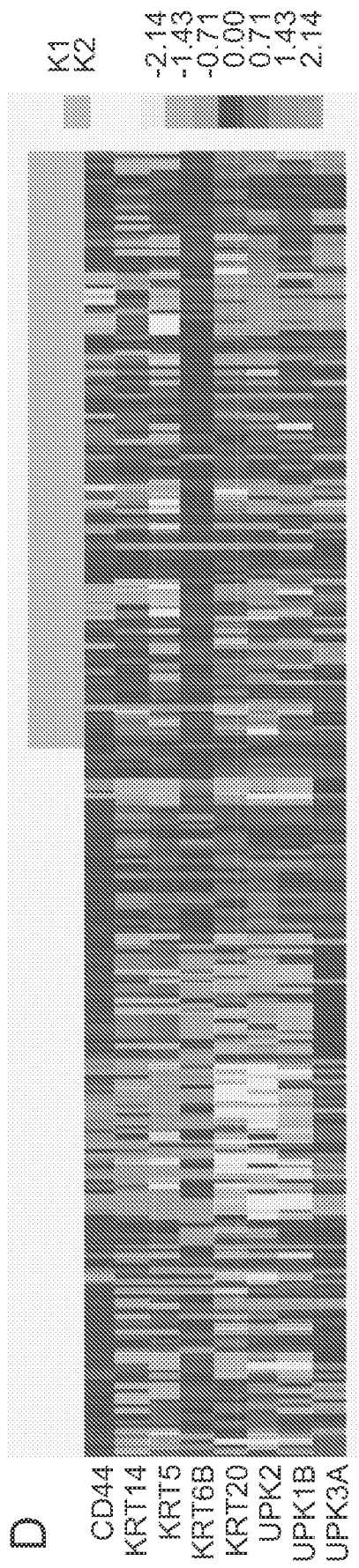


FIG. 1D

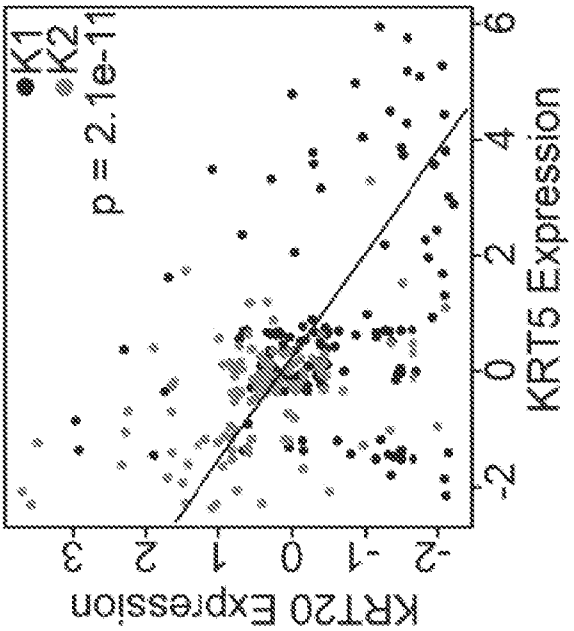


FIG. 1F

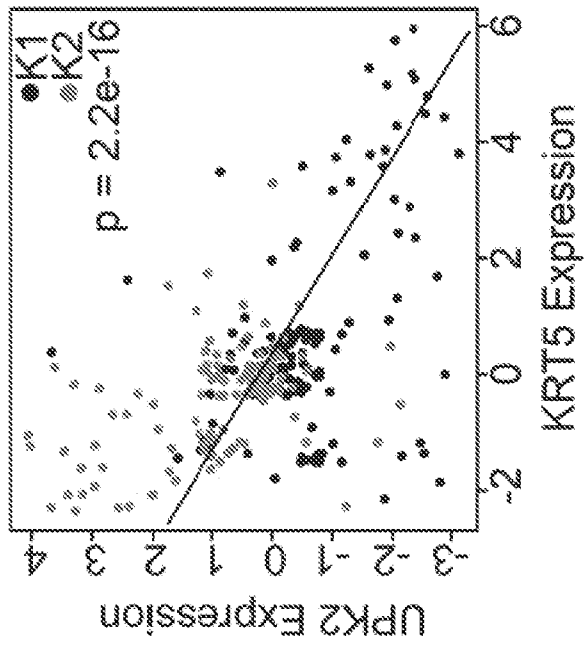


FIG. 1E

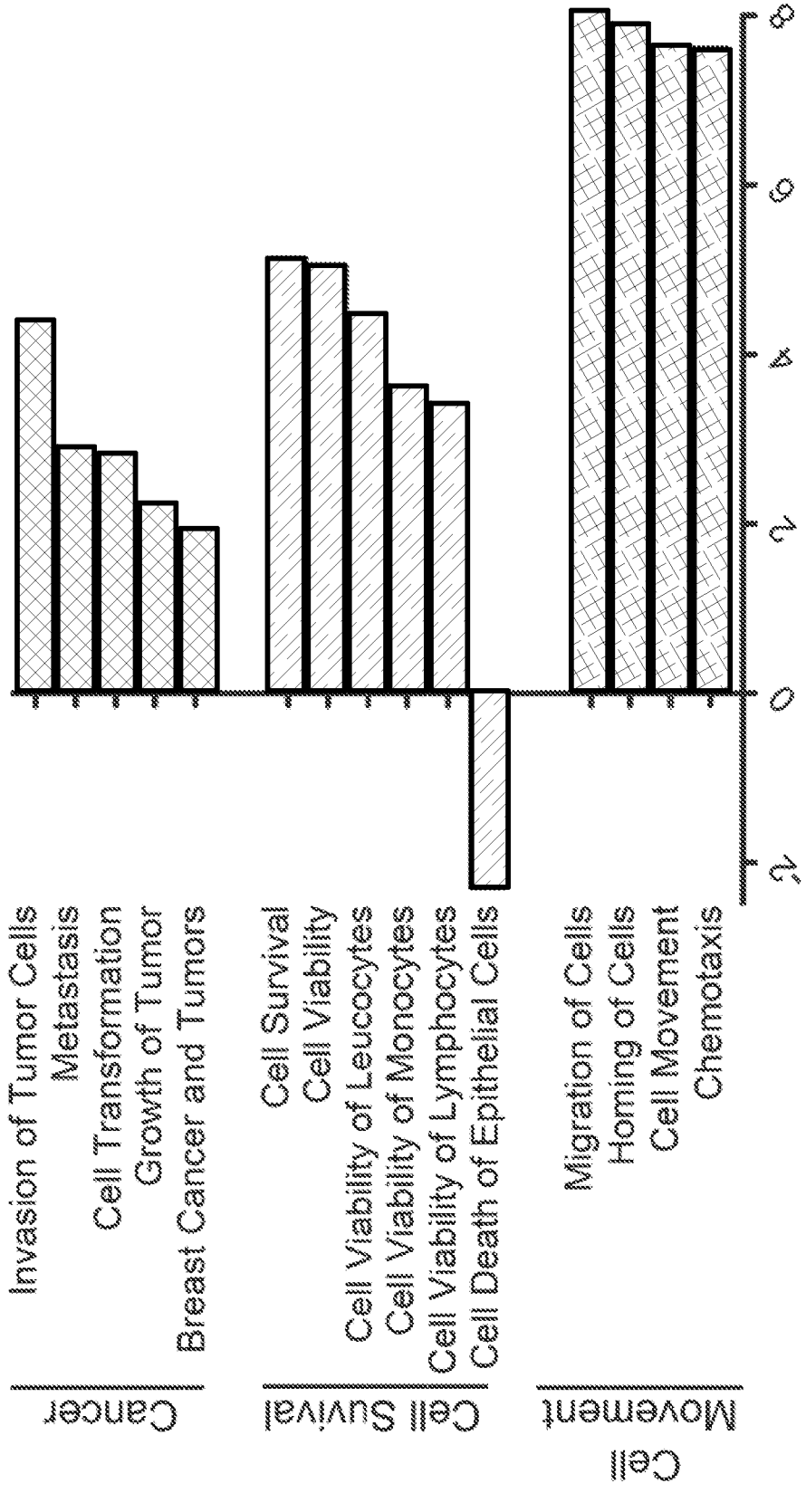

Cancer
Invasion of Tumor Cells
Metastasis
Cell Transformation
Growth of Tumor
Breast Cancer and Tumors
Cell Survival
Cell Viability
Cell Viability of Leucocytes
Cell Viability of Monocytes
Cell Viability of Lymphocytes
Cell Death of Epithelial Cells
Cell Movement
Migration of Cells
Homing of Cells
Cell Movement
Chemotaxis
-2
0
2
4
6
8
Activation z-score

FIG. 1G

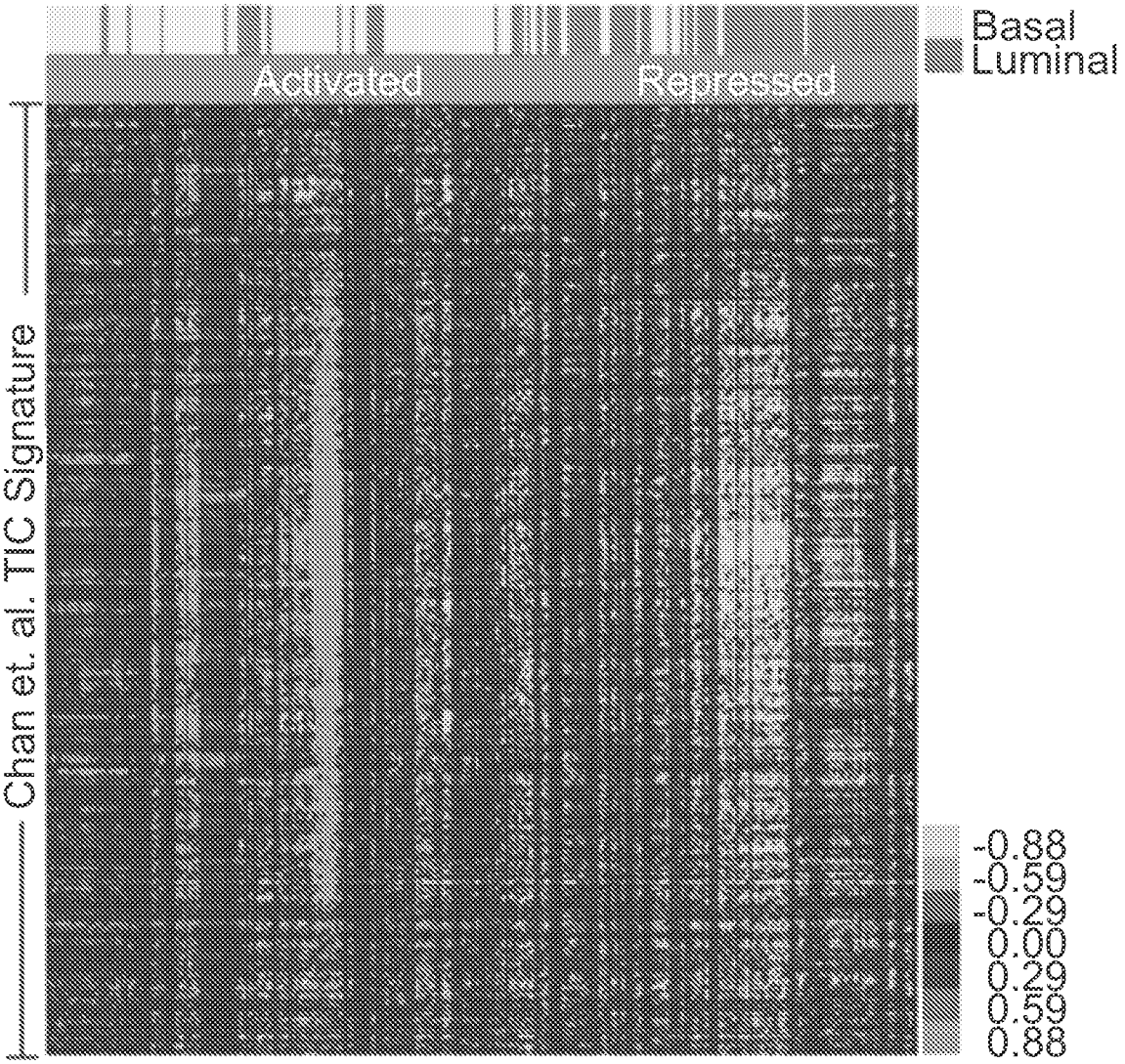

Basal
Luminal
Activated
Repressed
Chan et. al. TIC Signature
-0.88
-0.59
-0.29
0.00
0.29
0.59
0.88

FIG. 2A

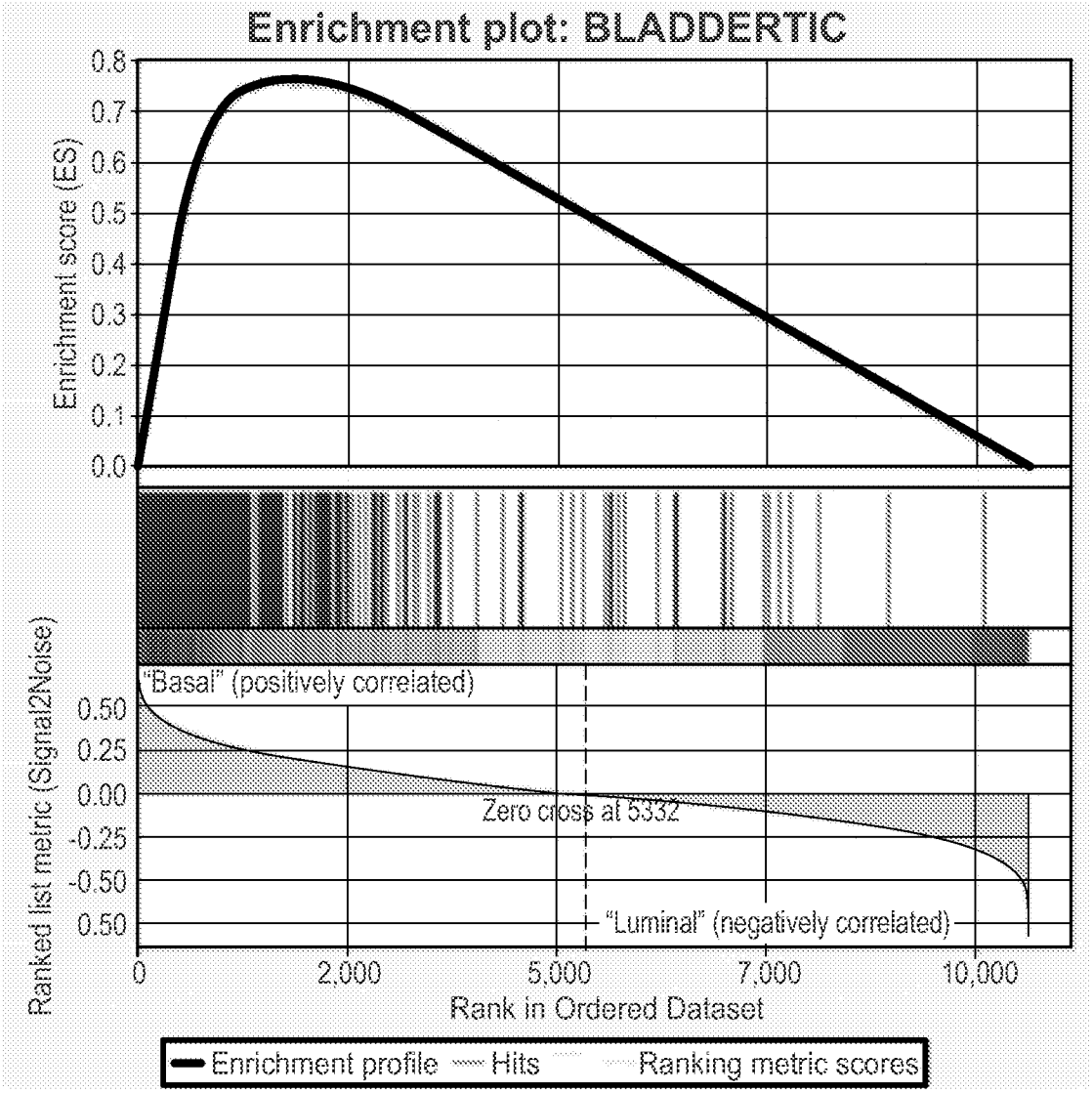
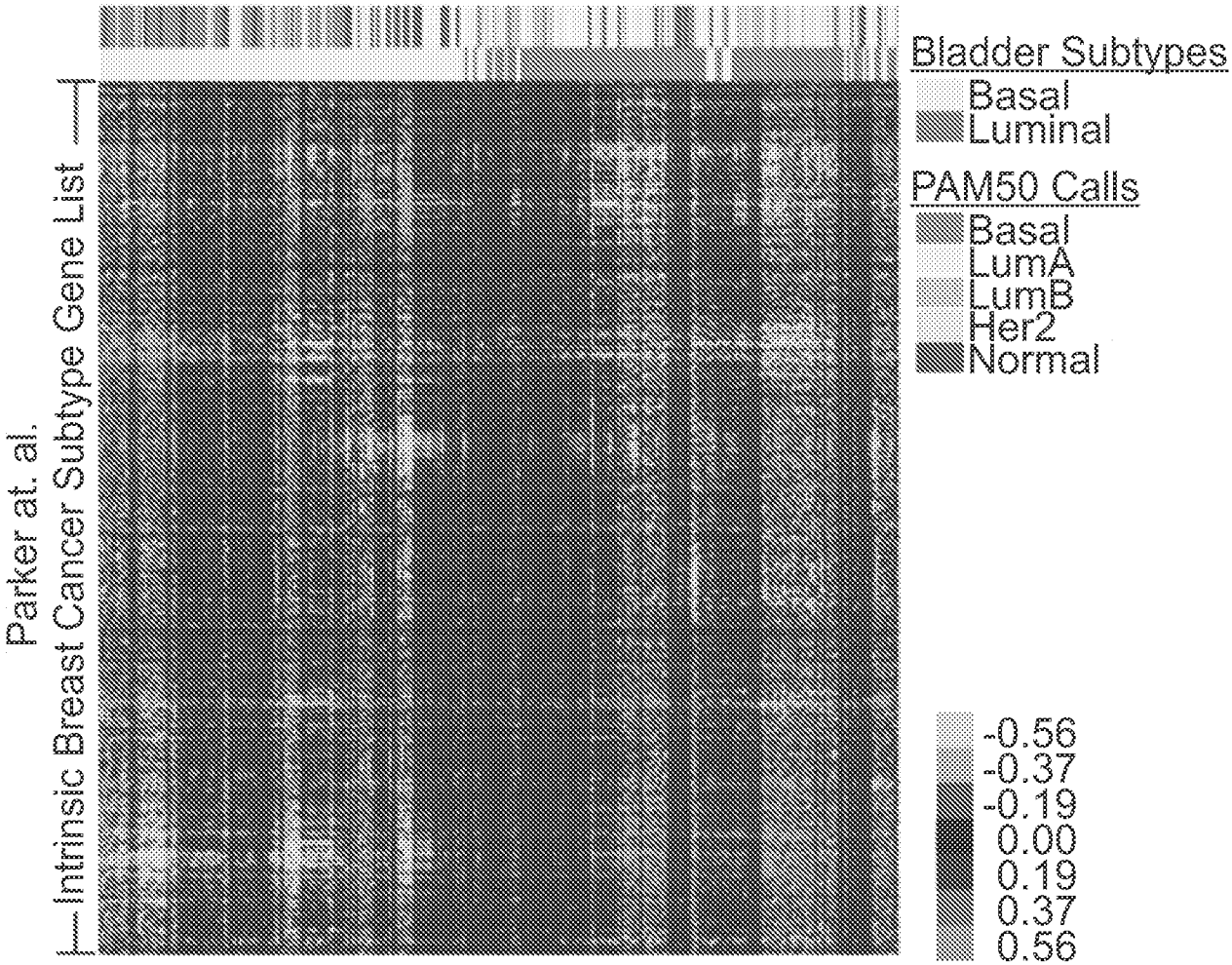

Enrichment plot: BLADDERTIC
Enrichment score (ES)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
"Basal" (positively correlated)
Ranked list metric (Signal2Noise)
0.50
0.25
0.00
-0.25
-0.50
0.50
Zero cross at 5332
"Luminal" (negatively correlated)
0
2,000
5,000
7,000
10,000
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

FIG. 2B



	Cluster1	Cluster2
Basal	125	17
Luminal	2	118

FIG. 2C

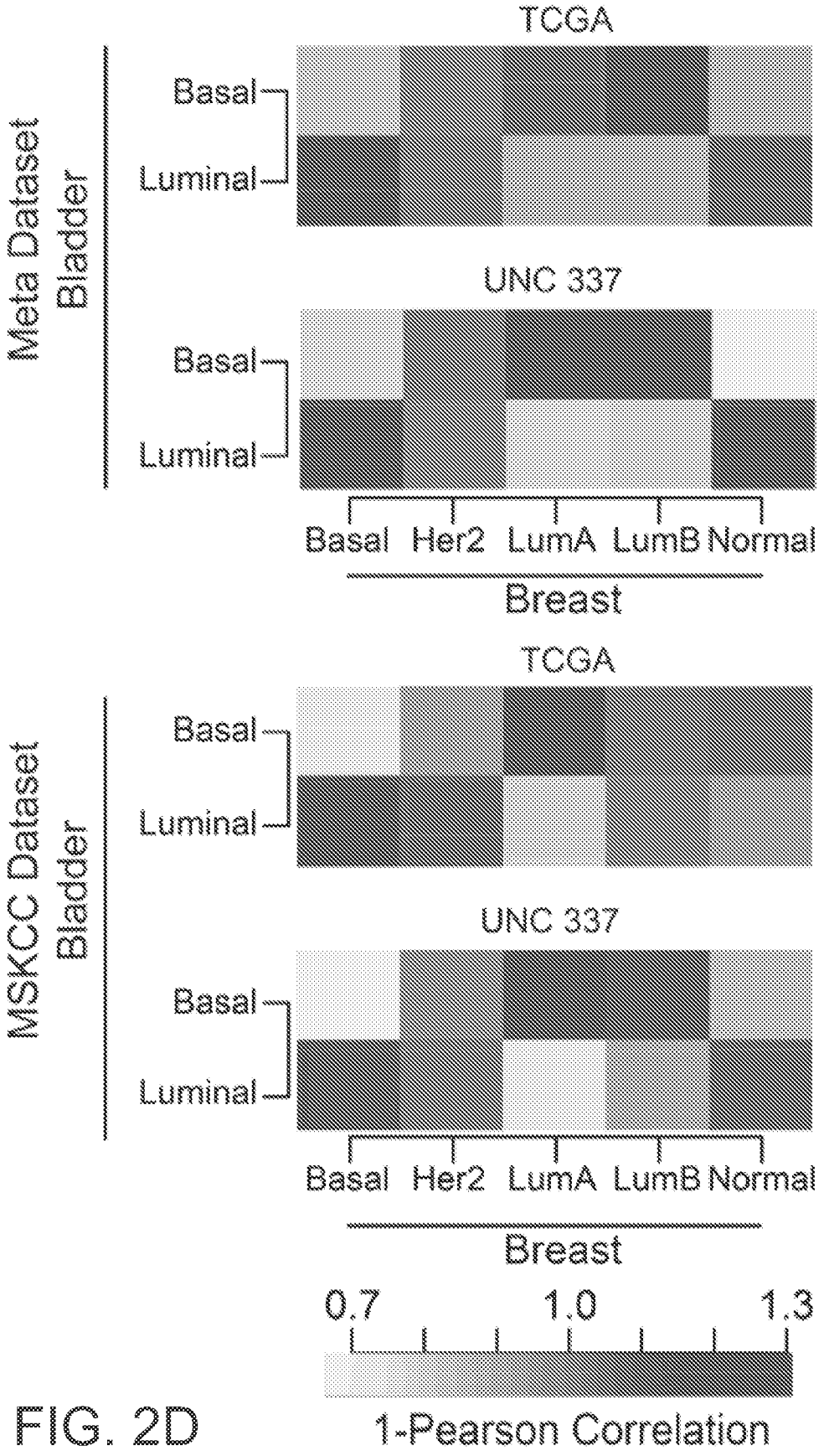

TCGA
Basal
Luminal
Meta Dataset
Bladder
UNC 337
Basal
Luminal
Basal Her2 LumA LumB Normal
Breast
TCGA
Basal
Luminal
MSKCC Dataset
Bladder
UNC 337
Basal
Luminal
Basal Her2 LumA LumB Normal
Breast
0.7 1.0 1.3
1-Pearson Correlation

FIG. 2D

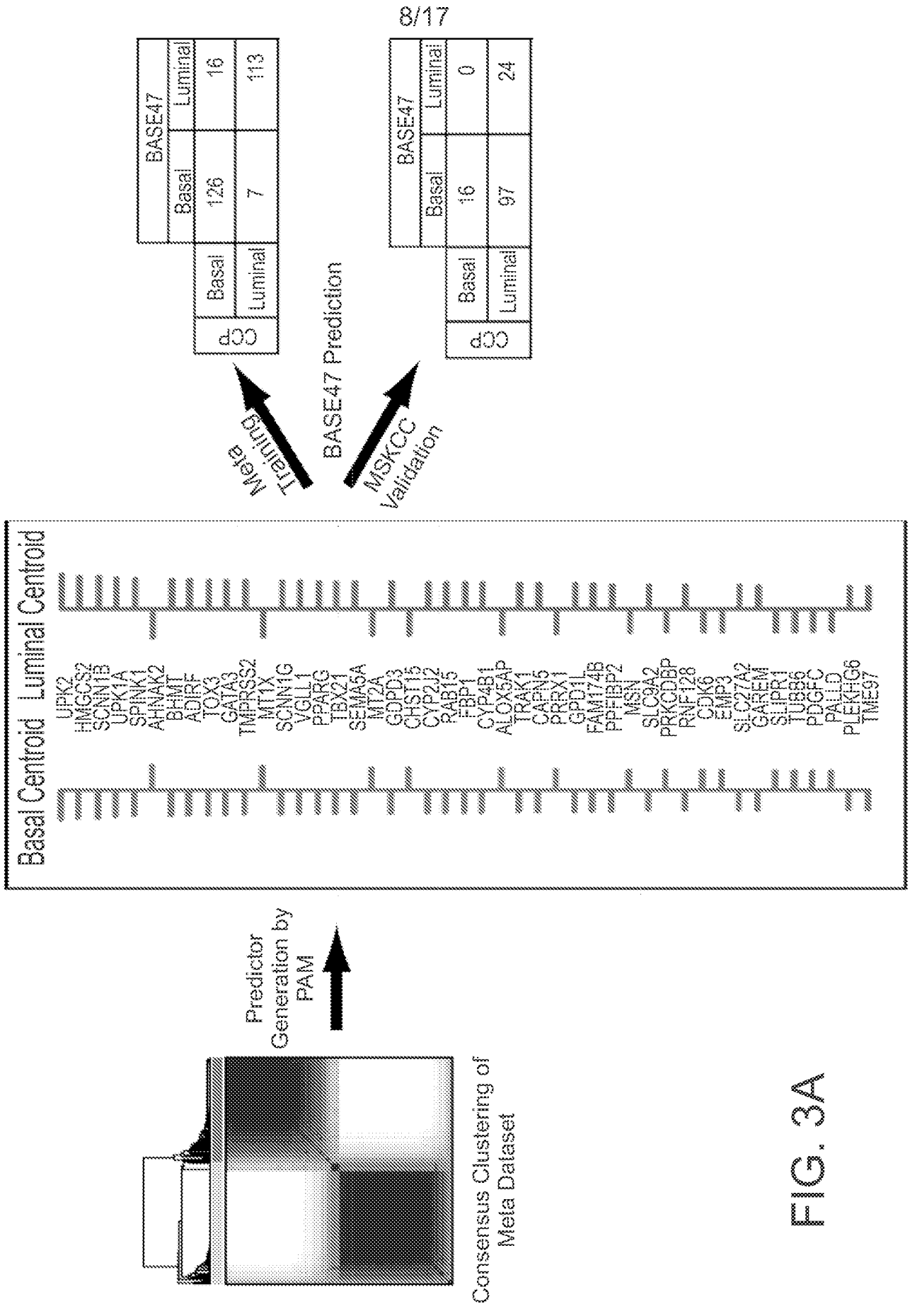

8/17
Consensus Clustering of Meta Dataset
Predictor Generation by PAM
Basal Centroid Luminal Centroid
UPK2
HMGCS2
SCNN1B
UPK1A
SPINK1
AHNAK2
BHMT
ADIRF
TOX3
GATA3
TMPRSS2
MT1X
SCNN1G
VGLL1
PPARG
TBX21
SEMA5A
MT2A
GDPD3
CHST15
CYP2J2
RAB15
FBP1
CYP4B1
ALOX5AP
TRAK1
CAPN5
PRRX1
GPD1L
FAM174B
PPFIBP2
MSN
SLC9A2
PRKCDBP
RNF128
CDK6
EMP3
SLC27A2
GAREM
SLIPR1
TUBB6
PDGFC
PALLD
PLEKHG6
TME97
Meta Training
BASE47 Prediction
MSKCC Validation
BASE47
Basal
Luminal
Basal
126
7
16
113
BASE47
Basal
Luminal
Basal
16
97
Luminal
0
24

FIG. 3A

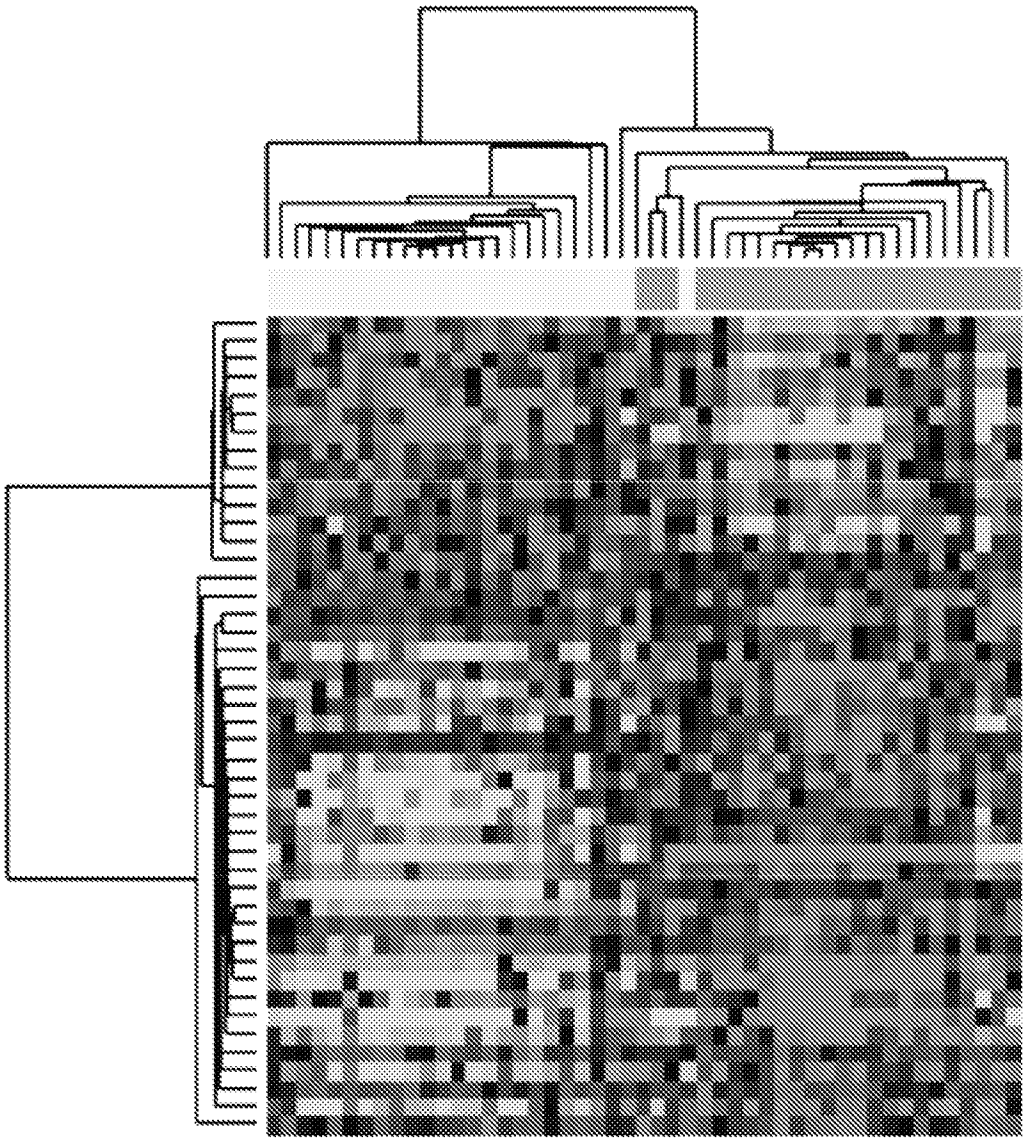


FIG. 3B

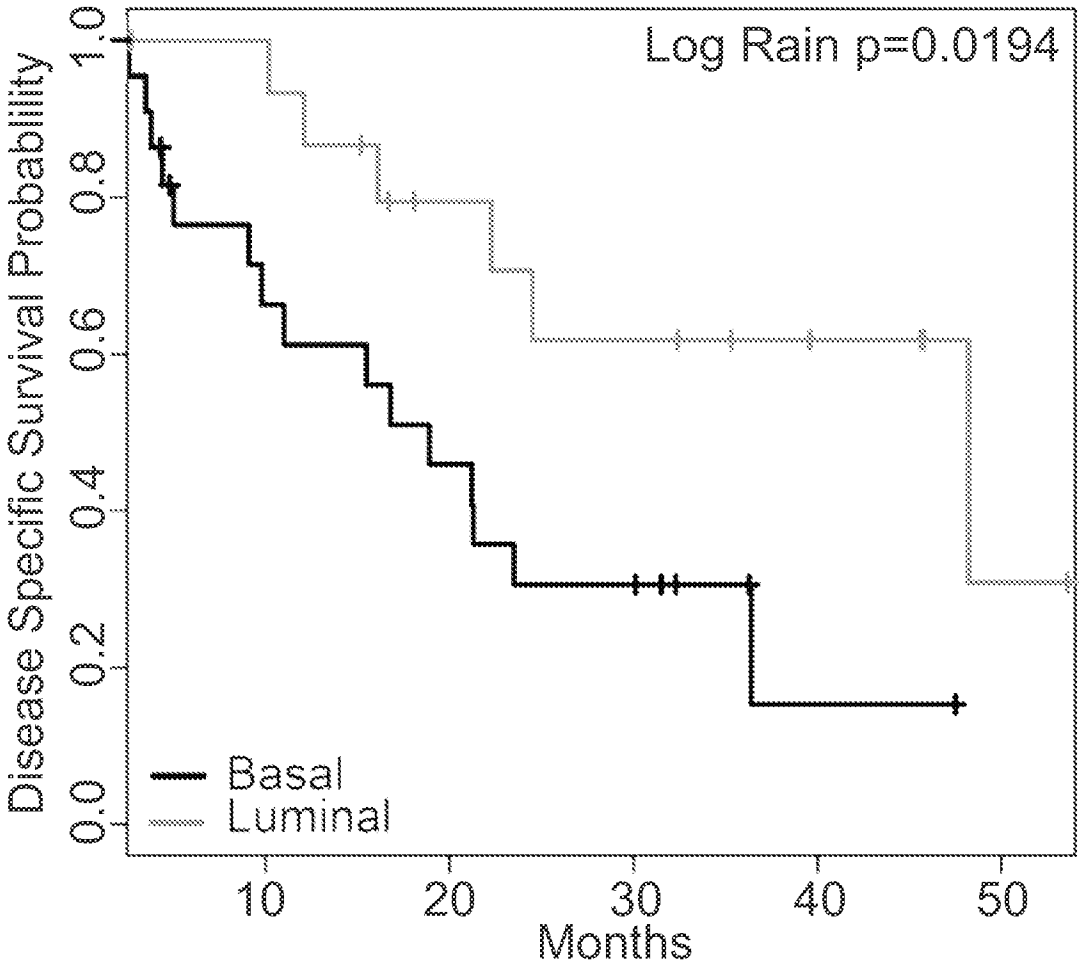


FIG. 4A

Superficial Tumors

	Basal-Like	Luminal-Like
GSE31684	8	7
GSE13507	71	32
GSE32894	116	97

FIG. 4B

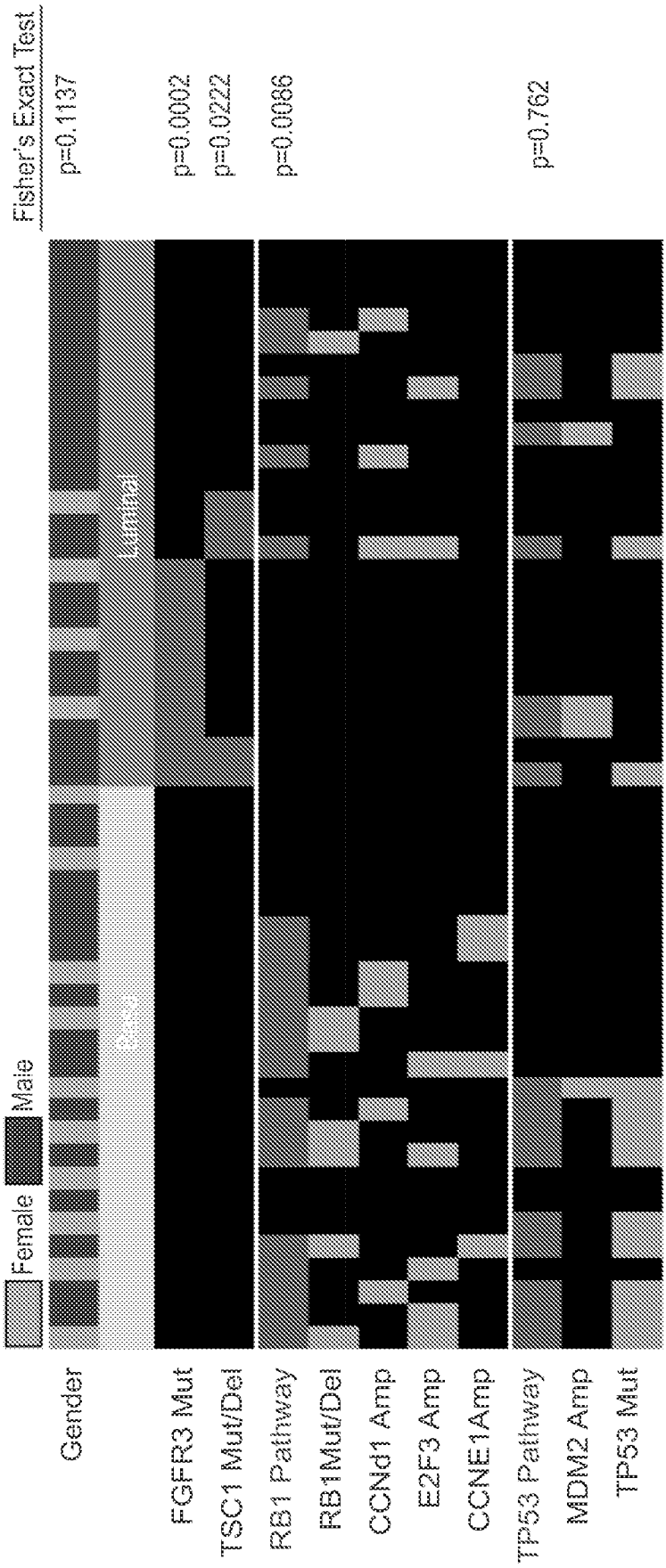

Fisher's Exact Test
p=0.1137
p=0.0002
p=0.0222
p=0.0086
p=0.762
Female
Male
Basal
Luminal
Gender
FGFR3 Mut
TSC1 Mut/Del
RB1 Pathway
RB1Mut/Del
CCNd1 Amp
E2F3 Amp
CCNE1Amp
TP53 Pathway
MDM2 Amp
TP53 Mut

FIG. 4C

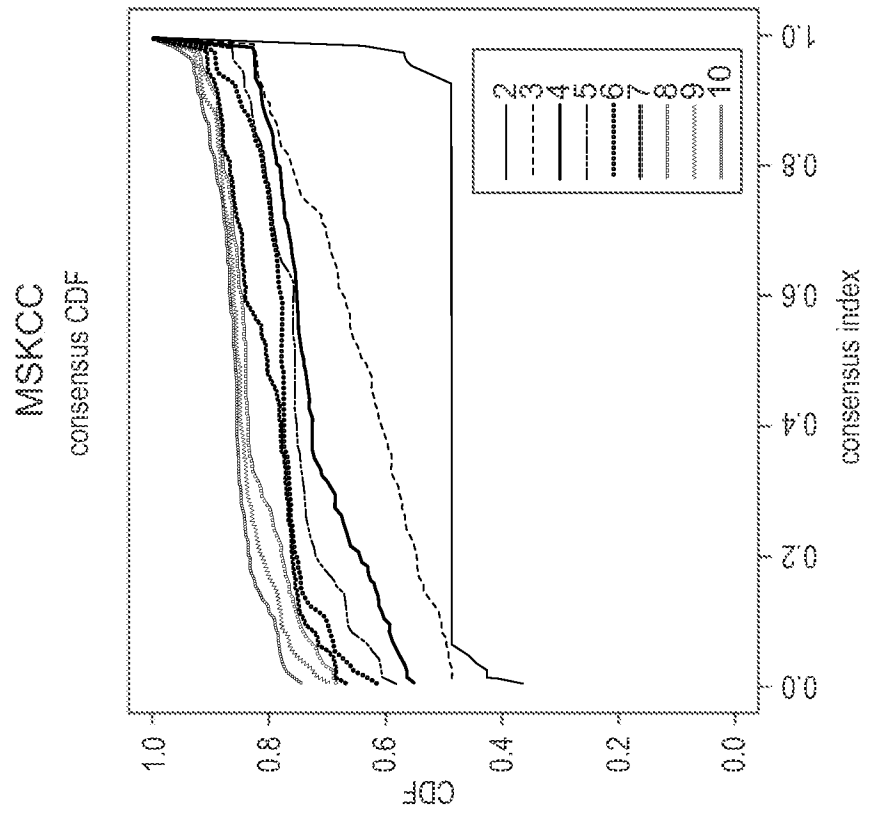

MSKCC
consensus CDF
CDF
consensus index
2
3
4
5
6
7
8
9
10
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 5B

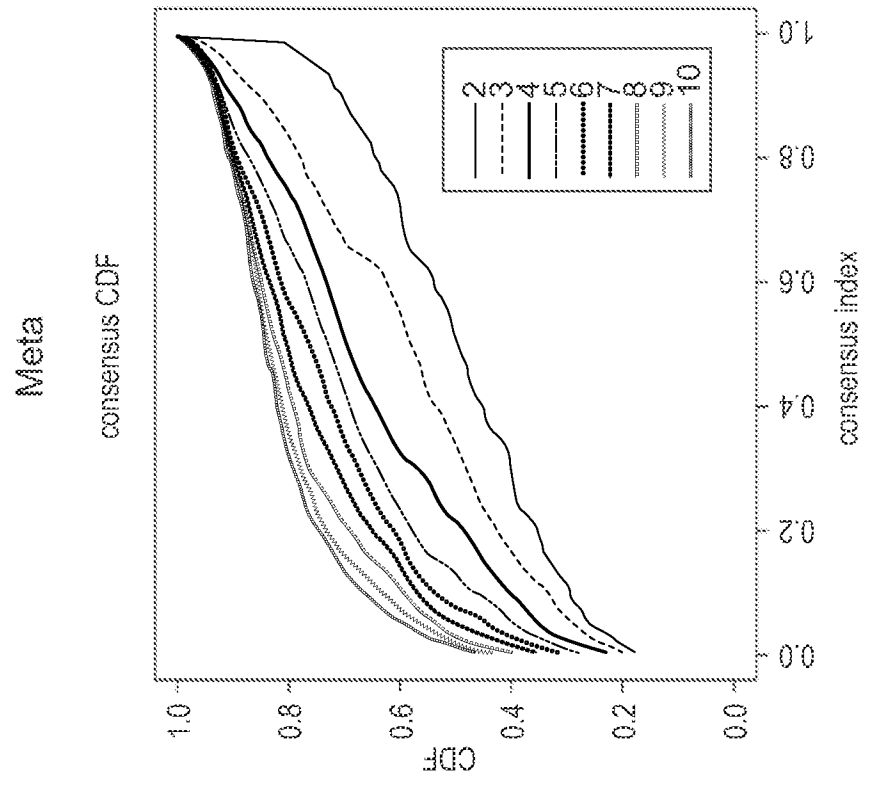

Meta
consensus CDF
CDF
consensus index
2
3
4
5
6
7
8
9
10
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 5A

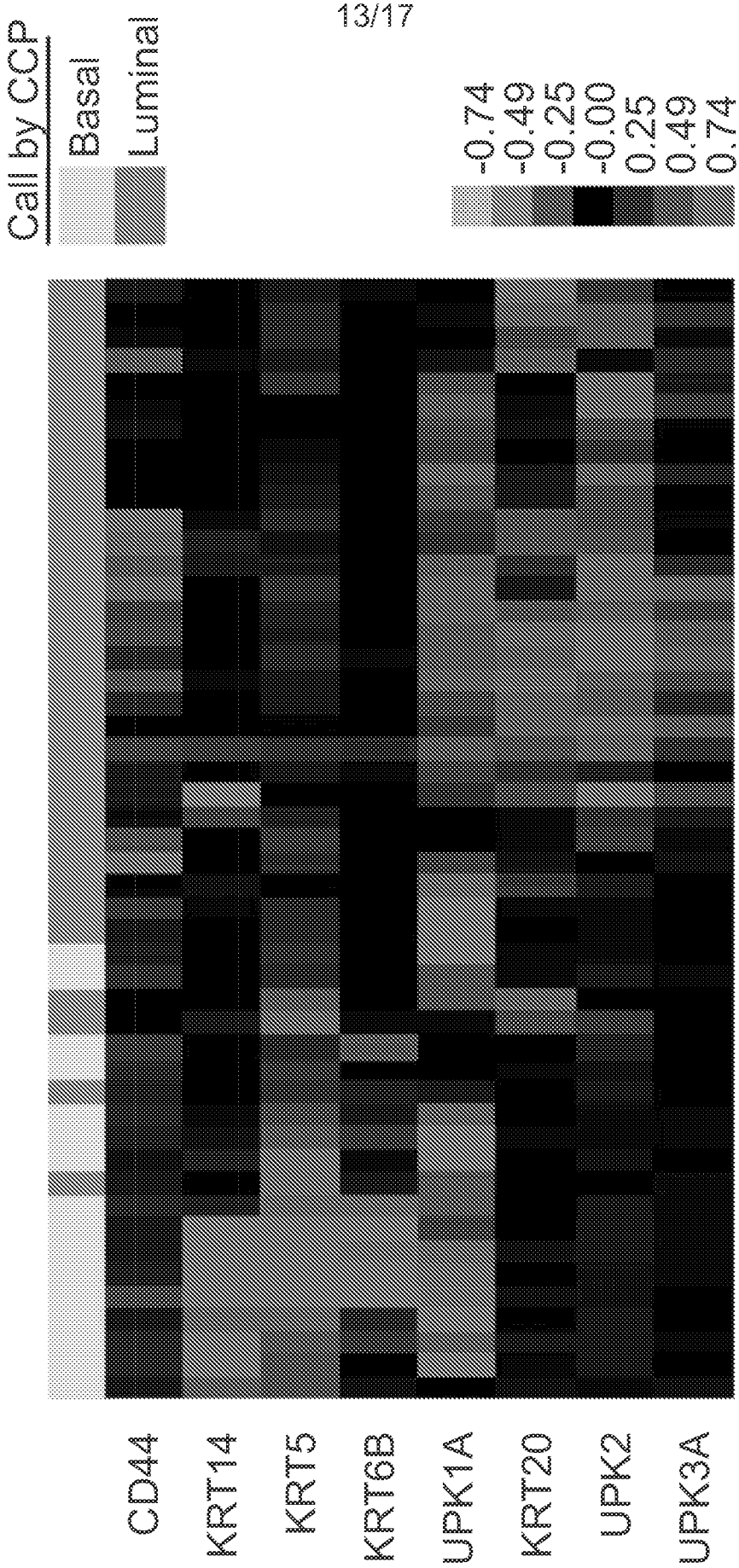

Call by CCP
Basal
Luminal
-0.74
-0.49
-0.25
-0.00
0.25
0.49
0.74
CD44
KRT14
KRT5
KRT6B
UPK1A
KRT20
UPK2
UPK3A

FIG. 5C

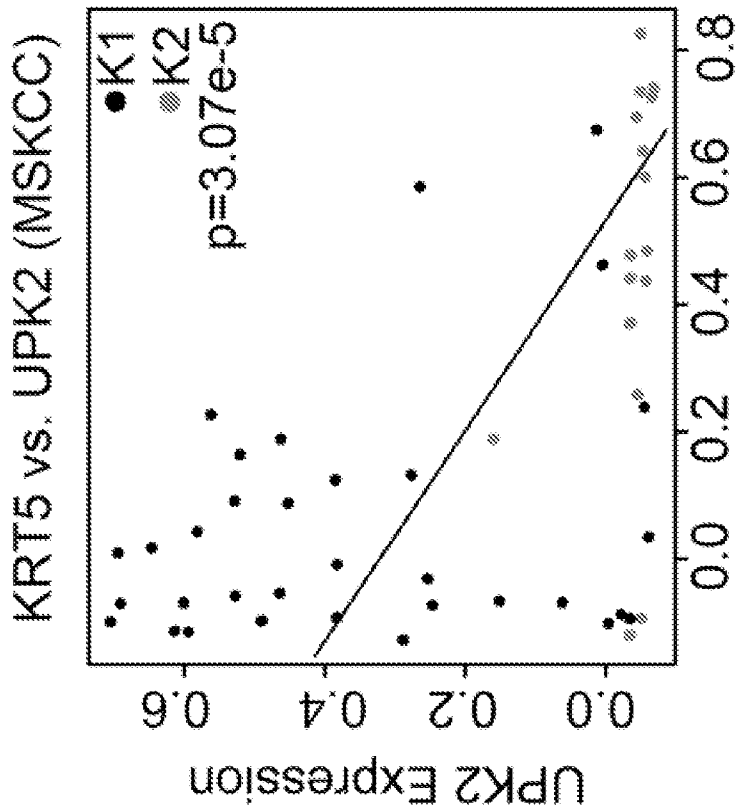

KRT5 vs. UPK2 (MSKCC)
K1
K2
p=3.07e-5
UPK2 Expression
KRT5 Expression
0.0
0.2
0.4
0.6
0.8

FIG. 5E

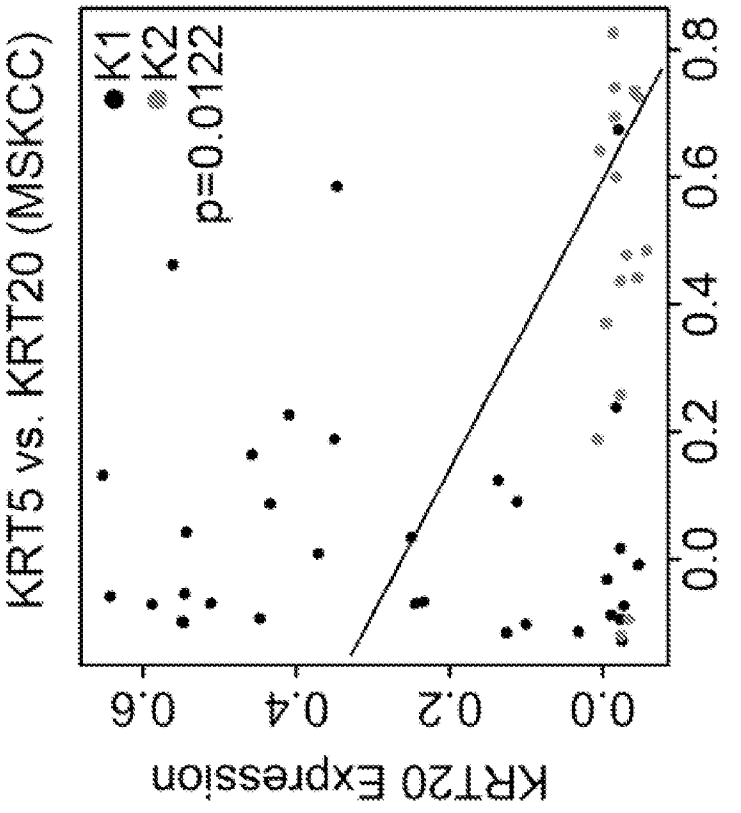
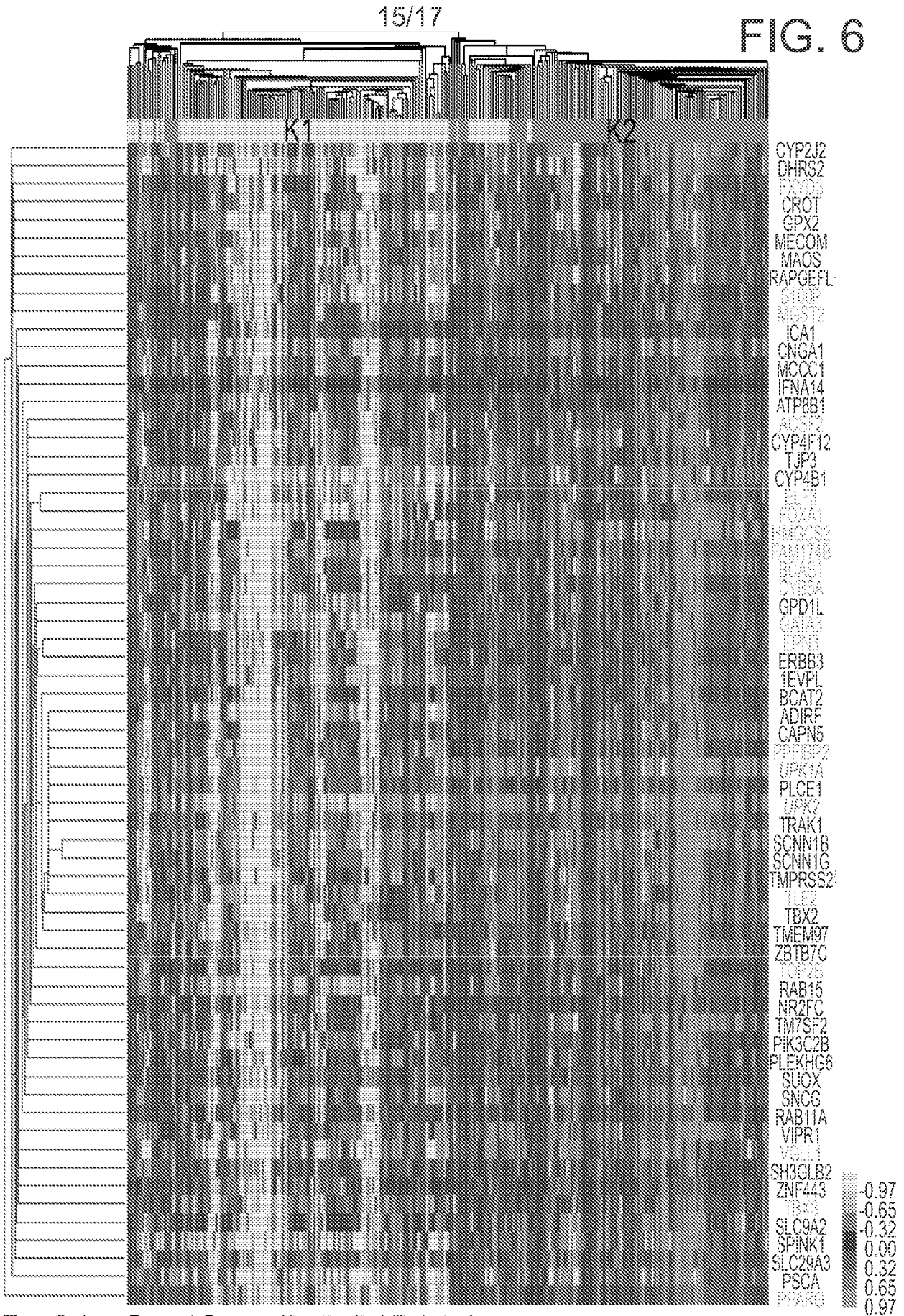

KRT5 vs. KRT20 (MSKCC)
K1
K2
p=0.0122
KRT20 Expression
KRT5 Expression
0.0
0.2
0.4
0.6
0.8

FIG. 5D

FIG. 6



Text Color : *Breast Cancer* Urothelial Related

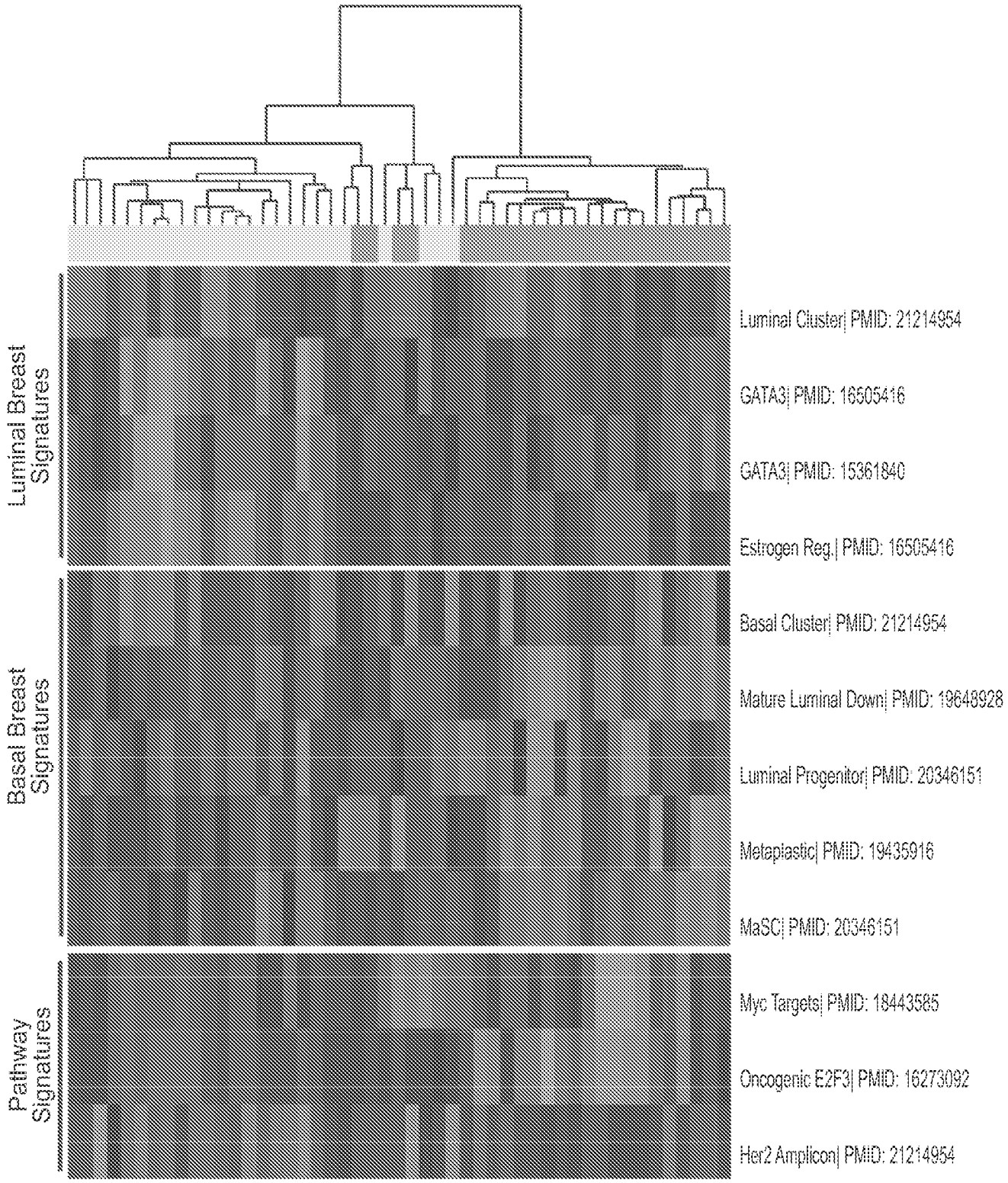
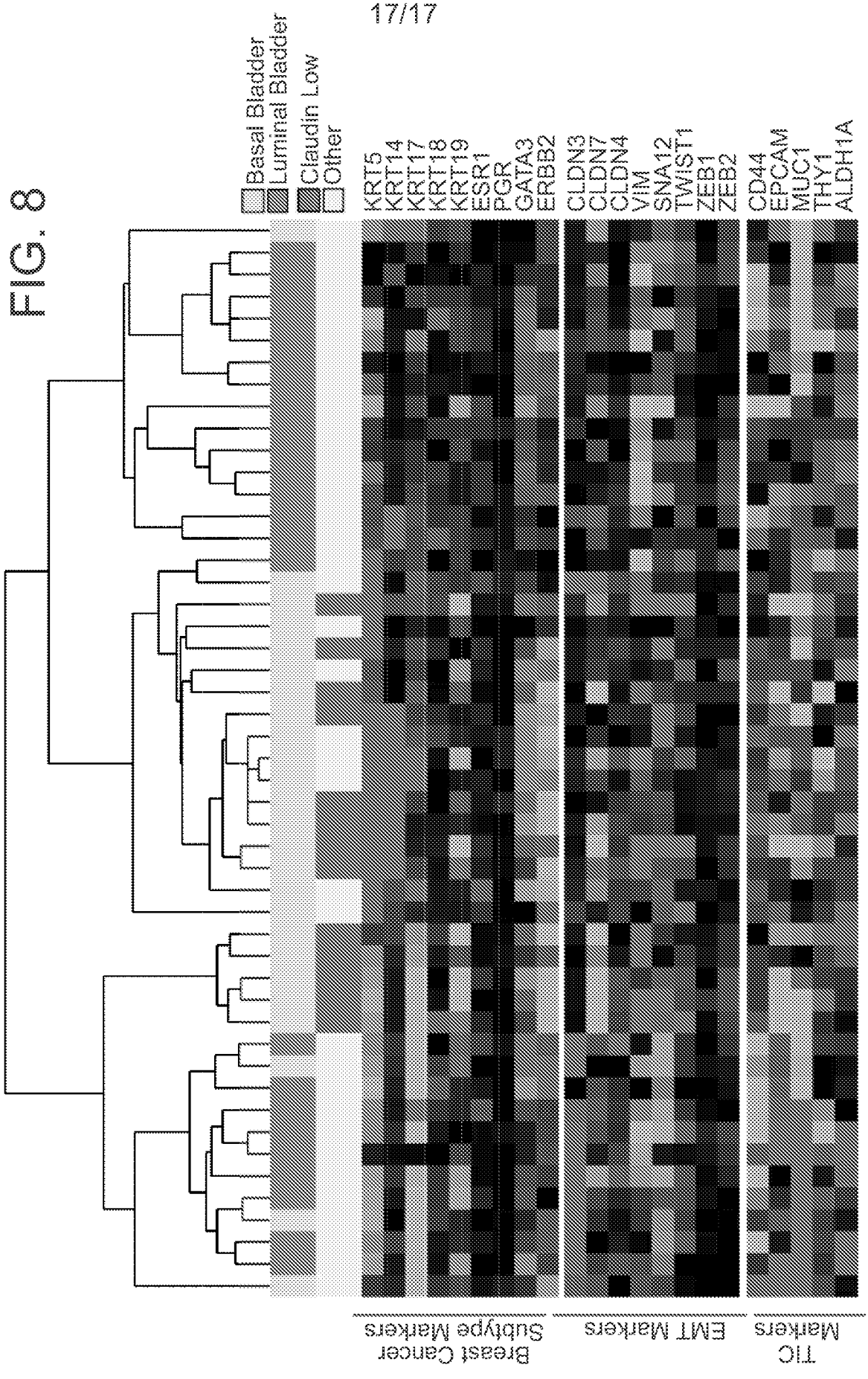

Luminal Breast Signatures
Basal Breast Signatures
Pathway Signatures
Luminal Cluster| PMID: 21214954
GATA3| PMID: 16505416
GATA3| PMID: 15361840
Estrogen Reg.| PMID: 16505416
Basal Cluster| PMID: 21214954
Mature Luminal Down| PMID: 19648928
Luminal Progenitor| PMID: 20346151
Metaplastic| PMID: 19435916
MaSC| PMID: 20346151
Myc Targets| PMID: 18443585
Oncogenic E2F3| PMID: 16273092
Her2 Amplicon| PMID: 21214954

FIG. 7

FIG. 8


Basal Bladder
Luminal Bladder
Claudin Low
Other
KRT5
KRT14
KRT17
KRT18
KRT19
ESR1
PGR
GATA3
ERBB2
CLDN3
CLDN7
CLDN4
VIM
SNA12
TWIST1
ZEB1
ZEB2
CD44
EPCAM
MUC1
THY1
ALDH1A
Breast Cancer Subtype Markers
EMT Markers
TIC Markers

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2014/065910

A. CLASSIFICATION OF SUBJECT MATTER

G01N 33/533 *(2006.01)*
G01N 33/574 *(2006.01)*
C12Q 1/68 *(2006.01)*

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)

G01N 33/533, 33/574, 33/68, C12S 3/20, C12Q 1/68

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAPATIS, PatSearch (RUPTO internal), Patentscope, Espacenet, PubMed, Google, eLIBRARY.ru

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	LINDGREN David et al. Combined Gene Expression and Genomic Profiling Define Two Intrinsic Molecular Subtypes of Urothelial Carcinoma and Gene Signatures for Molecular Grading and Outcome. Cancer Research, 2010, 70(9), pp. 3463-3472, abstract, fig. 1, 24	1, 4-7, 11, 17 2-3, 8-10, 12-16, 18, 19
X Y	WO 2004/010848 A2 (DIOMEDIA LIFE SCIENCES, INC. et al.) 05.02.2004, p. 6, line 11 - p. 11, line 19, p. 21, line 8 - p. 22, line 9, p. 23, line 22 - p. 29, line 9, p. 34, line 13 - p. 41, line 26, p. 43, line 15 - p. 61, line 27, examples, fig. 17, 18, tables 5-7	20, 23- 25 8-10, 12-16, 21, 22

[X] 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 document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
06 February 2015 (06.02.2015)	19 February 2015 (19.02.2015)
Name and mailing address of the ISA/RU: Federal Institute of Industrial Property, Berezhkovskaya nab., 30-1, Moscow, G-59, GSP-3, Russia, 125993 Facsimile No: (8-495) 531-63-18, (8-499) 243-33-37	Authorized officer E. Vinogradova Telephone No. (495)531-64-81

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2014/065910

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VOLKMER Jens-Peter et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. PNAS, February 7, 2012, Vol. 109, no. 6, pp. 2078-2083, p.2078-2083, fig. 2, 4	2, 18, 21
Y	AARON M. Gruver et al. Selective Immunohistochemical Markers to Distinguish Between Metastatic High-Grade Urothelial Carcinoma and Primary Poorly Differentiated Invasive Squamous Cell Carcinoma of the Lung. Arch Pathol Lab Med - Vol. 136, November 2012, pp. 1339-1346, p. 1339-1340, col. 1, line 19, p. 1341, col. 2, last paragraph - p. 1342, col. 1, line 10	3, 19, 22