

ARTICLE

Prognostic Power of a Tumor Differentiation Gene Signature for Bladder Urothelial Carcinomas

Qianxing Mo, Fotis Nikolos, Fengju Chen, Zoe Tramel, Yu-Cheng Lee, Kazukuni Hayashi, Jing Xiao, Jianjun Shen, Keith Syson Chan

Affiliations of authors: Department of Medicine (QM), Dan L. Duncan Comprehensive Cancer Center (QM, FC, KSC), Department of Molecular and Cellular Biology (FN, ZT, YCL, KH, JX, KSC), and Department of Epigenetics and Molecular Carcinogenesis (JS), University of Texas MD Anderson Cancer Center, Smithville, TX; Stem Cells and Regenerative Medicine (STaR) Center (KSC), Center for Drug Discovery (KSC), and Scott Department of Urology (KSC), Baylor College of Medicine, Houston, TX (KSC).

Correspondence to: Qianxing Mo, PhD, Department of Medicine, Baylor College of Medicine, Houston, TX (e-mail: qmo@bcm.edu); and Keith Syson Chan, PhD, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX (email: kc1@bcm.edu).

Abstract

Background: Muscle-invasive bladder cancers (MIBCs) cause approximately 150 000 deaths per year worldwide. Survival for MIBC patients is heterogeneous, with no clinically validated molecular markers that predict clinical outcome. Non-MIBCs (NMIBCs) generally have favorable outcome; however, a portion progress to MIBC. Hence, development of a prognostic tool that can guide decision-making is crucial for improving clinical management of bladder urothelial carcinomas.

Methods: Tumor grade is defined by pathologic evaluation of tumor cell differentiation, and it often associates with clinical outcome. The current study extrapolates this conventional wisdom and combines it with molecular profiling. We developed an 18-gene signature that molecularly defines urothelial cellular differentiation, thus classifying MIBCs and NMIBCs into two subgroups: basal and differentiated. We evaluated the prognostic capability of this “tumor differentiation signature” and three other existing gene signatures including the The Cancer Genome Atlas (TCGA; 2707 genes), MD Anderson Cancer Center (MDA; 2252 genes/2697 probes), and University of North Carolina at Chapel Hill (UNC; 47 genes) using five gene expression data sets derived from MIBC and NMIBC patients. All statistical tests were two-sided.

Results: The tumor differentiation signature demonstrated consistency and statistical robustness toward stratifying MIBC patients into different overall survival outcomes (TCGA cohort 1, $P = .03$; MDA discovery, $P = .009$; MDA validation, $P = .01$), while the other signatures were not as consistent. In addition, we analyzed the progression (Ta/T1 progressing to $\geq T2$) probability of NMIBCs. NMIBC patients with a basal tumor differentiation signature associated with worse progression outcome ($P = .008$). Gene functional term enrichment and gene set enrichment analyses revealed that genes involved in the biologic process of immune response and inflammatory response are among the most elevated within basal bladder cancers, implicating them as candidates for immune checkpoint therapies.

Conclusions: These results provide definitive evidence that a biology-prioritizing clustering methodology generates meaningful insights into patient stratification and reveals targetable molecular pathways to impact future therapeutic approach.

Bladder cancer is among the most underpublicized and understudied cancer types. Yet, bladder cancer is estimated to cause approximately 76 960 new cases and 16 390 deaths per year in the United States (1). Muscle-invasive bladder cancer (MIBC; tumor stages II–IV) accounts for the majority of patient mortality (2). No major progress has been made in improving the survival of MIBC patients over the past 30 years. The current standard-

of-care treatment is limited to neoadjuvant chemotherapy and radical cystectomy in the United States (2–4). Recently, anti-PD-L1 immunotherapy was granted accelerated approval by US Food and Drug Administration (FDA) to treat metastatic bladder cancer (5).

In the past few years, The Cancer Genome Atlas (TCGA) consortium, others, and our group (6–13) have reported

Table 1. Data sets used for the comparative analysis of MIBC gene signatures*

Data Set	Accession	No. of samples	No. of survival data	Platform	Median survival, mo	Median follow-up, mo	Censored, %
TCGA	TCGA	408	408	Illumina Hiseq for RNAseq	34	17.6	55.9
MDA discovery	GSE48277	73	73	Illumina HumanHT-12 V3.0	37.2	30.4	38.4
MDA validation	GSE48277	57	57	Illumina HumanHT-12 WG-DASL V4.0 R2	79.2	38.1	54.4
Lund	GSE32894	308	224	Illumina HumanHT-12 V3.0	NA	35.1	88.9
CancerCell2016 (NMIBC)	E-MTAB-4321	460	460	Illumina Hiseq for RNAseq	NA	33 (progression-free survival)	93.3

*MIBC = muscle-invasive bladder cancers; NMIBC = non-MIBC; TCGA = The Cancer Genome Atlas.

subclassification of bladder cancer into distinct subtypes based on unique molecular signatures (8–13). We previously reported a basal subtype of MIBCs characterized by high expression of cytokeratins (KRT14/KRT5) and cell surface receptors (Thy-1/CD44) that are generally expressed in stem and basal cells of the normal bladder urothelium (6,7). Independently, TCGA reported four clusters of MIBCs with various characteristics (9). Cluster I/II tumors resembled papillary histology, and Cluster III/IV tumors expressed high levels of cytokeratins (KRT14/KRT5), along with other surface proteins reported to mark stem/basal cells (9). Additional studies from Lund University, the MD Anderson Cancer Center (MDA), and the University of North Carolina at Chapel Hill (UNC) reported MIBCs with basal and luminal characteristics (8,10,11). Despite the independent clustering methodologies employed, there are overlaps between the different subclassifications proposed. A consensus meeting held in 2015 concluded that MIBCs can be classified into at least two major subtypes: basal/squamous-like (BASQ) and differentiated/luminal (14). State-of-the-art clinical trials now evaluate whether cancer subtyping plays a role in understanding therapeutic responses (15–18). While subclassification of MIBCs into subtypes reveals an exciting area of research, we rationalized that a methodology that assimilates biology during data analysis may provide valuable insights. Clinically, tumor grading is a pathological evaluation of tumor differentiation that often associates with patient outcome. The current study extrapolates this conventional wisdom into molecular profiling by developing a biology-based clustering methodology from a panel of 18 genes that define various stages of urothelial cell differentiation (7). The prognostic value of this newly defined “tumor differentiation signature” was compared with other published signatures in stratifying clinical outcome. The overarching goal is to demonstrate the relevance to consider tumor biology during classification of tumors into subtypes.

Methods

Statistical Methods for Clustering and Classification Analysis of the Bladder Cancer Data Sets

We evaluated the prognostic capability of four classifiers, including TCGA (2707 genes), MDA (2252 genes/2697 probes), UNC (47 genes), and the tumor differentiation (18 genes) signatures using five gene expression data sets derived from human MIBC and NMIBC patients (Table 1). The TCGA MIBC data were generated by RNA-seq, and gene expression was quantified by the RSEM method and normalized within sample to a fixed upper quartile (9). The other four data sets were generated by Illumina microarrays, and the normalized data were directly downloaded

from the GEO website (<http://www.ncbi.nlm.nih.gov/geo/>). For microarray data sets, when there were multiple probes for a gene, we chose the one with largest variance to represent the gene. To make the data sets comparable, for each gene in each data set, we standardized its expression values by subtracting the sample mean and then dividing by the sample standard deviation. We used the TCGA data set and hierarchical clustering algorithm (with 1 – Pearson correlation coefficient as the distance between samples and the “complete” agglomeration method) to generate the expression signature for the 18-gene tumor differentiation classifier in which the patient samples were clustered into two subgroups with distinct expression patterns. Using the clustered TCGA data set as the training data set and Euclidean distance based on the 18 genes as the metric to define the neighborhood among samples, we then classified the samples in the other three data sets into the two subgroups using the 1-nearest neighbor algorithm. For the other three gene signatures, if the samples in each data set had been classified, we directly used the results for comparison (ie, MDA discovery data sets). Otherwise, the 1-nearest neighbor algorithm was used for the classification analyses.

Survival and Functional Enrichment Analyses

Kaplan-Meier method was used to generate survival curves for the subgroups in each data set, and the log-rank test was used to test if they were statistically different. Moderated t-statistics were used to test if genes were differentially expressed between the groups of interest, and the Benjamini-Hochberg method was used to estimate false discovery rate (FDR) (19). Genes with an FDR of less than 0.05 were considered statistically significant. Gene functional term enrichment analysis was performed using the DAVID bioinformatics tools (<https://david.ncifcrf.gov/>) for the differentially expressed genes. Gene set enrichment analysis was performed using GSEA software (<http://software.broadinstitute.org/gsea/index.jsp>). All the statistical analyses were performed using R (<https://www.r-project.org/>), and the P values were two-sided. P values of less than .05 were considered statistically significant.

Results

Compiling a Biology-Based Tumor Differentiation Gene Signature

Tumor grade is defined by pathologic evaluation of tumor cell differentiation, which historically associates with disparate biologic behaviors (7). While all MIBCs are pathologically evaluated as high-grade cancers (T2 and higher), we hypothesized that

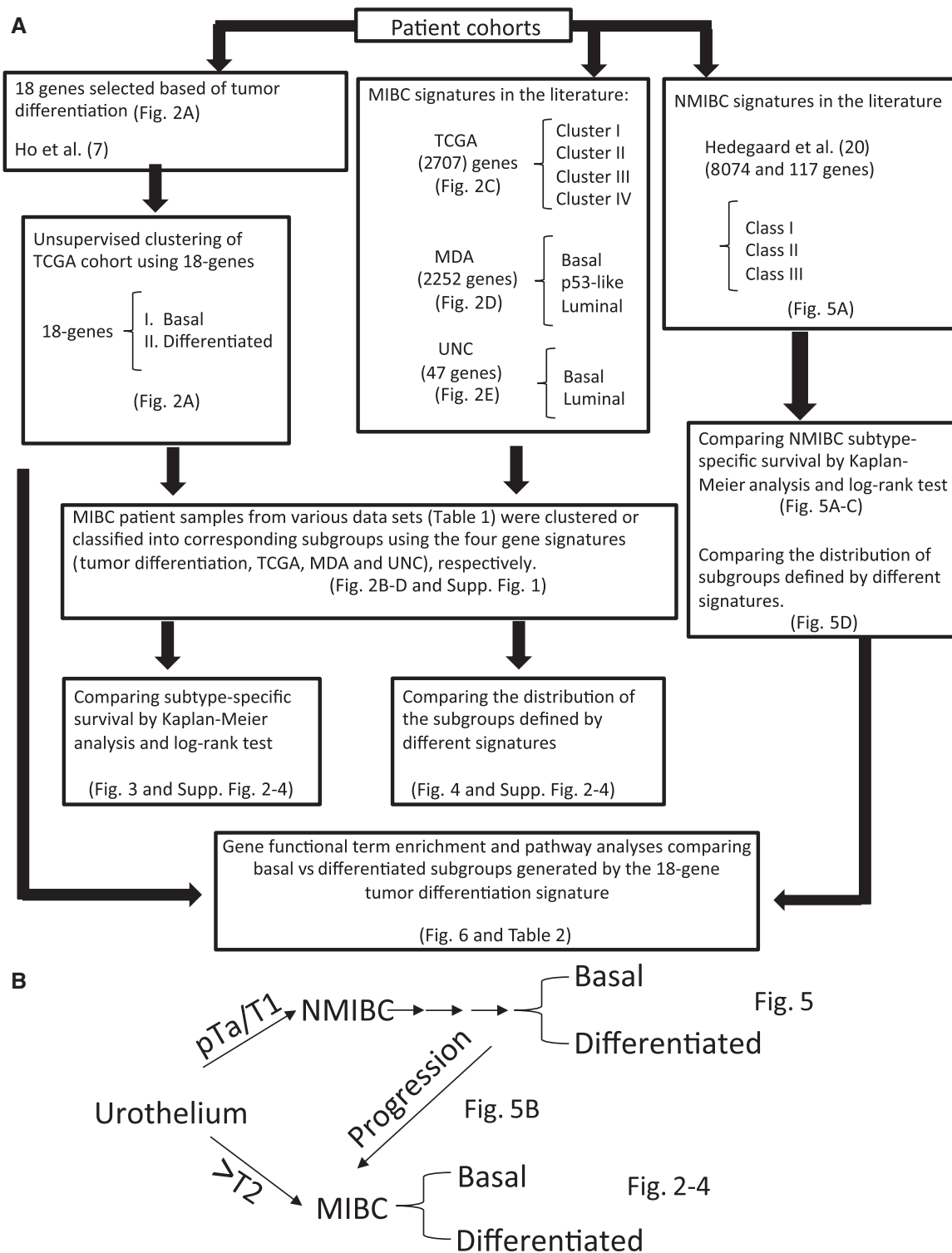


Figure 1. Flowchart of the current study. **A)** This figure summarizes the overall rationale and the sequence of statistical analyses performed to generate each of the figures and tables in the current study. **B)** Schematic of urothelial carcinoma molecular subtypes. MIBC = muscle-invasive bladder cancer; NMIBC = non-MIBC; PFS = progression-free survival.

expression of genes associated with urothelial cell differentiation (within a tumor) might provide molecular information to its corresponding differentiation status. Thus, such molecular information likely complements pathologic grading and reflects

pathobiology that might better predict outcome (6,7). First, we developed a classifier from 18 genes differentially expressed in various layers of the bladder urothelium (ie, different cellular differentiation stages) (Figure 2A). Established cytokeratins,

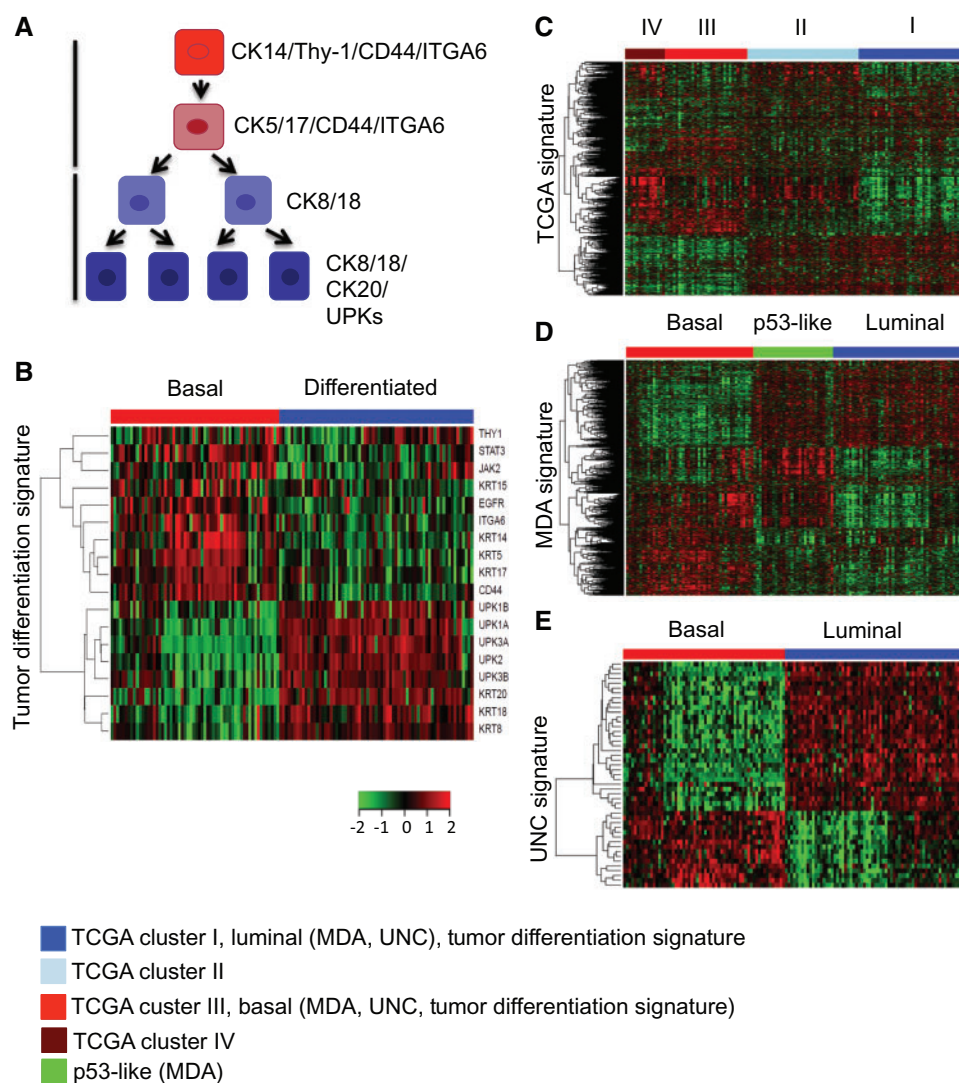


Figure 2. A biology-based tumor differentiation classifier. **A)** Schematic summary of the selection of genes expressed in basal and differentiated cells within normal bladder urothelium—for compiling the 18-gene tumor differentiation classifier. **B)** Hierarchical clustering analysis of the 129 muscle-invasive bladder cancer (MIBC) patients from The Cancer Genome Atlas (TCGA) cohort 1 using the 18 tumor differentiation genes reveals two MIBC subtypes: basal (red) and differentiated (blue). On the heatmap, green to red color indicates low to high gene expression. **C)** Published gene signature by TCGA (2707 genes) stratifies MIBC specimens into four clusters. **D and E)** Classification of MIBCs from TCGA data set into subgroups, using the 1-nearest neighbor algorithm and gene signatures from MD Anderson Cancer Center (2252 genes) and University of North Carolina at Chapel Hill (47 genes), respectively. MDA = MD Anderson Cancer Center; TCGA = The Cancer Genome Atlas consortium; UNC = University of North Carolina at Chapel Hill.

urolakins, and clusters of differentiation (CD) markers were selected on the basis of their expression corresponding to various differentiation stages or cell layers of the bladder urothelium (Figure 2A). For instance, KRT5/17 and CD44 are reported to be expressed in basal cells of human urothelium (7,21,22). KRT14, originally thought to define urothelial squamous differentiation (21,23), was also reported to be expressed in early stem/progenitor cells in the urothelial basal layer (6). A recent study employing a KRT14 Cre reporter mouse followed by lineage-tracing in bladder urothelium provided definitive evidence demonstrating that KRT14 could mark stem/progenitor cells, with the capacity of giving rise to all urothelial layers (24).

Cytokeratins 8/18 are expressed in the intermediate and more differentiated urothelial cells, while cytokeratin 20 and urolakins are expressed in terminally differentiated umbrella cells that are adjacent to the lumen (Figure 2A) (25,26). We

previously reported that constitutive expression of Stat3 drives invasive progression of MIBCs in a mouse model (27). STAT3 is known to maintain stem cell self-renewal (28), and its expression closely associates with basal KRT14 expression (27). Therefore, STAT3 and its upstream regulators, including EGFR and JAK2, were also included in the tumor differentiation signature (12).

Subclassification of Human MIBCs Into Basal and Differentiated Subtypes

We employed this 18-gene tumor differentiation signature to stratify the published TCGA cohort of 129 MIBC patients with RNA sequencing data (9) and identified two distinct subtypes (Figure 2B). The first MIBC subtype expressed a higher level of

basal cytokeratins and other genes associated with urothelial basal cells, while expression of differentiation genes was relatively lower, and was therefore designated as “basal” subtype. The second MIBC subtype expressed a relatively lower level of basal genes and higher level of differentiation genes, and was therefore designated as “differentiated” subtype.

Several molecular classification studies on human MIBCs have been published in the recent years. For instance, a published TCGA study performed unsupervised clustering analysis of 129 MIBC patient samples using 2707 genes, which stratified patients into four “clusters” (Clusters I, II, III, and IV) (Figure 2C). The MDA group performed unsupervised clustering analysis of 73 primary fresh-frozen MIBC specimens using 2252 genes (2697 probes) that generated three subtypes including basal, p53-like, and luminal (10). The UNC group performed consensus clustering analysis of a meta-data set that identified basal and luminal subtypes and further identified 47 genes as the subtype predictor (11). Using the 18-gene tumor differentiation signature and the other three existing gene signatures (TCGA/MDA/UNC), we classified patient specimens in the testing data sets into corresponding subtypes using the 1-nearest neighbor method with Euclidean distance. The MDA gene signature and the UNC gene signature presented by the TCGA data set are shown in Figure 2, D and E, respectively, where patient specimens were classified using the 1-nearest neighbor method and genes were arranged by hierarchical clustering using 1 – Pearson correlation coefficient as distance and complete agglomeration algorithm. The heatmap representation of these four gene signatures presented by different published data sets is shown in Supplementary Figure 1A (available online).

Prognostic Value of Tumor Differentiation Signature in Stratifying Patient Survival in Comparison With Other Existing Gene Signatures

Next, we investigated whether the basal and differentiated subtypes defined by the tumor differentiation signature using the published TCGA cohort (9) had prognostic value. The basal MIBC subtype had worse overall survival compared with the differentiated subtype ($P = .03$) (Figure 3A). Similarly, the UNC basal subtype also had worse overall survival compared with the luminal subtype ($P = .04$) (Figure 3J). In contrast, in terms of overall survival, the TCGA subtypes were not statistically significantly different ($P = .19$) (Figure 3D), and the MDA subtypes were marginally statistically significantly different ($P = .09$) (Figure 3G).

We further evaluated the prognostic value of the tumor differentiation signature in additional MIBC cohorts (Table 1). The tumor differentiation signature reproducibly stratifies overall survival in these cohorts, with the basal subtype associated with a statistically significantly worse overall survival ($P = .009$, $P = .01$, $P = .001$) (Figure 3, B and C; Supplementary Figure 4A, available online). The other three signatures demonstrated a trend in stratifying overall survival, and there were instances where some signatures demonstrated statistical significance (Figure 3, E and F, H and I, and K–L; Supplementary Figures 2A and 4A, available online). Taken together, these results demonstrated the statistical robustness and consistency of the tumor differentiation signature in stratifying overall survival across multiple data sets. Moreover, as the general perception was that TCGA Cluster I/II and Cluster III/IV tumors resemble differentiated (or luminal) and basal tumors, respectively, we performed additional survival analysis by combining Cluster I/II and Cluster III/IV tumors (Supplementary Figure 3, A–D). The results

revealed that the basal and differentiated tumors defined by the tumor differentiation signature are considerably different from combining Cluster III/IV and Cluster I/II, respectively. Combination of Cluster I/II tumors is not consistent in stratifying patient survival from Cluster III/IV tumors.

Comparison of MIBC Subtype Assignments by Different Gene Signatures

While in reviewed literature, basal and luminal (or differentiated) subtypes of MIBCs stratified by different signatures correspond well (29); raw data supporting such claims have not been available, and therefore warrant careful evaluation.

Figure 4 summarizes the relative cancer subtype assignments by each of the four gene signatures evaluated in the current study and illustrates how they relate to their counterparts in other studies. For every subtype or cluster, we determined the proportion of specimens that would be stratified as basal or differentiated subtypes based on the tumor differentiation signature. These data demonstrate that the TCGA cluster III and IV tumors correspond well with the basal cancer subtypes classified by MDA, UNC, and the tumor differentiation signature (Figure 4A; Supplementary Figures 2B and 4B, available online). In the TCGA and MDA discovery cohorts, we found that most of the Cluster III and IV tumors had a gene signature that was considered basal (Figure 4D; Supplementary Figures 2C and 4C, available online).

Within the MDA discovery and validation cohorts, the tumor differentiation signature classified specimens to basal or differentiated, which highly overlap with the basal and luminal subtypes assigned by the MDA group, respectively; however, the “p53-like tumors” were divided between basal and differentiated subtypes (Figure 4, A–C). While earlier reports suggested that p53-like tumors were mainly luminal, these findings further corroborate recent studies recognizing that the “p53-like” subtype is composed of both luminal and basal tumors (30). More than 50.0% of the “p53-like tumors” were classified as basal tumors by the UNC signature within the MDA validation cohort (Figure 4C); 73.2% of the TCGA Cluster I tumors and 83.3% of the TCGA Cluster II tumors had a gene signature that was considered “differentiated” by the tumor differentiation signature (Figure 4, A and D; Supplementary Figures 2, B and C, and 4, B and C, available online). These results are insightful in laying out the variability among subtype assignments by different signatures in different patient cohorts.

Tumor Differentiation Signature as a Predictor of PFS in NMIBC

NMIBCs have a favorable prognosis. However, 15.0% to 20.0% of NMIBCs eventually progress to aggressive MIBCs (20,31). Predicting NMIBC progression and the underlying molecular pathways involved has been an overarching goal. Here, we examined the power of the tumor differentiation signature to predict disease progression in NMIBC patients using a published data set of 460 NMIBC patients (20). The tumor differentiation gene signature stratified patients into basal and differentiated groups (Figure 5A), with 83.3% (383/460) of NMIBC patients clustered into the differentiated group. The basal tumors were strongly associated with worse progression-free survival ($P = .008$) (Figure 5B) and progression to T2 stage (basal: 12.9%, 95% CI = 7.1% to 22.1%; vs differentiated: 5.1%, 95% CI = 3.1% to 7.9%; $P = .009$) (Figure 5C). The distributions of the NMIBC patients

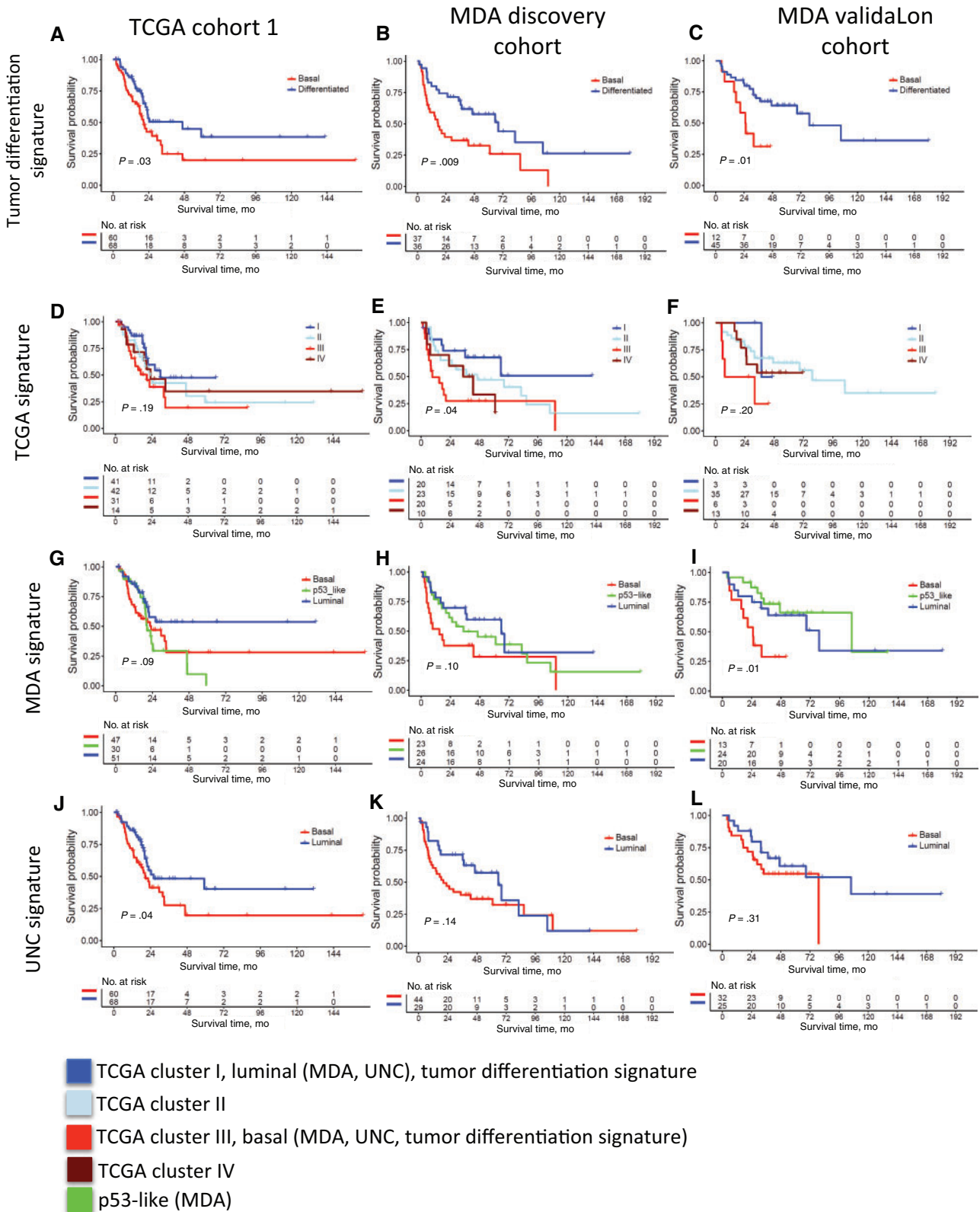


Figure 3. Comparison of tumor differentiation signature to other existing signatures in stratifying overall survival of muscle-invasive bladder cancer (MIBC) patients. A–C) Application of the tumor differentiation signature in The Cancer Genome Atlas consortium (TCGA), MD Anderson Cancer Center (MDA) discovery, and MDA validation cohorts of MIBC patients in direct comparison with the TCGA (D–F), MDA (G–I), and University of North Carolina at Chapel Hill (J–L) signatures in stratifying overall survival of MIBC patients. P values are two-sided, log-rank test. MDA = MD Anderson Cancer Center; MIBC = muscle-invasive bladder cancer; TCGA = The Cancer Genome Atlas consortium; UNC = University of North Carolina at Chapel Hill.

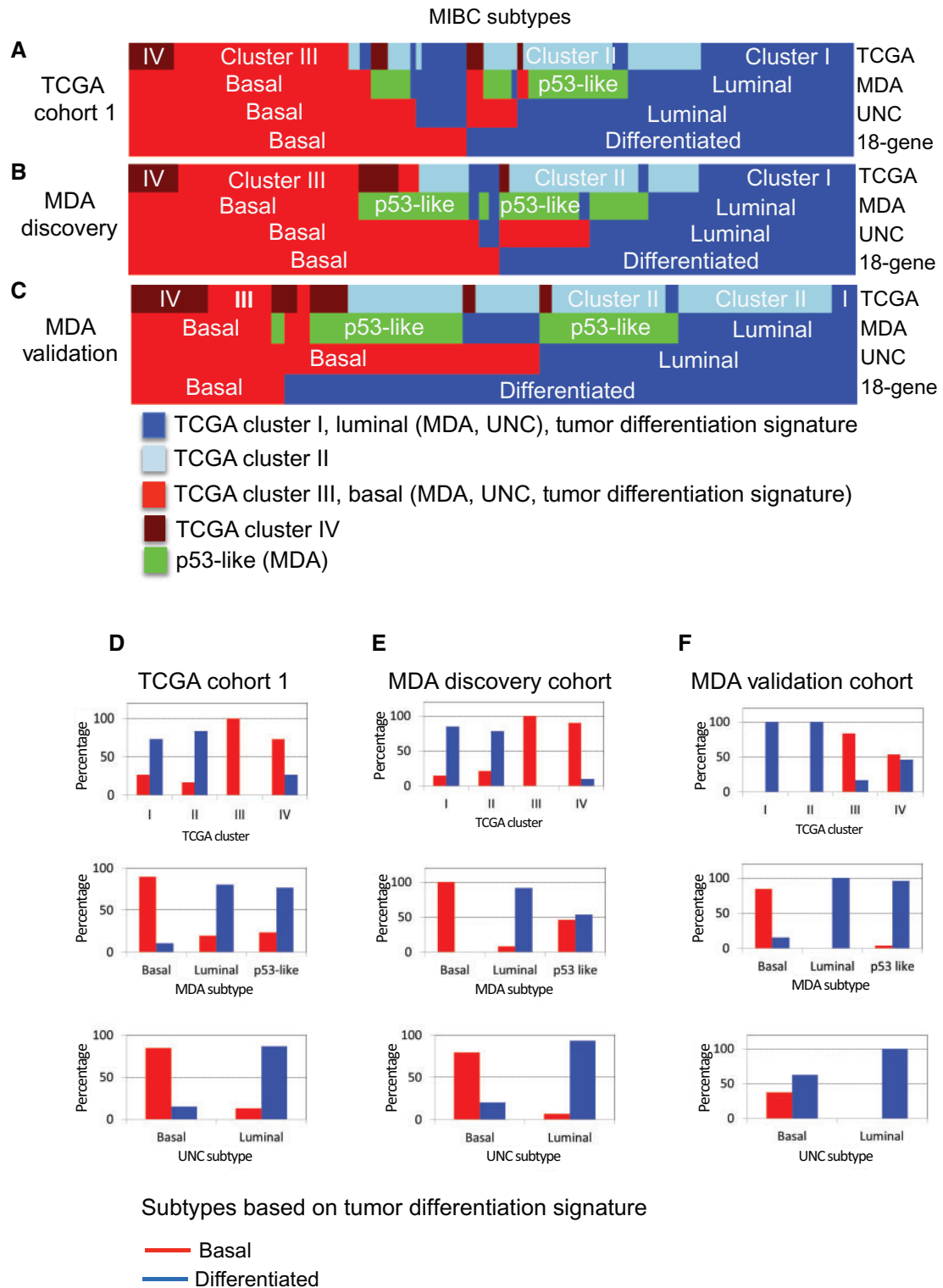


Figure 4. Muscle-invasive bladder cancer (MIBC) subtype assignments by different gene signatures. A–C) Colorimetric chart demonstrating the distribution of MIBC patients assigned by The Cancer Genome Atlas consortium (TCGA)/MD Anderson Cancer Center (MDA)/University of North Carolina at Chapel Hill (UNC), and the tumor differentiation signatures and their relation to other signatures. D–F) The percentage of basal and differentiated MIBCs assigned by the tumor differentiation signature in the subtype assignments by the TCGA/MDA/UNC signatures in multiple cohorts of patients. MDA = MD Anderson Cancer Center; MIBC = muscle-invasive bladder cancer; TCGA = The Cancer Genome Atlas consortium; UNC = University of North Carolina at Chapel Hill.

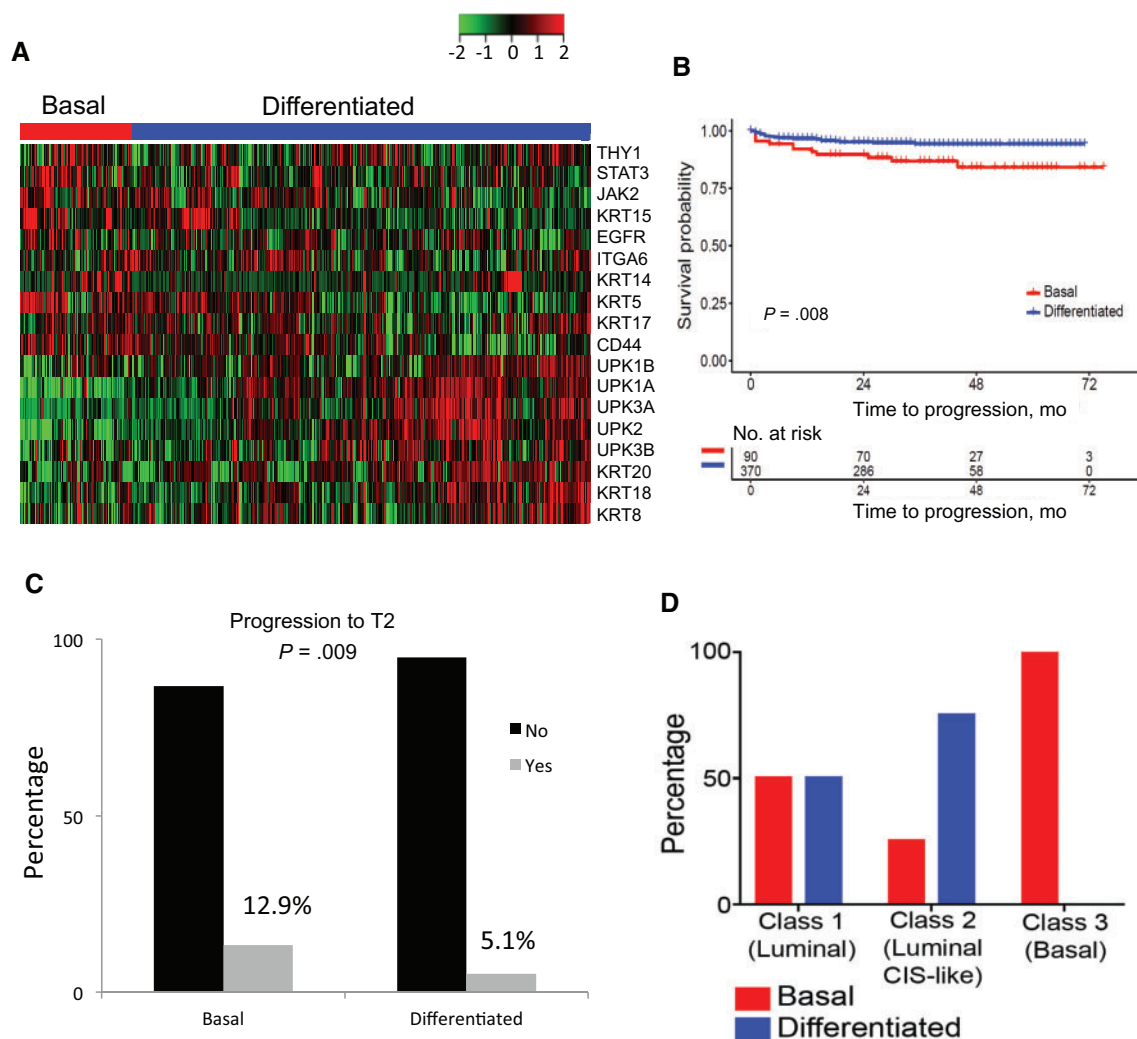


Figure 5. Tumor differentiation signature as a predictor of progression-free survival (PFS) in non-muscle-invasive bladder cancer (NMIBC). **A**) The 460 NMIBC patients from Hagedaard et al. (20) are classified to two subtypes: basal (red) and differentiated (blue) using the tumor differentiation signature. On the heatmap, green to red color indicates low to high gene expression. **B**) Kaplan-Meier analysis of progression-free survival (PFS) in the NMIBC cohort ($P = .008$, log-rank test, two-sided). **C**) Percentage of patients progressing to T2 (basal: 12.9%, 95% confidence interval [CI] = 7.1% to 22.1%; differentiated: 5.1%, 95% CI = 3.1% to 7.9%; $P = .009$, Fisher exact test, two-sided). **D**) Percentage of basal and differentiated subtypes in the NMIBC patients that progressed to T2.

who progressed to T2 (20) in the basal and differential subtypes are shown in Figure 5D. Collectively, these results indicate that the tumor differentiation status of NMIBCs is one determining factor that influences tumor progression, but clearly not the sole factor.

Pathways Associated With Basal and Differentiated Subtypes Defined by the Tumor Differentiation Signature

As the basal and differentiated MIBC subtypes were associated with overall survival, we further investigated the molecular pathways associated with each subtype. Gene functional term enrichment analysis of upregulated genes revealed many overlaps of the top 10 pathways enriched in the basal subtype of MIBCs and NMIBCs (Figure 6, A–E). These results indicate that common pathways, such as those related to epidermis development, cell adhesion, and inflammatory, immune, and defense

responses, are activated in the basal subtype. Gene set enrichment analysis (GSEA) identified members of cytokines and the cytokine receptor family, as well as toll-like receptors (TLRs) and the chemokine (CXCL) family, as elevated in basal MIBCs (Figure 6F, Table 2), indicating an inflammatory and immune-infiltrating microenvironment (32). This is further supported by co-expression of genes corresponding to various T-cell subsets, as well as a master regulator of cancer inflammation—STAT3 (Figure 6G) (33). While further analysis and functional experiments are needed to determine the pro- or antitumoral roles of these immune infiltrates, we found that basal MIBCs also highly expressed genes encoding cytotoxic T-lymphocyte-associated protein 4 (CTLA4) and programmed death-ligand 1 (PD-L1 or CD274) with established roles in immunosuppression, as well as other genes with immune regulatory roles (Figure 6, G and H; Supplementary Figure 5, available online) (34–36).

For the differentiated subtype of MIBCs, among the most statistically significant pathways are fatty acid metabolism, drug

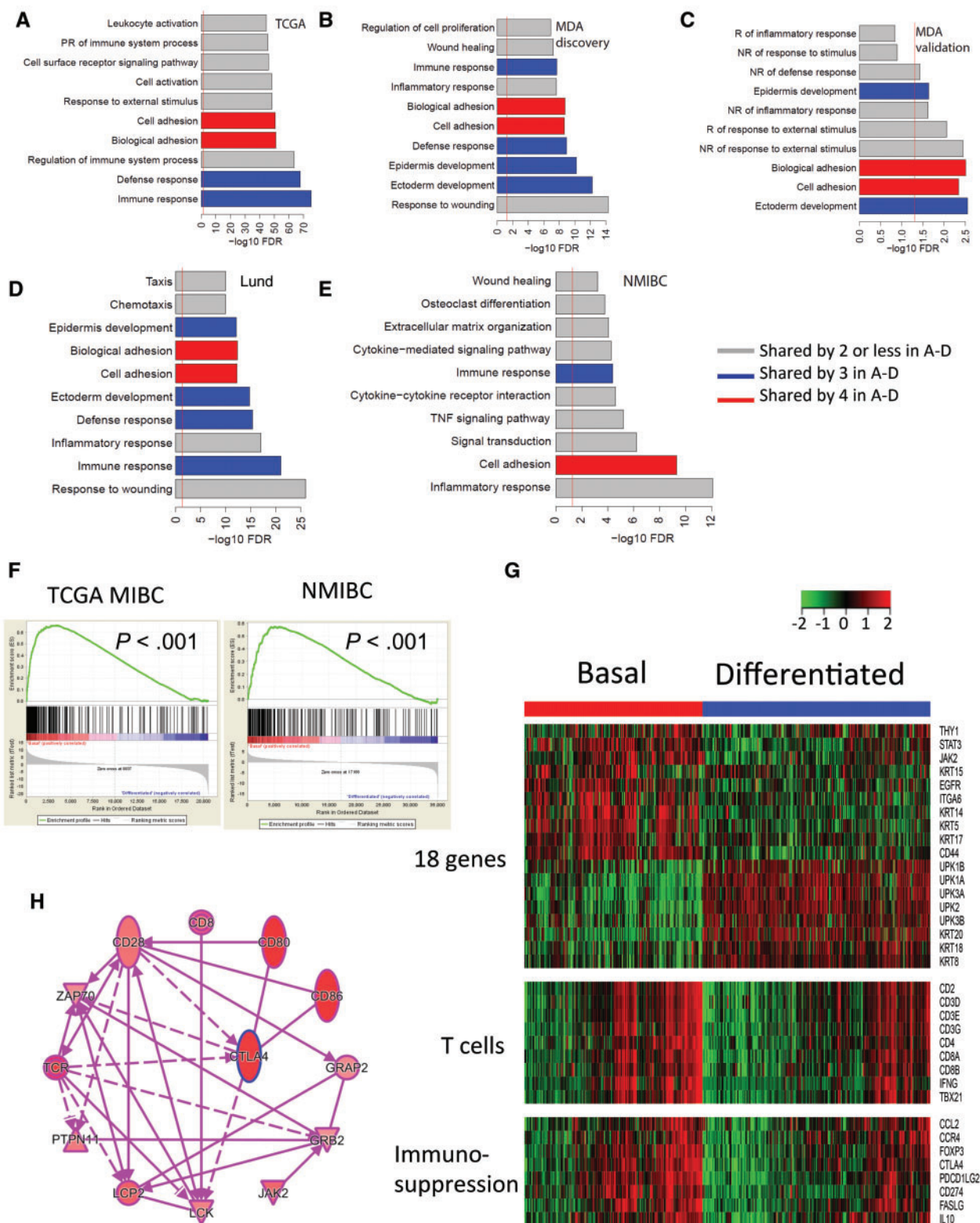


Figure 6. Summary of gene functional terms and gene set enrichment analysis (GSEA). **A–D)** The top 10 most statistically significantly represented biologic processes within the basal subtype of muscle-invasive bladder cancers (MIBCs) and non-MIBCs (**E**) from each of the data sets analyzed (The Cancer Genome Atlas [TCGA]/MD Anderson Cancer Center [MDA] discovery/MDA validation/Lund, NMIBC). **F)** Immune gene enrichment analysis in the TCGA and NMIBC cohorts. The *P* values are two-sided, GSEA test. **G)** Heatmap of immune cell-specific gene clusters in the TCGA cohort (a gene can belong to different gene clusters). **H)** Pathway analysis summarizing the interacting regulatory role for a panel of upregulated genes with CTLA4 within the “immune response” and “inflammatory response” groups. Lund = University of Lund; MDA = MD Anderson Cancer Center; MIBC = muscle-invasive bladder cancer; NR = negative regulation; PR = positive regulation; R = regulation; TCGA = The Cancer Genome Atlas consortium; UNC = University of North Carolina at Chapel Hill.

Table 2. KEGG pathway enrichment analysis of basal and differentiated MIBCs from the TCGA cohort with 408 samples*

Name	Size	NES	FDR
KEGG pathways enriched in basal subtype			
CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	264	2.63	<0.001
GRAFT_VERSUS_HOST_DISEASE	37	2.58	<0.001
LEISHMANIA_INFECTION	70	2.54	<0.001
HEMATOPOIETIC_CELL_LINEAGE	84	2.51	<0.001
NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	132	2.47	<0.001
ALLOGRAFT_REJECTION	35	2.45	<0.001
AUTOIMMUNE_THYROID_DISEASE	50	2.45	<0.001
ANTIGEN_PROCESSING_AND_PRESENTATION	81	2.41	<0.001
TYPE_I_DIABETES_MELLITUS	41	2.33	<0.001
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	62	2.33	<0.001
JAK_STAT_SIGNALING_PATHWAY	155	2.30	<0.001
TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	102	2.28	<0.001
ASTHMA	28	2.17	<0.001
INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	46	2.15	<0.001
CHEMOKINE_SIGNALING_PATHWAY	188	2.09	<0.001
PRIMARY_IMMUNODEFICIENCY	35	2.07	<0.001
ECM_RECEPTOR_INTERACTION	83	2.06	<0.001
CYTOSOLIC_DNA_SENSING_PATHWAY	55	2.06	<0.001
KEGG pathways enriched in differentiated subtype			
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	69	-2.31	<0.001
DRUG_METABOLISM_CYTOCHROME_P450	71	-2.20	<0.001
RETINOL_METABOLISM	63	-2.16	<0.001
STEROID_HORMONE_BIOSYNTHESIS	54	-2.13	<0.001
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	27	-2.08	<0.001
LINOLEIC_ACID_METABOLISM	28	-1.97	<0.001
PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	40	-1.91	0.0016
ASCORBATE_AND_ALDARATE_METABOLISM	24	-1.90	0.0015
FATTY_ACID_METABOLISM	40	-1.83	0.0042
TYROSINE_METABOLISM	42	-1.72	0.015
KEGG_STARCH_AND_SUCROSE_METABOLISM	49	-1.72	0.014
KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	18	-1.69	0.017
KEGG_ARACHIDONIC_ACID_METABOLISM	57	-1.68	0.019
PPAR_SIGNALING_PATHWAY	69	-1.56	0.057

*FDR = false discovery rate; KEGG = Kyoto Encyclopedia of Genes and Genomes; MIBC = muscle-invasive bladder cancers; TCGA = The Cancer Genome Atlas; NES = normalized enrichment score.

metabolism, and peroxisome proliferator activated receptor gamma (PPARG) pathways (Table 2). Drug-induced activation of PPARG has been previously reported to promote cellular differentiation in normal human urothelial cells, demonstrated by upregulation of differentiation markers such as UPKs and KRT20 (7,37,38). Such a phenotype is reminiscent of the differentiated subtype of MIBCs, which express high levels of differentiated genes, for example, UPKs and KRT20 (Figure 2B).

Discussion

Molecular classification of human epithelial cancers (eg, breast cancer) into distinct molecular subtypes has opened an avenue to personalized medicine, for example, herceptin for human epidermal growth factor receptor 2 (HER2)-positive breast cancer patients (39). Such clinical advances brought enthusiasm to genomic profiling of other cancer types. Until recently, bladder urothelial carcinomas lagged behind in terms of molecular subtyping. TCGA and multiple groups have published molecular signatures that classify bladder cancers into distinct subgroups (8–13). Each of these signatures has their own validity, as they were either geared toward a specific clinical outcome or utilized

a conventional unsupervised hierarchical clustering strategy without necessarily considering cancer biology during data analysis. Our study is unique in that it shows that differentiation state can be used to classify MIBC tumors into biologically meaningful subtypes: namely “basal” and “differentiated” subtypes. When applied to multiple published data sets, this biology-based gene signature is highly consistent in stratifying patient overall survival.

An important goal of subtyping is the identification of molecular pathways deregulated within each cancer subtype, which provides opportunities for precision medicine. Our pathway analysis of basal and differentiated tumor subtypes revealed FDA-approved drug targets for basal MIBCs (eg, CTLA-4 and PD-L1), as well as master transcriptional regulators of cellular differentiation that associate with each subtype (eg, Stat3 for basal subtype and PPARG for differentiated subtype), supporting our conception that bladder cancer is a disease of aberrant cellular differentiation (7). In fact, a mouse model overexpressing active Stat3 in urothelial basal cells demonstrated features of human basal bladder cancers (27).

While anti-CTLA4 and anti-PD-L1 immunotherapies are highly effective in certain patient subpopulations, a major clinical challenge is the lack of predictive markers available to

stratify responders and nonresponders. Our findings indicate that patients with basal MIBC are candidates for immune checkpoint therapies, as they express high levels of CTLA-4 and PD-L1, with other genes implicating heavy immune infiltration. Recent clinical trials testing Atezolizumab (anti-PD-L1 antibody) in chemotherapy-refractory (16) and cisplatin-ineligible MIBC patients (18) revealed partial support. Greater response to Atezolizumab was seen among patients classified as TCGA Cluster II (within differentiated/luminal) and Cluster III (within basal) (16,18). These results suggest that there are MIBC patient populations within basal and differentiated subtypes that respond differently to Atezolizumab. Our data also led us to speculate that the differentiated and basal subtypes may be further divided into subgroups. For instance, certain patients within the differentiated subtype revealed a considerable expression of stem/basal cell genes, for example, Thy-1, CD44, and ITGA6 (6,40), and genes indicative of immune infiltration, which might correspond to Cluster II responders to Atezolizumab. These observations raise the notion that an interplay between stem/basal cancer cells and tumor-infiltrating immune cells may modulate response to immune checkpoint therapies. It will be important to analyze gene expression data from these checkpoint inhibition clinical trials to validate our hypothesis. Additional findings from the Atezolizumab trials revealed that Cluster III/basal tumors responded to immunotherapy. However, they are not necessarily the best responders, posing an urgent need to identify additional uncharacterized immunosuppressive mechanisms in basal MIBC patients.

Recently, another meticulous study performed an integrated analysis across 12 cancer types and found that certain cancer subtypes are closely related to their tissue of origin (41). In their objective and comparative analysis, bladder cancer was found to be one of the most diverse cancer types, with specimens clustered into seven out of 11 major subtypes (41). Here, by employing a biology-based clustering method, we could define such a molecularly diverse cancer type (41) into two biologically distinct subtypes that statistically associated with clinical prognosis. Importantly, this tumor differentiation gene signature outperforms the TCGA signature in predicting patient survival. This concept is further supported by recent studies in other cancer types, for example, acute myeloid leukemia (42) and prostate cancer (43), where stem cell genes were successfully used to stratify patients into groups with differential survival outcome (42) or to identify patients associated with pathologically aggressive cancers (43), respectively. With further identification and characterization of study models that recapitulate the bladder cancer subtypes (as in [44–46]), pre-clinical studies can be designed to reveal valid drug targets (47,48), which can impact patient care for bladder urothelial carcinomas.

Our study has its limitations. We utilized published data sets, which renders our study design retrospective. Therefore, it is essential to validate our findings in prospective studies of treatment-naïve MIBC patients. Another limitation is that the analysis is primarily based on treatment-naïve MIBC patient cohorts. Classification of patient cohorts treated with different regimens (eg, chemotherapy or immunotherapy) using the tumor differentiation signature will provide important information for understanding its usefulness in selecting patients that might benefit from the given therapies.

Funding

This work was supported in part by the National Cancer Institute (National Institutes of Health [NIH]; R01CA175397

to KSC) and the Cancer Prevention Research Institute of Texas (CPRIT; CPRIT RP140252 to KSC).

QM and FC are supported in part by the National Cancer Institute Center Core Grants 5P30CA125123 (to CKO). KSC is supported in part by the V Scholar Award. KH is supported in part by NIH T32GM088129, and JJS is supported in part by CPRIT Core Facility Support Awards RP120348 and RP170002.

Notes

The funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

The authors thank the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performance computing resources. All authors have nothing to disclose.

References

1. American Cancer Society. January 5, 2017. <http://www.cancer.org/cancer/bladdercancer/detailedguide/bladder-cancer-key-statistics>. Accessed November 7, 2017.
2. Prasad SM, Decastro GJ, Steinberg GD, Medscape. Urothelial carcinoma of the bladder: Definition, treatment and future efforts. *Nat Rev Urol*. 2011;8(11):631–642.
3. Knowles MA, Hurst CD. Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat Rev Cancer*. 2015;15(1):25–41.
4. Costello JC, Theodorescu D. Decade in review-bladder cancer: International progress: From cytology to genomics. *Nat Rev Urol*. 2014;11(11):609–610.
5. Powles T, Eder JP, Fine GD, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature*. 2014;515(7528):558–562.
6. Volkmer JP, Sahoo D, Chin RK, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc Natl Acad Sci U S A*. 2012;109(6):2078–2083.
7. Ho PL, Kurtova A, Chan KS. Normal and neoplastic urothelial stem cells: Getting to the root of the problem. *Nat Rev Urol*. 2012;9(10):583–594.
8. Sjodahl G, Lauss M, Lovgren K, et al. A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res*. 2012;18(12):3377–3386.
9. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315–322.
10. Choi W, Porten S, Kim S, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell*. 2014;25(2):152–165.
11. Damrauer JS, Hoadley KA, Chism DD, et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A*. 2014;111(8):3110–3115.
12. Rebouissou S, Bernard-Pierrot I, de Reynies A, et al. EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med*. 2014;6(244):244ra91.
13. Biton A, Bernard-Pierrot I, Lou Y, et al. Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. *Cell Rep*. 2014;9(4):1235–1245.
14. Lerner SP, McConkey DJ, Hoadley KA, et al. Bladder cancer molecular taxonomy: Summary from a consensus meeting. *Bladder Cancer*. 2016;2(1):37–47.
15. Seiler R, Ashab HA, Erho N, et al. Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur Urol*. In press.
16. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: A single-arm, multicentre, phase 2 trial. *Lancet*. 2016;387(10031):1909–1920.
17. McConkey DJ, Choi W, Shen Y, et al. A prognostic gene expression signature in the molecular classification of chemotherapy-naïve urothelial cancer is predictive of clinical outcomes from neoadjuvant chemotherapy: A phase 2 trial of dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin with bevacizumab in urothelial cancer. *Eur Urol*. 2016;69(5):855–862.
18. Balar AV, Galsky MD, Rosenberg JE, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: A single-arm, multicentre, phase 2 trial. *Lancet*. 2017;389(10064):67–76.
19. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3: Article3.

20. Hedegaard J, Lamy P, Nordentoft I, et al. Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer Cell*. 2016;30(1):27–42.
21. Southgate J, Hutton KA, Thomas DF, Trejdosiewicz LK. Normal human urothelial cells in vitro: Proliferation and induction of stratification. *Lab Invest*. 1994;71(4):583–594.
22. Southgate J, Trejdosiewicz LK, Smith B, Selby PJ. Patterns of splice variant CD44 expression by normal human urothelium in situ and in vitro and by bladder-carcinoma cell lines. *Int J Cancer*. 1995;62(4):449–456.
23. Harnden P, Southgate J. Cytokeratin 14 as a marker of squamous differentiation in transitional cell carcinomas. *J Clin Pathol*. 1997;50(12):1032–1033.
24. Papafiotou G, Paraskevopoulou V, Vasilaki E, Kanaki Z, Paschalidis N, Klinakis A. KRT14 marks a subpopulation of bladder basal cells with pivotal role in regeneration and tumorigenesis. *Nat Commun*. 2016;7:11914.
25. Southgate J, Harnden P, Trejdosiewicz LK. Cytokeratin expression patterns in normal and malignant urothelium: A review of the biological and diagnostic implications. *Histol Histopathol*. 1999;14(2):657–664.
26. Wu XR, Kong XP, Pellicer A, Kreibich G, Sun TT. Uroplakins in urothelial biology, function, and disease. *Kidney Int*. 2009;75(11):1153–1165.
27. Ho PL, Lay EJ, Jian W, Parra D, Chan KS. Stat3 activation in urothelial stem cells leads to direct progression to invasive bladder cancer. *Cancer Res*. 2012;72(13):3135–3142.
28. Niwa H, Burdon T, Chambers I, Smith A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev*. 1998;12(13):2048–2060.
29. Choi W, Czerniak B, Ochoa A, et al. Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol*. 2014;11(7):400–410.
30. Dadhania V, Zhang M, Zhang L, et al. Meta-analysis of the luminal and basal subtypes of bladder cancer and the identification of signature immunohistochemical markers for clinical use. *EBioMedicine*. 2016;12:105–117.
31. Lamm D, Persad R, Brausi M, et al. Defining progression in nonmuscle invasive bladder cancer: It is time for a new, standard definition. *J Urol*. 2014;191(1):20–27.
32. Rakoff-Nahoum S, Medzhitov R. Toll-like receptors and cancer. *Nat Rev Cancer*. 2009;9(1):57–63.
33. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: A leading role for STAT3. *Nat Rev Cancer*. 2009;9(11):798–809.
34. Allison JP. Immune checkpoint blockade in cancer therapy: The 2015 Lasker-DeBakey Clinical Medical Research Award. *JAMA*. 2015;314(11):1113–1114.
35. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. *J Clin Oncol*. 2015;33(17):1974–1982.
36. Mahoney KM, Rennert PD, Freeman GJ. Combination cancer immunotherapy and new immunomodulatory targets. *Nat Rev Drug Discov*. 2015;14(8):561–584.
37. Varley CL, Stahlschmidt J, Lee WC, et al. Role of PPARgamma and EGFR signalling in the urothelial terminal differentiation programme. *J Cell Sci*. 2004;117(pt 10):2029–2036.
38. Varley CL, Southgate J. Effects of PPAR agonists on proliferation and differentiation in human urothelium. *Exp Toxicol Pathol*. 2008;60(6):435–441.
39. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
40. Chan KS, Espinosa I, Chao M, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc Natl Acad Sci U S A*. 2009;106(33):14016–14021.
41. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929–944.
42. Ng SW, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433–437.
43. Smith BA, Sokolov A, Uzunangelov V, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci U S A*. 2015;112(47):E6544–E6552.
44. Neve RM, Chin K, Fridlyand J, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 2006;10(6):515–527.
45. Daemen A, Griffith OL, Heiser LM, et al. Modeling precision treatment of breast cancer. *Genome Biol*. 2013;14(10):R110.
46. Garay JP, Gray JW. Omics and therapy - a basis for precision medicine. *Mol Oncol*. 2012;6(2):128–139.
47. Kurtova AV, Xiao J, Mo Q, et al. Blocking PGE2-induced tumour repopulation abrogates bladder cancer chemoresistance. *Nature*. 2015;517(7533):209–213.
48. Chan KS. Molecular Pathways: Targeting cancer stem cells awakened by chemotherapy to abrogate tumor repopulation. *Clin Cancer Res*. 2016;22(4):802–806.