Title

A cell of origin gene signature indicates human bladder cancer has distinct cellular progenitors.

Journal

Stem Cells, 32(4)

Authors

Dancik, Garrett

Owens, Charles

Iczkowski, Kenneth

et al.

Publication Date

2014-04-01

DOI

10.1002/stem.1625

Peer reviewed

Published in final edited form as:
Stem Cells. 2014 April ; 32(4): 974–982. doi:10.1002/stem.1625.

A Cell Of Origin gene signature indicates human bladder cancer has distinct cellular progenitors

Garrett M. Dancik[a,b,^], **Charles R. Owens**[a], **Kenneth A. Iczkowski**[c], and **Dan Theodorescu**[a,b,c,*]

[a]Department of Surgery, University of Colorado, Aurora, Colorado, 80045 USA

[b]Department of Pharmacology, University of Colorado, Aurora, Colorado, 80045 USA

[c]Department of Pathology, University of Colorado, Aurora, Colorado, 80045 USA

[d]University of Colorado Comprehensive Cancer Center, Aurora, Colorado, 80045 USA

Abstract

There are two distinct forms of urothelial (bladder) cancer: muscle-invasive (MI) and non-muscle invasive (NMI) disease. Since it is currently believed that bladder cancer arises by transformation of urothelial cells of the basal layer, bladder cancer stem cells (CSCs) have been isolated based on expression markers found in such cells. However, these CSCs have only been identified in MI tumors raising the intriguing hypothesis that NMI tumor progenitors do not arise from the basal compartment. To test this hypothesis, we carried out genomewide expression profiling of laser capture microdissected basal and umbrella cells, the two most histologically distinct cell types in normal urothelium and developed a Cell Of Origin (COO) gene signature that distinguishes these. The COO signature was a better predictor of stage and survival than other bladder, generic or breast CSC signatures and bladder cell differentiation markers in multiple patient cohorts. To assess whether NMI and MI tumors arise from a distinct (DPC) or common (CPC) progenitor cell, we developed a novel statistical framework that predicts COO score as a function of known genetic alterations (*TP53, HRAS, KDM6A*and *FGFR3*) that drive either MI or NMI bladder cancer and compared this to the observed COO score of the tumor. Analysis of 874 patients in 5 cohorts established the DPC model as the best fit to the available data. This observation supports distinct progenitor cells in NMI and MI tumors and provides a paradigm shift in our understanding of bladder cancer biology that has significant diagnostic and therapeutic implications.

Keywords

Bladder cancer; Cancer stem cells; Urothelial tumorigenesis; DNA microarray

[*]**Corresponding author:** Dan Theodorescu University of Colorado Comprehensive Cancer Center, Aurora, CO 80045, Tel: 303-724-7135, Fax: 303-724-3162, dan.theodorescu@ucdenver.edu.
[^]Current address: Mathematics and Computer Science Department, Eastern Connecticut State University, Willimantic, Connecticut, 06226 USA

Author Contributions

G.M.D, C.R.O. and K.A.I performed the experimental work, data analysis and wrote the manuscript.
D.T. conceived the initial idea and wrote the manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest regarding subject matter of this manuscript

INTRODUCTION

Bladder cancer is the 6th most common cancer in the United States and the 4th most common cancer in males [1]. There are two major types. Approximately 70% of bladder cancer patients present with non-muscle invasive (NMI) tumors, which recur frequently and are usually not fatal but make bladder cancer one of the most expensive cancers to treat [2]. The remaining 30% of patients present with muscle-invasive (MI) tumors. These individuals have a 5 year survival rate of approximately 50% [1] due to the lethal metastasis that can occur. NMI and MI tumors have distinct mutational genomic landscapes, with activating *RAS* mutations common in NMI tumors and *TP53/RB1* impairment common in MI tumors [3].

It is currently believed that all bladder cancer is derived from a common cancer stem cell (CSC) that arises from transformation of a normal urothelial progenitor. A CSC is a cell with the ability to self-renew and reconstitute the heterogeneity of the tumor [4]. CSCs have been isolated from many tumor types including leukemias [5], melanomas [6] and epithelial tumors [7]. Identification and therapeutic targeting of CSCs is clinically important since CSCs may drive recurrence and metastasis [8]. In bladder cancer, the genesis of the CSC is postulated to occur from the normal urothelial basal cell layer that includes long-lived and proliferative cells that are thought to give rise to an intermediate layer of “transient amplifying cells” [9] and in turn a superficial umbrella (luminal) layer of fully differentiated cells. Bladder CSCs have hence been isolated on the basis of expression of markers such as CD44 found on the surface of normal basal urothelial cells.

However, this model of bladder cancer has two major conceptual limitations. The first is the assumption that basal cells give rise to umbrella cells. This notion is not supported by recent evidence showing that normal umbrella cells can develop independently of the basal cell layer [10]. The second is that to date, bladder CSCs have only been identified in MI tumors [11–16]. Together this raises the intriguing hypothesis that not all bladder cancer derives from a common progenitor, or more specifically, that NMI tumors have a non-basal cell of origin.

Given the distinct origin of basal and umbrella cells [10], here we reason that a gene expression signature that distinguishes basal from umbrella cells in normal urothelium can be used in combination with mutation signatures for genes associated with the two types of bladder cancer to determine if human bladder tumors arise from a common progenitor or not. First we compared transcriptional profiles of normal basal and umbrella cells to derive a Cell Of Origin (COO) gene signature. We then evaluated the ability of the COO to predict stage and outcome in bladder cancer patients and compared this predictive performance to that of other CSC signatures and known bladder differentiation markers. A novel statistical framework was then developed using the COO signature to determine if tumors arise from distinct or common progenitors, concluding that MI and NMI bladder cancers arise from distinct progenitors.

MATERIALS AND METHODS

Gene expression profiling and Quantitative RT-PCR of basal and umbrella samples

The Applied Biosystems® ArcturusXT™ LCM System (Life Technologies, Carlsbad, CA, USA) was used to isolate basal cells and umbrella cells from frozen nonmalignant human urothelial tissue specimens. Total RNA was harvested from the laser capture micro-dissected basal and umbrella cells using the ARCTURUS® PicoPure® RNA Isolation Kit (Life Technologies, Carlsbad, CA, USA) and the mRNA was converted to cDNA using the IScript cDNA Synthesis kit (Bio-Rad Laboratories, Hercules, CA, USA). Gene expression

analysis was carried out using the Affymetrix U219 Array Plate platform. Expression levels of mRNA of OSR2, RHOJ, and SLFN11 were determined by quantitative RT-PCR using the SYBR Green SuperMix protocol on an iQ5 Cycler (Bio-Rad Laboratories, Hercules, CA, USA) and QuantiTect® primer sets QT00044793, QT00092078, and QT01028671 respectively (Qiagen Inc., Valencia, CA, USA). Levels of GAPDH mRNA were determined in parallel and used to normalize the relative mRNA expression levels shown, using specific quantitative RT-PCR primers: forward 5'-TCTTTTGCGTCGCCAGCCGA-3' and reverse 5'-ACCAGGCGCCCAATACGACC-3'. For the calculation of gene expression, the ΔΔCT method was used. Expression was normalized to the umbrella cells to calculate gene expression in the basal and umbrella cells. The investigator was blinded to other results.

Gene expression datasets, sample processing, signatures, and signature scores

All gene expression datasets used in the study are summarized in Supporting Information Table S1. The processed publicly available data was used for all cohorts. Missing values in the Blaveri cohort were imputed using the *impute* package (*impute.knn* function in *R*) with default parameters [17]. Replicate samples in Dyrskjot cohort were averaged to produce a single expression profile for each patient.

The Cell Of Origin (COO) signature includes all genes differentially expressed between basal and umbrella samples at a false discovery rate of 10%, using the significance analysis of microarrays (*samr*) package [18]. The T-IC signature consists of 50 genes differentially expressed between CD44+ and CD44- tumor cell pairs sorted from three patients [11], at an FDR of 10%, using *samr*. The HTC signature consists of 1079 genes differentially expressed between 67LR+ and 67 LR- xenografted SW780 cells [12]. The *KDM6A, HRAS*and *TP53* mutation signatures consist of the differentially expressed genes between mutant and wild-type patients for each respective gene at a FDR of 10% (Supporting Materials and Methods). The *FGFR3* signature is simply *FGFR3* mRNA expression. Other signatures include the embryonic stem cell signature ESC, consisting of 335 genes commonly expressed in mouse and human embryonic stem cells [19]; the ES exp1 signature, consisting of 380 genes overexpressed in human embryonic stem cells [20], an induced pluripotent (iPS) stem cell signature, consisting of 90 genes differentially expressed between iPS stem cells and human fibroblasts [21]; the invasiveness gene signature (IGS), consisting of 174 genes differentially expressed between CD44+CD24-/low tumorigenic breast CSCs and normal breast cells [22]; and the consensus stemness ranking (CSR) signature, consisting of genes differentially expressed between those having consistently high and low ESC, ES exp1, IGS and iPS score rankings [23].

Signature gene symbols were updated according to www.genenames.org. For genes with multiple probes in a dataset, the probe with the highest mean expression value was used [24]. For each gene, expression values are *z*-normalized across samples to have a mean expression of zero and a standard deviation of one. We defined an upregulated (downregulated) gene as a signature gene having higher (lower) expression in basal (compared to umbrella) cells, CD44+ (compared to CD44- cells), 67LR bright (compared to 67LR dim) cells, CD44+CD24-/low tumorigenic breast cells (compared to normal breast cells) and high CSR (compared to low CSR) cells, or patients with mutations in *KDM6A, TP53*, or *HRAS* (compared to wild-type patients) in the cohort used to derive the signature. *FGFR3* is considered 'upregulated' in the *FGFR3* signature, and all ESC and ES exp1 signature genes are upregulated in stem cells. For all signatures, the signature score is equal to the sum of all normalized up-regulated genes minus the sum of all normalized down-regulated genes.

Common progenitor cell (CPC) and distinct progenitor cell (DPC) models

Simple linear regression models are used to model Cell Of Origin (COO) signature score as a function of *KDM6A*, *HRAS*, *FGFR3*, and *TP53* signature scores, and the progenitor cell as follows:

$$s_i = \mu + \gamma x_{stage,i} + \alpha_{KBM6A} x_{KDM6A,i} + \alpha_{HRAS} x_{HRAS,i} + \alpha_{FGFR3} x_{FGFR3,i} + \alpha_{TP53,i} + \varepsilon_i$$

where s_i is the COO score for patient i, $i = 1, \ldots n$; μ is the intercept; γ is the increase (decrease) in COO signature score associated with a tumor being derived from an MI (NMI) progenitor cell, and α_{KDM6A}, α_{HRAS}, α_{FGFR3}, and α_{TP53} are the increase in COO score associated with high *KDM6A*, *HRAS*, *FGFR3*, and *TP53* signature scores (defined below), respectively. Note that under the CPC model $\gamma = 0$ while under the DPC model $\gamma \neq 0$. The variable $x_{stage,i}$ is set equal to +1 if patient i has an MI tumor and −1 otherwise; $x_{g,i}$ is an indicator and is 1 if patient i has a high signature score for gene g, $g \in \{$*KDM6A*, *HRAS*, *FGFR3*, *TP53*$\}$, or 0 otherwise. Measurement error is represented by ε_i and these are independently and normally distributed with mean 0. For each signature score, StepMiner was used to calculate a threshold separating high and low signature scores, when scores were ordered from lowest to highest [25]. StepMiner identifies the threshold that maximizes the F-statistic testing its significance [26]. For each cohort, the statistical model is fit to the data and an F-test is used to test the hypothesis that $\gamma = 0$, i.e., that the CPC model fits the patient COO scores significantly better than the DPC model.

Additional statistical methods

Gene set enrichment analysis was carried out using Database for Visualization and Annotated Discovery (DAVID, Ref #[27]) to identify overrepresented Gene Ontology (GO) terms associated with COO signature genes. Default DAVID parameters were used for the analysis and GO terms having a false discovery rate (FDR) <20% identified. Enrichment was with respect to the Homo sapiens whole genome background. The predictive ability of signature scores and differentiation markers to distinguish between NMI and MI tumors was assessed by the area under the receiver operating characteristics curve (AUC), with AUC > 0.50 indicating the signature or marker score is higher in MI tumors than NMI tumors, unless specified otherwise. Statistical significance was assessed by the Wilcoxon rank-sum test that calculates a p-value comparing AUC to 0.5 (i.e., what would be expected by chance). Survival analysis was carried out using Cox Proportional Hazards regression models. With the exception of DAVID, all statistical analyses were conducted using the *R* programming language [28].

RESULTS

Derivation and characteristics of a Cell Of Origin (COO) gene signature from normal basal and umbrella urothelial cells

Eight samples of normal urothelium were used to harvest basal and umbrella cells using laser capture microdissection (LCM) (Fig. 1). These were pooled in 2 groups and RNA extracted followed by gene expression profiling using the Affymetrix U219 platform. We identified a Cell Of Origin (COO) gene signature consisting of 186 genes (224 probes) differentially expressed between basal and umbrella cells (Supporting Information Table S2, FDR < 0.10). We selected the top 3 genes for internal validation by qPCR. Top genes were defined as genes with multiple significant probes having differential expression (FDR < 0.10) and with one probe having an FDR of 0. Internal validation of these genes (RHOJ, SLFN11, OSR2) by qPCR confirmed that these signature genes have higher expression in basal than umbrella cells (Supporting Information Fig. S1A). Notably, all genes were higher

in basal compared to umbrella cells. Gene Ontology (GO) analysis found the COO signature enriched in genes relating to cell proliferation (FDR < 0.001), consistent with the finding that umbrella cells are quiescent whereas a subset of basal cells divide [29]. The COO signature is also enriched in cell-cell signaling, skeletal system development and cell activation (FDR < 20%, Supporting Information Table S3). A COO "score" was developed consisting of the average expression of the 186 signature genes (see **Materials and Methods**) upregulated in basal cells, so that samples with high COO scores are more "basal-like" (or have high "basalness"), and samples with low COO scores are more "umbrella like".

The COO gene signature is associated with "stemness" indicating that in muscle invasive bladder cancer CSCs derive from normal basal urothelial progenitors

Because basal cell markers have been used to isolate bladder CSCs, we evaluated the relationship between COO score and bladder CSCs by evaluating the ability of the COO and published bladder CSC gene signatures or those we derived from published data (as described below) to discriminate between basal and umbrella cells or between tumorigenic (defined as bladder CSCs) and non-tumorigenic bladder cancer cells isolated from patients with MI tumors. Berman et al. sorted SW170 xenograft cells on the basis of the basal marker 67LR, and found that 67LR bright cells ("Highly Tumorigenic Cells"; HTC) were more tumorigenic *in vivo* compared to 67LR dim cells. They identified a 1079 gene signature (HTC signature) that distinguished 67LR bright from dim cells [12]. Chan et al. sorted patient tumors on the basis of the basal marker CD44, and found that $CD44^+$ cells ("Tumor-Initiating Cells"; T-IC) were more tumorigenic *in vivo* compared to $CD44^-$ cells [11]. Using the gene expression profiles of $CD44^+$ and $CD44^-$ cells graciously provided to us by the authors, we developed a 50 gene signature (T-IC signature) that distinguished $CD44^+$ from $CD44^-$ cells. For both HTC and T-IC signatures, high scores (see Methods) were associated with a bladder CSC phenotype or "stemness" (i.e., 67LR bright or $CD44^+$).

The COO, HTC, and T-IC signatures were then compared for their association with basalness and stemness in Fig. 2. All signature scores differentiate between the two cell types in the datasets they were derived from (basal vs. umbrella for COO; bright vs. dim for HTC; CD44 positive vs. negative for T-IC), confirming that high signature scores reflect basalness (COO) or stemness (HTC and T-IC). The COO score is higher in bladder CSCs (67LR bright, $CD44^+$) than in non-tumorigenic cells (67LR dim, $CD44^-$) (Fig. 2A, $P < 0.05$). However, the HTC and T-IC signature scores are not significantly higher in basal cells than in umbrella cells, suggesting that these signatures report on signals that do not drive normal urothelial differentiation. HTC scores are not significantly different between $CD44^+$ and $CD44^-$ cells ($P = 0.40$, Fig. 2B), while the T-IC score is not significantly different between 67LR bright and 67LR dim cells ($P = 0.50$, Fig. 2C), suggesting that 67LR and $CD44^+$ cells represent distinct subsets of bladder CSCs isolated from MI tumors. In contrast, the COO analysis is different in these groups (Fig. 2A) suggesting that these bladder CSC isolation procedures identified CSCs from MI tumors that are derived from normal basal urothelial cells supporting previous findings [11, 12]. However as stated previously, bladder CSCs from NMI tumors have not been identified.

COO signature has superior performance in predicting tumor stage and patient outcome

Gene signatures derived from $CD44^+$ and 67LR bright bladder CSCs can distinguish NMI from MI tumors and are associated with bladder cancer patient outcomes [11, 12]. To assess the clinical relevance of COO score we evaluated its predictive ability for stage (NMI vs. MI) and outcome in bladder cancer and compared its ability to that of known urothelial differentiation markers and additional stem cell signatures. In addition to the HTC and T-IC bladder CSC signatures described above, we also considered the basal cell marker KRT14,

which associates with poor prognosis in bladder cancer [25] and the basal marker p63 which is required for normal basal urothelium development [30] but has high expression in NMI tumors [31]. We also considered three 'generic' embryonic stem cell signatures and two breast cancer stem cell signatures. The ESC signature consists of 335 genes commonly expressed in mouse and human embryonic stem cells [19], the ES exp1 signature consists of 380 genes overexpressed in human embryonic stem cells [20], and the induced pluripotent (iPS) stem cell signature consists of 90 genes differentially expressed between iPS stem cells and human fibroblasts [21]. The invasiveness gene signature (IGS) consists of genes differentially expressed between $CD44^{+}CD24^{-}$/low tumorigenic breast CSCs and normal breast cells [22] and the consensus stemness ranking (CSR) signature includes genes differentially expressed between those having consistently high and low ranks in ESC, ES exp1, IGS and iPS scores [23]. High signature scores reflect basalness (for bladder signatures) or stemness (for generic and breast signatures).

The bladder-specific COO, HTC, and T-IC scores were the only scores significantly predictive of stage in all five bladder cancer cohorts (N = 874, Supporting Information Table S1) we examined (Fig. 3A, B, $P < 0.05$), with high scores associated with MI tumors, as expected. The COO score was the best predictor of stage in 4 of the 5 cohorts (AUC range 0.72 - 0.88, $P < 0.05$), and maintained its predictive ability when controlling for grade (Supporting Information Fig. S1B). Therefore, although the COO signature is enriched in genes associated with proliferation, which is known to associate with tumor grade [32], the ability of COO score to distinguish between NMI and MI tumors is not a function of its ability to distinguish between tumors of low and high grades. The CSR signature was predictive of stage in 4 cohorts, with MI tumors having higher scores (AUC range 0.63 – 0.80, $P < 0.05$). Although we expected high stem cell scores to be associated with MI tumors, high ESC scores were significantly associated with NMI tumors in one cohort (MSKCC, AUC = 0.37, $P < 0.05$), as were high IGS scores (Dyrskjot, AUC = 0.27, $P < 0.001$). Interestingly, high iPS scores were associated with NMI tumors in all cohorts, with a statistically significant association in three (AUC range = 0.22–0.34, $P < 0.05$). When the signatures were categorized according to tissue type (bladder, generic, and breast), the bladder signatures had higher average AUC values in 4 out of 5 cohorts (Supporting Information Fig. S1C, $P < 0.05$ by one-tailed binomial test).

The COO score was the best predictor of disease-specific survival in the 3 datasets with this endpoint and was predictive in all 3 cohorts (Fig. 3C,D, $P < 0.05$). ES exp1 and CSR scores were also predictive in 3 cohorts but with lower hazard ratios than those obtained from COO scores. T-IC scores were predictive in two datasets, while KRT14 (probed in 2 cohorts) and TP63 (probed in 1 cohort) expression were not predictive in any. These results demonstrate that COO score is the best predictor of stage and outcome in bladder cancer patients, and that there are tissue-specific aspects of the COO signature that drive its predictive value which cannot be captured by the "generic" embryonic and breast CSC signatures we considered. Furthermore, it is clear that MI tumors have higher COO scores than NMI tumors.

Evaluation of common and distinct progenitor cell models predicting patient COO scores indicates that bladder cancers arise from distinct normal cellular progenitors

Mutations in MI or NMI tumors in bladder cancer [33–35] suggest these tumors arise through divergent genetic pathways but whether this occurs from a common or distinct normal progenitor cells remains to be established. *TP53* mutations occur in >50% of high grade MI tumors while *HRAS* and *FGFR3* mutations occur in ~30% and 70% of low grade NMI tumors, respectively, while *KDM6A* is more frequently mutated in NMI tumors [36]. The observation that basal-like bladder CSC signatures are associated with stage [11, 12]

and COO scores are higher in MI than in NMI tumors (Fig 3A) suggest that NMI tumors have a non-basal progenitor cell. To address this issue, we considered two possibilities that would explain these findings: 1) NMI and MI tumors arise through a common progenitor cell (CPC) and oncogenic alterations determine invasiveness (i.e., stage) of the tumor, and 2) NMI and MI tumors have distinct progenitors (DPC), umbrella and basal cells respectively, with different inherent levels of basalness determining invasiveness.

We evaluated these two possibilities using a novel statistical framework that models COO score as a function of tumor stage and genomic alterations in NMI or MI disease. This framework and model evaluation process is illustrated in Fig. 4. According to the CPC model, a tumors' COO score is driven by the genetic pathways involving *KDM6A*, *HRAS*, *FGFR3*, and *TP53* whose activity is quantified by gene expression signature scores as described below (Fig. 4B). In contrast, in the DPC model, the tumors' COO scores are also dependent on the distinct (MI or NMI) progenitor cell type, with the mutation signature scores not able to adequately explain the tumors' COO scores. The difference between the CPC and DPC models is encapsulated mathematically by the single parameter γ which is proportional to the difference in the expected COO scores between NMI and MI progenitor cells; $\gamma = 0$ under the CPC model and $\gamma \neq 0$ under the DPC model.

For both models, *KDM6A*, *HRAS*, and *TP53* mutation status is estimated using mutation signature scores derived from patient tumors (wild-type vs. mutant) that are predictive of *KDM6A* and *TP53* mutation status and of *HRAS* overexpression in manipulated cell lines, as well as of stage and outcome in multiple cohorts (Supporting Results and Discussion). For each cohort and each signature score, the StepMiner algorithm [26] was used to identify a threshold separating high and low scores which define a mutated and wild-type state, respectively. The *FGFR3* signature is simply *FGFR3* mRNA expression according to the probe with the highest mean expression across samples. *FGFR3* expression is positively correlated with *FGFR3* mutation status and is high in NMI tumors [37]. We note that although *KDM6A*, *HRAS*, and *FGFR3* mutation is known to be associated with NMI tumors and *TP53* mutation with that of MI tumors (and are illustrated this way in Fig. 4B), the CPC and DPC models make no assumptions about the relationship between *KDM6A*, *HRAS*, *FGFR3*, and *TP53* mutation status signatures scores and invasiveness.

For each patient, the COO, *KDM6A*, *HRAS*, *FGFR3*, and *TP53* mutation status signature scores are calculated for each patient (Fig. 4A), input into the statistical model along with tumor stage (Fig. 4B) and model parameters are estimated using ordinary least squares (Fig. 4C) We fit the CPC and DPC models (Fig. 4C and **Methods**) to five bladder cancer cohorts using gene expression profiles from 874 bladder cancer patients (Table 1) and evaluated whether each DPC model more accurately explained actual patient COO scores than the corresponding CPC model (Fig. 4D). In the CPC models, high *TP53* scores are associated with high basal scores ($\alpha_{TP53} > 0$), while in general low *FGFR3*, *HRAS*, and *KDM6A* scores are associated with high basal scores (α_{HRAS}, α_{FGFR3}, α_{KDM6A}, < 0) and the coefficients of all signature scores except for *HRAS* were significant in multiple cohorts. In the DPC models, the MI progenitor has a higher level of basalness than the NMI progenitor ($\gamma > 0$, $P < 0.05$) and the DPC model fits the data significantly better than the CPC model does ($P \leq 0.01$). Results were similar when continuous *KDM6A*, *HRAS*, *FGFR3*, and *TP53* signature scores were used in the analysis rather than selecting a threshold to identify low and high-scoring samples (Supporting Information Table S5), although the DPC model was no longer a significantly better fit than the CPC model at $P < 0.05$ in one cohort (Dyrskjot, $P = 0.06$). Interestingly, low *KDM6A* scores are associated with high COO scores in all cohorts ($P < 0.05$) in the DPC model, and low *FGFR3* scores are associated with high COO scores in multiple cohorts ($P < 0.05$). These results strongly suggest that MI and NMI tumors have

distinct progenitor cells because the COO scores in NMI and MI tumors are not adequately explained by genetic alterations represented by mutation scores.

DISCUSSION

Observational studies of bladder cancer patients and experimental studies in mouse models indicate divergent genetic pathways are associated with NMI and MI tumors [3]. However, it remains unknown whether NMI and MI tumors have a common or distinct progenitor cell type. Attempts to isolate bladder CSCs based on basal cell markers presume there is a single bladder CSC. However, although CD44 is a basal cell and bladder CSC marker, in a tissue array of >300 specimens roughly 60% stained negative for the basal cell marker CD44 [11]. Furthermore, $CD44^-$ primary patient tumors formed xenografts in mice, suggesting that CSCs in $CD44^-$ tumors are not from the basal layer [38]. In addition, attempts to isolate bladder CSCs from NMI tumors based on basal cell markers have been unsuccessful and the finding that umbrella-like cells can develop independently of basal progenitor cells [10] suggest that distinct progenitor cell types exist.

Here, we present a novel statistical model to address the fundamental question of whether genetic alterations in common or distinct progenitors lead to the two types of human bladder cancer. We answer this question by developing a Cell Of Origin (COO) signature from normal urothelium cells that is reflective of the two putative progenitor cell types (basal and umbrella) and find that MI tumors have higher COO scores (i.e., are more basal like) than NMI tumors. To untangle whether this difference is due to genetic alterations or distinct progenitor cell types, we develop gene mutation signatures for genes associated with stage and combine these with the COO signature to evaluate whether or not NMI and MI tumors have a common (CPC) or distinct (DPC) progenitor. Strikingly, we found the DPC model was a better fit than the CPC model in 5 cohorts, suggesting that differences in COO scores between patients is not adequately explained by genetic alterations alone. Interestingly, the association of low *KDM6A* and *FGFR3* scores with high COO scores indicates that alterations in these genes could promote differentiation. However, the relationship between *KDM6A* and *FGFR3* and urothelial differentiation is unknown and this result could simply indicate that stage and *FGFR3*, *KDM6A*, and COO scores are highly correlated. On the other hand, *KDM6A* loss inhibits endoderm differentiation in human embryonic stem cells [39], while *FGFR3* inhibition induces differentiation in human myeloma cell lines [40] making their involvement in bladder cancer differentiation a possibility.

Our current CPC/DPC statistical model has limitations. First, we do not have a mutation signature for *RB1*, which is more frequently mutated in MI tumors [41]. However, because *RB1* is the only known driver associated with invasiveness not included in our model and because *RB1* and *TP53* mutations frequently co-occur [42], we expect that inclusion of *RB1* would not significantly change our results. Second, it is possible that the tumor microenvironment could influence differentiation (as measured by COO score) and this is not modeled. In bladder cancer, 67LR is a basal cell marker that colocalizes with the basal marker C17 in xenografts but not in culture, suggesting that tumor stromal cells influence differentiation [12]. While alterations in the microenvironment that independently influence invasiveness would be supportive of the CPC model, it is currently not known whether such influences affect differentiation independently of the genes in our model.

The identification of bladder CSCs has important diagnostic and therapeutic implications. Our results suggest that bladder CSCs in NMI tumors arise from non-basal cells, although further studies will be necessary to determine whether the cell of origin in these tumors reside in the umbrella or intermediate layers. Identification and characterization of such NMI CSCs using genes in our signature could lead to clinically significant improvements in

the prognostication and treatment of patients with NMI tumors, which are notorious for their high rates of recurrence [43]. Our results also suggest that the $CD44^+$ [11] and 67LR bright [12] bladder CSCs that have previously been identified from MI tumors are distinct CSC subsets, both derived from basal progenitor cells. In conclusion, this work supports a new conceptual framework for investigating and understanding bladder cancer, one in which two rather than one distinct progenitor cell type drives tumor biology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Keith S. Chan and Irving Weissman for providing the raw gene expression data for $CD44^+$ (T-IC) and $CD44^-$ samples. This work was supported in part by National Institutes of Health grants CA075115 and CA104106. The authors wish to thank Nicole Spoelstra in the Department of Pathology at the University of Colorado for assistance with laser capture microdissection.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. 2013; 63:11–30. [PubMed: 23335087]
2. Botteman MF, Pashos CL, Redaelli A, et al. The health economics of bladder cancer: a comprehensive review of the published literature. Pharmacoeconomics. 2003; 21:1315–1330. [PubMed: 14750899]
3. Wu XR. Urothelial tumorigenesis: a tale of divergent pathways. Nat Rev Cancer. 2005; 5:713–725. [PubMed: 16110317]
4. Clarke MF, Dick JE, Dirks PB, et al. Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Research. 2006; 66:9339–9344. [PubMed: 16990346]
5. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature Medicine. 1997; 3:730–737.
6. Quintana E, Shackleton M, Sabel MS, et al. Efficient tumour formation by single human melanoma cells. Nature. 2008; 456:593-U533. [PubMed: 19052619]
7. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:3983–3988. [PubMed: 12629218]
8. Li XX, Lewis MT, Huang J, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008; 100:672–679. [PubMed: 18445819]
9. Farsund T. Cell-Kinetics of Mouse Urinary-Bladder Epithelium .2. Changes in Proliferation and Nuclear-DNA Content during Necrosis Regeneration, and Hyperplasia Caused by a Single Dose of Cyclophosphamide. Virchows Archiv B-Cell Pathology Including Molecular Pathology. 1976; 21:279–298.
10. Signoretti S, Pires MM, Lindauer M, et al. p63 regulates commitment to the prostate cell lineage. Proc Natl Acad Sci U S A. 2005; 102:11355–11360. [PubMed: 16051706]
11. Chan KS, Espinosa I, Chao M, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:14016–14021. [PubMed: 19666525]
12. He XB, Marchionni L, Hansel DE, et al. Differentiation of a Highly Tumorigenic Basal Cell Compartment in Urothelial Carcinoma. Stem Cells. 2009; 27:1487–1495. [PubMed: 19544456]
13. Ning ZF, Huang YJ, Lin TX, et al. Subpopulations of Stem-like Cells in Side Population Cells from the Human Bladder Transitional Cell Cancer Cell Line T24. Journal of International Medical Research. 2009; 37:621–630. [PubMed: 19589244]

14. She JJ, Zhang PG, Wang ZM, et al. Identification of side population cells from bladder cancer cells by dyecycle violet staining. Cancer Biology & Therapy. 2008; 7:1663–1668. [PubMed: 18787416]
15. Su Y, Qiu Q, Zhang XQ, et al. Aldehyde Dehydrogenase 1 A1-Positive Cell Population Is Enriched in Tumor-Initiating Cells and Associated with Progression of Bladder Cancer. Cancer Epidemiology Biomarkers & Prevention. 2010; 19:327–337.
16. Yang YM, Chang JW. Bladder cancer initiating cells (BCICs) are among EMA(-)CD44v6(+) subset: Novel methods for isolating undetermined cancer stem (initiating) cells. Cancer Investigation. 2008; 26:725–733. [PubMed: 18608209]
17. Hastie T, Ribshirani R, Balasubramanian N, et al. impute: Imputation for microarray data. 2010
18. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98:5116–5121. [PubMed: 11309499]
19. Wong DJ, Liu H, Ridky TW, et al. Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell. 2008; 2:333–344. [PubMed: 18397753]
20. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene 14 expression signature in poorly differentiated aggressive human tumors. Nature Genetics. 2008; 40:499–507. [PubMed: 18443585]
21. Park IH, Zhao R, West JA, et al. Reprogramming of human somatic cells to pluripotency with defined factors. Nature. 2008; 451:141-U141. [PubMed: 18157115]
22. Liu R, Wang XH, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. New England Journal of Medicine. 2007; 356:217–226. [PubMed: 17229949]
23. Shats I, Gatza ML, Chang JT, et al. Using a Stem Cell-Based Signature to Guide Therapeutic Selection in Cancer. Cancer Research. 2011; 71:1772–1780. [PubMed: 21169407]
24. Miller JA, Cai C, Langfelder P, et al. Strategies for aggregating gene expression data: the collapseRows R function. BMC Bioinformatics. 2011; 12:322. [PubMed: 21816037]
25. Volkmer JP, Sahoo D, Chin RK, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes (vol 109, pg 2078, 2012). Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:3600–3600.
26. Sahoo D, Dill DL, Tibshirani R, et al. Extracting binary signals from microarray timecourse data. Nucleic Acids Research. 2007; 35:3705–3712. [PubMed: 17517782]
27. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
28. [Accessed July 1, 2010] R: A Language and Environment for Statistical Computing. Available at: http://www.R-project.org
29. Kurzrock EA, Lieu DK, deGraffenried LA, et al. Label-retaining cells of the bladder: candidate urothelial stem cells. American Journal of Physiology-Renal Physiology. 2008; 294:F1415–F1421. [PubMed: 18367656]
30. Karni-Schmidt O, Castillo-Martin M, Shen TH, et al. Distinct expression profiles of p63 variants during urothelial development and bladder cancer progression. American Journal of Pathology. 2011; 178:1350–1360. [PubMed: 21356385]
31. Choi W, Shah JB, Tran M, et al. p63 Expression Defines a Lethal Subset of Muscle-Invasive Bladder Cancers. PLoS One. 2012; 7:453–462.
32. Lauss M, Ringner M, Hoglund M. Prediction of stage, grade, and survival in bladder cancer using genome-wide expression data: a validation study. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16:4421–4433. [PubMed: 20736328]
33. Bakkar AA, Wallerand H, Radvanyi F, et al. FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. Cancer Research. 2003; 63:8108–8112. [PubMed: 14678961]
34. Orntoft TF, Wolf H. Molecular alterations in bladder cancer. Urol Res. 1998; 26:223–233. [PubMed: 9759995]
35. van Rhijn BSW, van der Kwast TH, Vis AN, et al. FGFR3 and P53 characterize alternative genetic pathways in the pathogenesis of urothelial cell carcinoma. Cancer Research. 2004; 64:1911–1914. [PubMed: 15026322]

36. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nature Genetics. 2011; 43:875–878. [PubMed: 21822268]

37. Tomlinson DC, Baldo O, Hamden P, et al. FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. Journal of Pathology. 2007; 213:91–98. [PubMed: 17668422]

38. Chan KS, Volkmer JP, Weissman I. Cancer stem cells in bladder cancer: a revisited and evolving concept. Current Opinion in Urology. 2010; 20:393–397. [PubMed: 20657288]

39. Jiang W, Wang J, Zhang Y. Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. Cell Res. 2013; 23:122–130. [PubMed: 22907667]

40. Trudel S, Ely S, Farooqi Y, et al. Inhibition of fibroblast growth factor receptor 3 induces differentiation and apoptosis in t(4;14) myeloma. Blood. 2004; 103:3521–3528. [PubMed: 14715624]

41. Cairns P, Proctor AJ, Knowles MA. Loss of heterozygosity at the RB locus is frequent and correlates with muscle invasion in bladder carcinoma. Oncogene. 1991; 6:2305–2309. [PubMed: 1766677]

42. Cote RJ, Dunn MD, Chatterjee SJ, et al. Elevated and absent pRb expression is associated with bladder cancer progression and has cooperative effects with p53. Cancer Res. 1998; 58:1090–1094. [PubMed: 9515785]

43. Akagashi K, Tanda H, Kato S, et al. Recurrence pattern for superficial bladder cancer. Int J Urol. 2006; 13:686–691. [PubMed: 16834643]

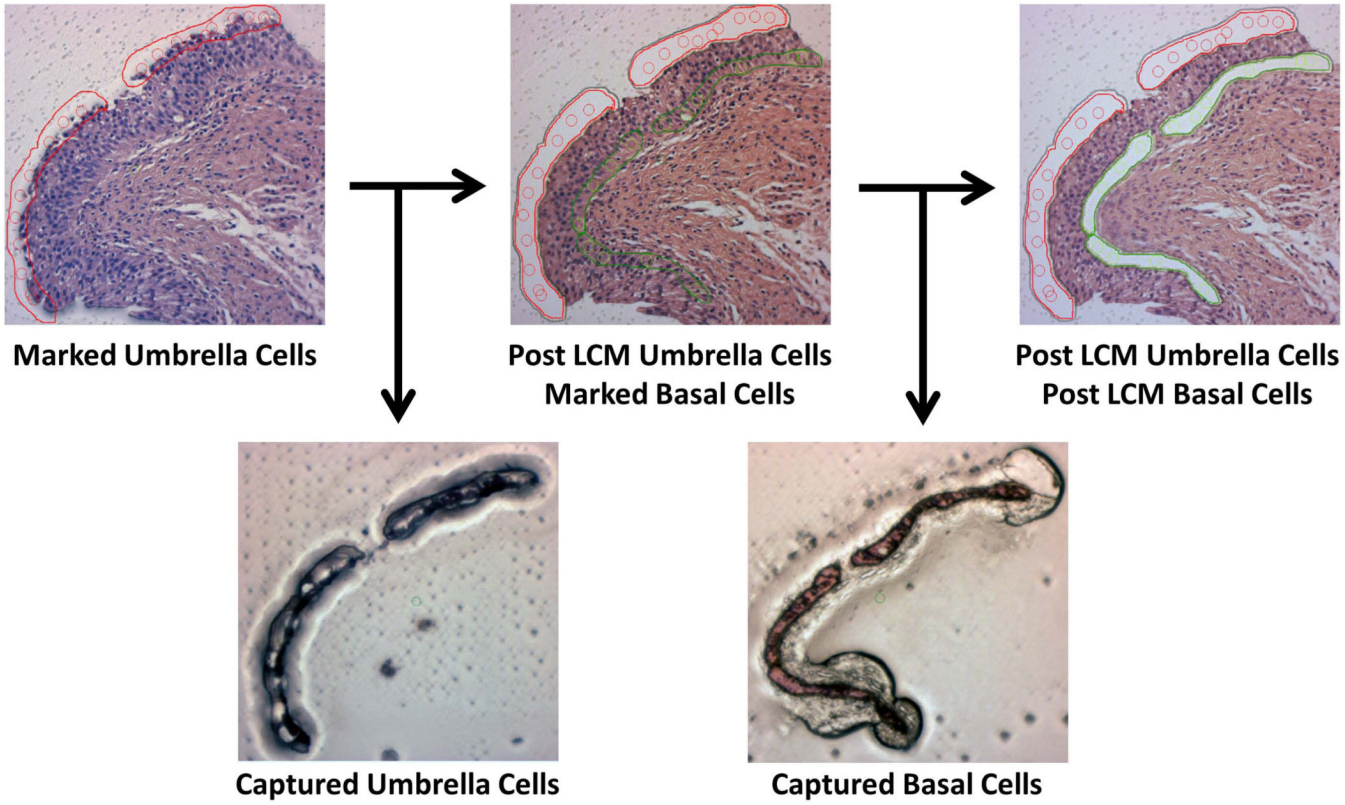

Figure 1. Laser capture microdissection (LCM) of umbrella and basal urothelial cells in normal human urothelium

Eight samples underwent LCM and were pooled in two groups (umbrella and basal) for gene expression analysis as described in **Methods**. Typical sample shown pre and post LCM of umbrella (red) and basal cell (green) layers (magnification 20X).

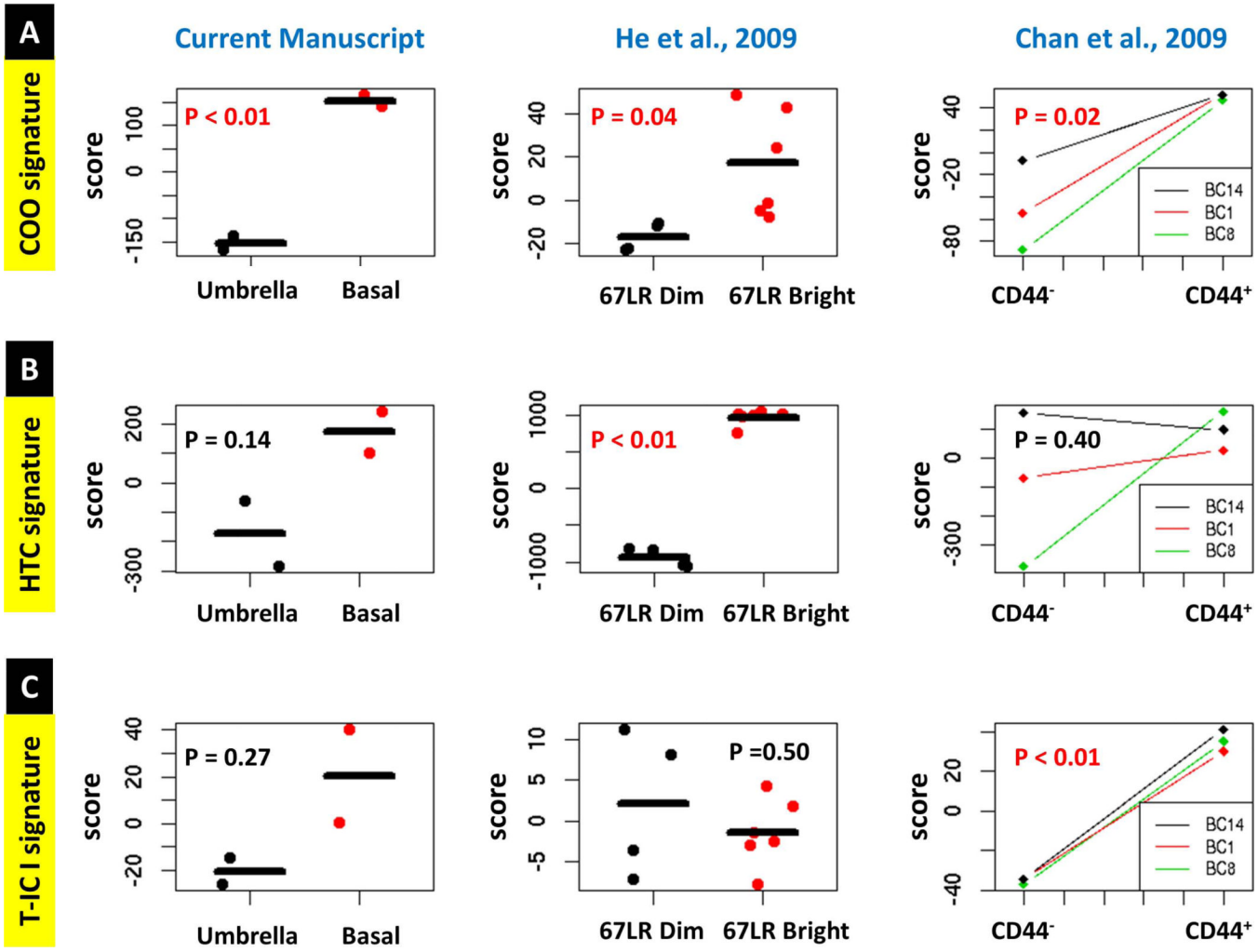

Figure 2. Ability of COO and bladder CSC signatures to discriminate between umbrella and basal cells and bladder CSCs in independent datasets

Comparison of COO **A)**, HTC **B)** and T-IC **C)** signature scores across normal umbrella and basal cells, non-tumorigenic 67LR dim and tumorigenic 67LR bright stem cells, and non-tumorigenic $CD44^{-}$ and tumorigenic $CD44^{+}$ stem cells. The COO, HTC, and T-IC signatures were derived from the current dataset, He *et al*., 2009 [12] and Chan *et al*., 2009 [11] datasets, respectively. P-values calculated by unpaired (umbrella vs. basal cells, 67LR bright vs. 67LR dim cells) and paired ($CD44^{+}$ vs. $CD44^{-}$ cells) two sample t-test.

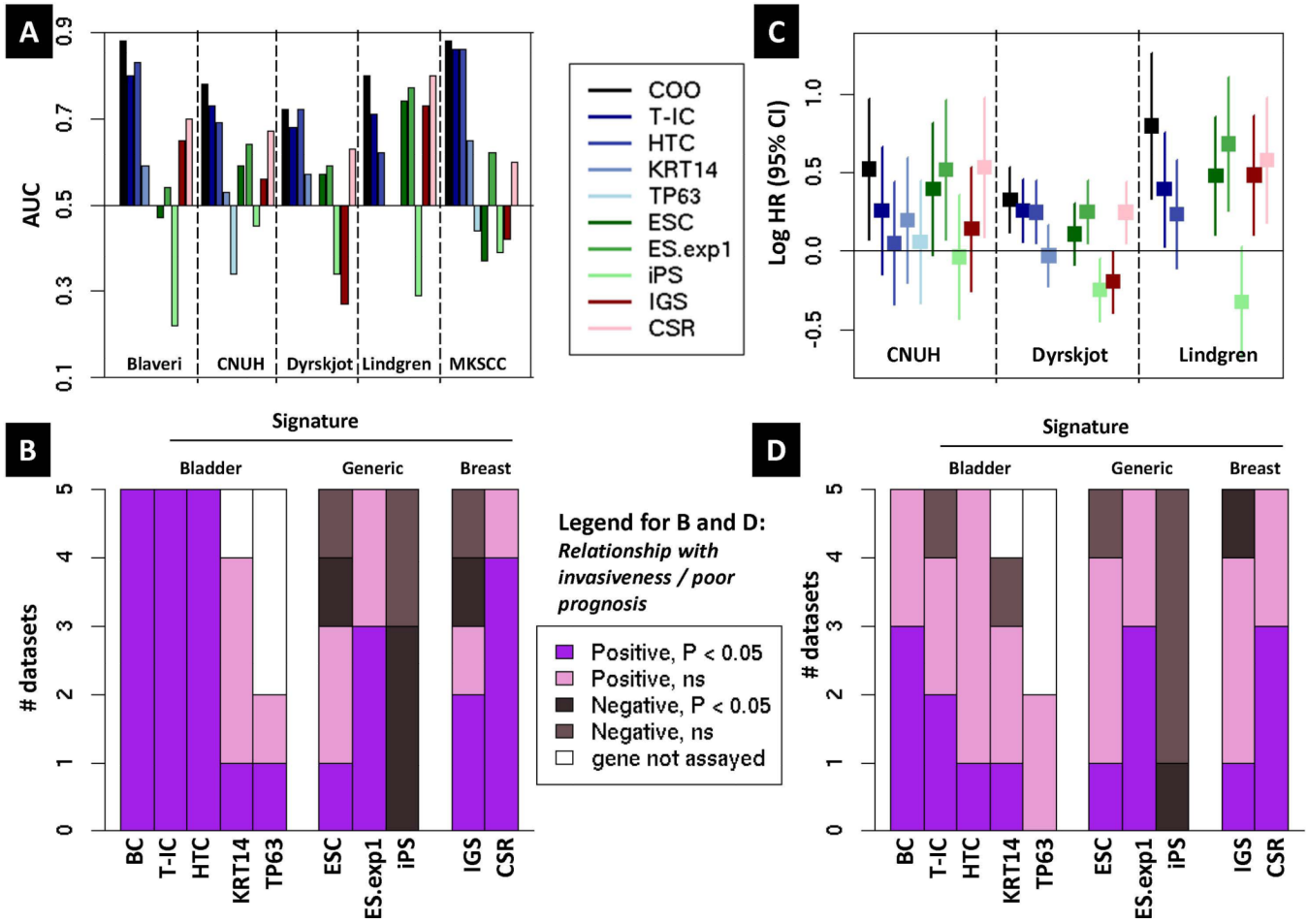

Figure 3. Predictive ability of various stem cell signatures and differentiation markers for stage and outcome in bladder cancer

A) Ability of signatures to discriminate between non-muscle invasive (NMI) and muscle invasive (MI) tumors, with the area under the receiver operating characteristic curve (AUC) corresponding to the probability of a randomly selected MI tumor having a higher signature score than a randomly selected NMI tumor. Signatures are color-coded by class: COO signature is black, bladder signatures or differentiation markers in blues, generic stem cell signatures in greens, and breast cancer stem cell signatures in reds. The following signatures are significant ($P < 0.05$) by Wilcoxon rank-sum test: COO, HTC, T-IC (all datasets); KRT14 (MSKCC); TP63 (CNUH); ES.exp1 (Lindgren and MSKCC); iPS (Blaveri, Dyrskjot, Lindgren); IGS (Blaveri, Dyrskjot, Lindgren); CSR (Blaveri, CNUH, Dyrskjot, Lindgren). Dashed line corresponds to AUC = 0.50 (i.e., the AUC expected by chance). **B)** Frequency of statistically significant signatures ($P < 0.05$) for stage prediction. Signatures with positive (negative) relationship to invasiveness have AUC values above (below) 0.50. **C)** Prognostic value of signatures for disease specific survival (DSS). Log hazard ratios (HR) and 95% confidence intervals are plotted. A confidence interval that does not cross the zero line corresponds to a statistically significant ($P < 0.05$) prognostic signature. **D)** Frequency of statistically significant signatures ($P < 0.05$) for DSS prediction. Signatures with positive (negative) relationship to poor prognosis have log HR values above (below) 0. Abbreviations: ns, not significant.

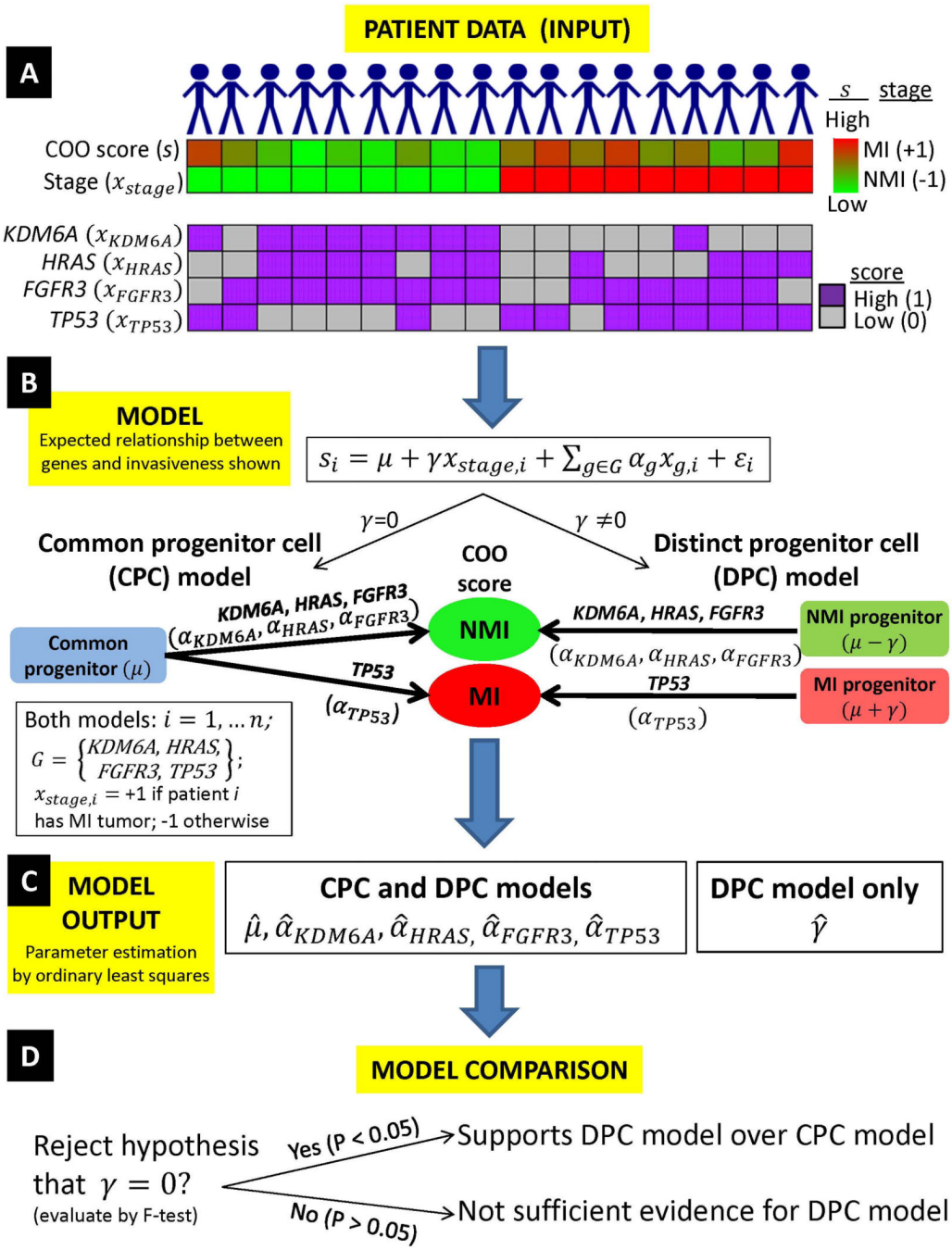

Figure 4. Overview of model and its use in determining whether urothelial cancer derives from a common or distinct progenitor cell

A) *Patient data (Input)* consists of the stage, Cell Of Origin (COO) score, and *KDM6A*, *HRAS*, *FGFR3*, and *TP53* signature scores for each patient. **B)** The COO score (*s*) of a sample is *modeled* as a function of *KDM6A*, *HRAS*, *FGFR3*, and *TP53* pathway alterations (signature scores) and the average COO score of the common (CPC model) or distinct (DPC model) progenitor cell of the tumor. Mutations in *KDM6A*, *HRAS*, and *FGFR3* are associated with NMI disease while *TP53* mutations are associated with MI disease (Ref #[3]) and are labeled this way for visualization purposes only. The statistical model makes no assumptions about whether signature scores are associated with NMI or MI tumors.

Mathematically, the CPC and DPC models differ only by γ. In the CPC model $\gamma = 0$ while in the DPC model $\gamma \neq 0$ and the NMI and MI progenitor cells have different basal levels (i.e., average COO scores). For details, see **Methods**. **C)** The model parameters *(model output,* e.g., γ) are estimated using ordinary least squares, and **D)** *Model comparison* is carried out using an F-test for the hypothesis that $\gamma = 0$. If this hypothesis is rejected ($P < 0.05$), then the DPC model fits the data significantly better than the CPC model. The DPC model fits patient COO scores significantly better than the CPC model does ($P < 0.05$) in five independent cohorts (see Table 1).

Table 1

Comparison of distinct (DPC) and common (CPC) progenitor cell models

Cohort	Model	γ	α_{KDM6A}	α_{HRAS}	α_{FGFR3}	α_{TP53}	P-value for DPC over CPC model
MSKCC	CPC	-	−71.2*	−7.5	−49.1*	6.74	0.01
	DPC	18.3*	−55.6*	−3.1	−44.2*	1.1	
CNUH	CPC	-	−50.9*	21.2*	−46.0*	22.77*	<0.001
	DPC	19.3*	−41.5*	18.0	−45.4*	5.8	
Lindgren	CPC	-	−7.8*	0.54	−.31	7.2*	0.01
	DPC	2.08*	−7.68*	0.58	1.01	6.02	
Dyrskjot	CPC	-	−1.05*	−.37	−.23	4.89*	<0.001
	DPC	1.38*	−.89	−.33	−.13	4.36*	
Blaveri	CPC	-	−7.44*	−5.43	−7.33*	2.45	0.001
	DPC	4.88*	−5.59*	−4.58	−5.23*	−1.45	

*P < 0.05.