Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer

Markus Riester*, 1 Jennifer M. Taylor*, 2 Andrew Feifer, 2 Theresa Koppie, 3 Jonathan E. Rosenberg, 4 Robert J. Downey, 5 Bernard H. Bochner, 2 and Franziska Michor 1†

1Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA. 2Urology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA. 3Department of Urology, University of California-Davis, Sacramento, CA. 4Bladder Cancer Center, Dana-Farber Cancer Institute, Boston, MA 02215, USA. 5Thoracic Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

*These authors contributed equally. †Author for correspondence. Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA. Tel: 671 632 5045. Fax: 617 632 4222. Email: michor@jimmy.harvard.edu.

Running head: Gene expression signature with nomogram improves the prediction of survival in high-risk bladder cancer

Keywords: Bladder cancer, gene expression analysis, molecular subtypes, survival analysis, bioinformatics

Word count: abstract: 241; text: 4,653

Number of figures and tables: 2 figures, 4 tables

Number of citations: 38
ABSTRACT

Purpose: We aimed to validate and improve prognostic signatures for high-risk urothelial carcinoma of the bladder.

Experimental Design: We evaluated microarray data from 93 bladder cancer patients managed by radical cystectomy to determine gene expression patterns associated with clinical and prognostic variables. We compared our results with published bladder cancer microarray datasets comprising 578 additional patients, and with 49 published gene signatures from multiple cancer types. Hierarchical clustering was utilized to identify subtypes associated with differences in survival. We then investigated whether the addition of survival-associated gene expression information to a validated post-cystectomy nomogram utilizing clinical and pathologic variables improves prediction of recurrence.

Results: Multiple markers for muscle invasive disease with highly significant expression differences in multiple datasets were identified, such as FN1, NNMT, POSTN and SMAD6. We identified signatures associated with pathologic stage and the likelihood of developing metastasis and death from bladder cancer, as well as with two distinct clustering subtypes of bladder cancer. Our novel signature correlated with overall survival in multiple independent datasets, significantly improving the prediction concordance of standard staging in all datasets (mean ΔC-statistic: 0.14, 95% CI 0.01-0.27; P < 0.001). Tested in our patient cohort, it significantly enhanced the performance of a postoperative survival nomogram (ΔC-statistic: 0.08, 95% CI -0.04-0.20; P < 0.005).

Conclusions: Prognostic information obtained from gene expression data can aid in post-treatment prediction of bladder cancer recurrence. Our findings require further validation in external cohorts and prospectively in a clinical trial setting.
**Translational Relevance:** Urothelial carcinoma of the urinary bladder is the fifth most common cancer in the Western world, representing 3% of the global cancer incidence. Bladder cancers exhibit a heterogeneous clinical behavior highlighted by frequent recurrences in patients with non-invasive tumors and the potential for the development of metastatic disease in those with invasive lesions. Due to this variation, better tools to predict prognosis and refine treatment are needed. Here we evaluated the utility and reproducibility of expression array differences in bladder tumors as potential prognostic indicators. Our analysis identified novel signatures associated with tumor stage, progression among patients with organ-confined disease, and molecular subtypes of tumors with relevance to disease-specific survival. Incorporation of our gene expression signatures furthermore significantly improved the prediction using a postoperative nomogram. Identifying patients at higher risk for progression or death provides additional rationale for selection of multimodality management, using chemotherapy and surgery, in an effort to improve survival.
INTRODUCTION

Urothelial carcinoma of the urinary bladder is the fifth most common cancer in the Western world, and the ninth most frequent cancer worldwide, representing 3% of cancers diagnosed globally (1). Treatment of bladder cancer has not changed significantly in over 20 years, and outcomes for patients remain suboptimal. Approximately 20-30% of newly diagnosed patients present with muscle-invasive (MI) disease (stages T2-4) or metastatic disease, while up to a third of patients with initially non-muscle-invasive (non-MI) disease (stages Ta/T1/Tis) later progress to MI or metastatic disease (2). Clinical features including stage and grade are strongly associated with outcome and play an important role in determining treatment. For example, the 5-year tumor-specific mortality rates range from less than 5% for low-stage and low-grade disease up to approximately 50% for all MI lesions (3, 4). However, even though grade and stage are important predictors of outcome, there remains significant variability in the prognosis of patients with similar characteristics. This highlights the need to identify additional tumor characteristics that predict clinical behavior.

In recent years, there has been a growing interest in the use of gene expression signatures for risk stratification of cancer patients (5). Multiple groups have produced urothelial cancer gene signatures predictive of a range of tumor characteristics and outcomes, including stage (6-8), molecular subtype (9, 10), likelihood of recurrence and progression of non-MI (11-13) and MI disease (14), and survival (7). Here we aimed to (i) explore if novel cancer gene expression patterns can be identified that stratify patients undergoing radical cystectomy for urothelial cancer by risk of recurrence and death; (ii) test our novel signatures and all previously developed signatures in all available bladder cancer microarray datasets; and (iii) investigate the gain in predictive accuracy of a validated clinical nomogram by combination with gene signature information.
MATERIALS AND METHODS

MSKCC Samples

Characteristics. We utilized previously unpublished cancer gene expression data from 93 patients undergoing radical cystectomy (RC) at Memorial Sloan-Kettering Cancer Center (MSKCC) between 1993 and 2004. Specimens were collected with MSKCC Institutional Review Board approval, and a waiver of authorization to review associated clinical data was obtained from the Board. The clinicopathologic characteristics are summarized in Table 1 (see Table S1 for details). In those 15 patients with non-MI disease on final pathologic analysis after RC, four had MI disease histologically in tissue obtained at the time of prior transurethral resection (TUR), and the remaining patients had high-risk features (e.g. extensive volume of disease, recurrent or BCG-refractory disease). Lymph node dissection was performed in 77 patients; no patient had metastatic disease at the time of RC. Chemotherapy was administered to 3 patients as neoadjuvant, 16 patients as adjuvant, and 19 patients as salvage for recurrent disease. Case selection was restricted to those with frozen specimens with measurable volume of malignancy and adequate percentage of tumor.

Clinical endpoints. Overall survival time was defined as the time from date of RC to date of death or last follow-up. Disease-specific survival was defined as the time from RC to death attributed to urothelial cancer, when death occurred with known and progressive metastatic disease. If death was recorded in the institutional database without knowledge of a recurrent cancer or with documentation of another malignancy or a non-malignancy cause, death was attributed to other causes. Recurrence was defined as pelvic if within the pelvis below the aortic bifurcation, distant if there was visceral metastasis or recurrence above the aortic bifurcation, and urothelial if within the remaining urinary tract (renal pelvis, ureter, urethra). In the analysis of the development of metastases, only pelvic and distance recurrences were included. Primary endpoint in this study was overall survival if not mentioned otherwise.

Sample preparation. All patient specimens collected at the time of surgery were processed expeditiously within the Department of Pathology and stored in an institutional biospecimen bank. Frozen bladder tissues were examined by a genitourinary pathologist to identify tumor content, which was microdissected. Ten 50-µm sections were cut, with confirmation of tumor
content by pathologic review. RNA was isolated with Trizol (Invitrogen) and cleaned with Rneasy mini kits (Qiagen) according to manufacturer protocols. Expression profiling was performed in the MSKCC Genomics Core Facility. In brief, 5-10 ug of RNA was transcribed utilizing a T7-oligo dT primer and converted to cDNA (Invitrogen cDNA synthesis kit). Biotinylated aRNA was produced from the cDNA using an in vitro transcription kit (Enzo Diagnostics). Following quality assessment using an Agilent Bioanalyzer, the aRNA was fragmented and hybridized onto Affymetrix U133 Plus 2.0 arrays, containing 54,675 probe sets. Data was uploaded to NCBI Geo (15) under accession number GSE31684.

**External datasets**

We obtained six independent bladder cancer microarray datasets (6, 7, 9, 10, 13, 16) (Table 2) totaling 578 additional patients and encompassing a broad spectrum of stages, grades, and histologic types. Overall survival data was available for the Blaveri et al., Sanchez-Carbayo et al., Lindgren et al., and Kim et al. cohorts (6, 7, 9, 13) (see Supplementary Data for preprocessing details). Disease-specific survival time information was only documented for patients in the cohort from Kim et al. The primary endpoint was therefore overall survival in the meta-analysis.

**Statistical analysis**

**Preprocessing.** Expression estimates were obtained by the GCRMA algorithm (17).

**Differential gene expression.** The gene expression profiles were evaluated for significant correlation with multiple clinical variables and outcomes, including stage, lymph node status, recurrence, and overall survival. For the identification of differentially expressed probe sets, we used the LIMMA method (18). If not stated otherwise, the cutoff of the adjusted p-values (the false discovery rate, FDR) (19) was set to 0.01 and the minimum fold change (FC) to 1.5. The probability of observing $n$ or more differentially expressed genes was estimated by permutation tests. For the external datasets, the lists of differentially expressed genes were downloaded from the supplementary material of the respective papers (6, 7, 9, 10, 13, 16) when available, otherwise calculated with LIMMA. When comparing lists of differentially
expressed genes with corresponding lists from external datasets, the significance of overlaps was calculated in Bioconductor (GSEABase), using 25,000 permutations.

**Machine learning.** Classifiers were generated with Fisher's linear discriminant and support vector machines (SVMs) and were leave-one-out cross-validated. The R packages MASS and e1071 were used for building the classifiers. We utilized a linear SVM kernel and default, untuned parameters. ROC curves were generated with the ROCR package (20).

**Risk category creation.** For each patient, a risk score was calculated by a Cox proportional hazards model that was fitted using the gene expression profiles of all remaining patient samples (leave-one-out cross-validation). The risk score was defined as the sum of the gene signature expression values, weighted by the Cox regression coefficients (21). Each individual was then classified into low-risk or high-risk categories, based on the median leave-one-out cross-validated risk score: patients were classified as low-risk when their risk score was smaller than the cohort median, otherwise as high-risk. This was done independently for all datasets for which survival information was available. The significance of survival curve differences of cross-validated models was estimated with permutation tests. The process from cross-validation to risk stratification was repeated 500 times with shuffled survival labels. This empirical chi-square distribution was then utilized to estimate p-values. For non-cross-validated models, the log-rank test was used. Multivariate survival prediction models were compared by the C-statistic, an estimator of the model concordance (22), and by likelihood-ratio tests. The concordance represents the probability that given two random non-censored individuals, the one with the higher risk score has a shorter survival time.

**Clustering analysis.** For unsupervised clustering, we used the Ward clustering method and the Pearson correlation distance as implemented in the pvclust R package (23). The significance of a cluster was reported as its bootstrap value, which is the proportion of 10,000 bootstrap samples showing this particular cluster topology. The clustering was applied to the six external datasets of bladder cancer samples.
Meta-analysis using published data sets

Published signatures. We compiled 49 published gene signatures (Table S2) associated with malignancy: 39 from Lauss et al. (24) as well as other bladder (11-14) and melanoma signatures (25, 26); the melanoma signatures were included to test whether signatures from other solid tumors would perform well in bladder cancer. The 49 gene signatures were tested in our dataset and the 4 external gene expression datasets with survival information. Associations of gene expression signatures with outcome and other covariates were calculated using globaltest (27); this test can be used to estimate whether the expression of a group of genes is significantly associated with a particular response variable, for example stage or survival. Gene signatures were further tested by leave-one-out cross-validation. To avoid problems with highly correlated covariates in multivariate Cox proportional hazards models, expression values were scaled by principal component analysis (PCA). The number of components was chosen so that at least 99% of the expression variance was included in the model, with a maximum of 20 components. The choice of this cutoff was examined by varying the maximum number of components from 3 to 30 (Table S2). The performance of these cross-validated models was reported as the C-statistic of a univariate Cox model with the risk score as covariate.

Feature selection. The signatures were optimized by stepwise selection, an iterative procedure which serially removes and adds probe sets from a pool of candidates. The procedure was terminated when adding, removing or replacing a probe set did not further significantly improve the mean leave-one-out cross-validated Area Under Receiver Operating Curve (AUC) (for the stage and subtype signatures) or C-statistic (for the survival signature) over all datasets (see Supplemental Data for details). Thus all datasets were used for training and validation. In total, over a million promising signatures were cross-validated in all datasets.

Addition of new survival signature to existing predictive model for recurrence-free survival

We calculated the recurrence-free survival probabilities in our patients using the postoperative nomogram developed by the International Bladder Cancer Nomogram Consortium (IBCNC) (28). This nomogram, developed from data of over 9,000 patients, includes age, sex, time from
diagnosis to surgery, pathologic tumor stage and grade, tumor histologic subtype, and regional lymph node status. The contribution of our newly curated gene signature for overall survival to the predictive ability of the nomogram was then evaluated. Multivariate modeling with nomogram information applied the leave-one-out cross-validated gene signature risk score as a second covariate. Thus two Cox models were used: the first model estimated the risk as described for the Kaplan-Meier analysis using the gene signature alone; the second model combined the nomogram and the signature score. The right endpoint of this second model was set to be recurrence-free survival. Concordance was estimated with the C-statistics approach (22) for a univariate model with the nomogram alone and then for a multivariate model with the addition of the survival gene signature. Prediction models were compared by the likelihood ratio test.

In Figure 1, we provide an overview of the analyses performed in this study.
RESULTS

Differentially expressed genes are associated with clinical features and outcome

Comparison by pathologic stage. When comparing non-MI with MI samples in our data set, we found 636 significantly differentially expressed probe sets, which significantly overlapped with the corresponding lists of genes in the six independent studies ($P < 0.0001$, Tables S3a, S4-S5, Figure S1). FN1, a member of the integrin signaling pathway, and several other members or close interaction partners of this pathway were upregulated in MI as compared to non-MI tumors in most datasets we investigated; these genes include ACTN1, COL1A2, COL3A1, COL5A2, COL6A3, COL11A1, COL16A1, FBN1, FLNA, LUM, TGFBI, and TNC. Furthermore, we found that members of the transforming growth factor-beta (TGF-β) signaling pathway were present in most lists of differentially expressed genes; for example, SMAD3, SMAD6, and BMP7 were overexpressed in non-MI as compared to MI tumors, while INHBA, NNMT, and POSTN were overexpressed in MI samples. To further characterize the association of gene expression with pathologic stage, we created a classifier based on consistently differentially expressed genes among all datasets. The optimized classifier (Table S6) identified MI tumors with a mean accuracy of 89.0% (88.4% SVM, Figure S2) over all datasets.

Comparison by recurrence status. We then compared the gene expression patterns of patients who developed metastasis (pelvic or distant recurrence) and died of bladder cancer with those of patients who did not. When considering samples of all stages, we identified no significantly differentially expressed genes. However, when restricting this comparison to patients with pathologically organ-confined MI (pT2N0) tumors (17 patients, 5 of whom developed metastasis), we identified 53 differentially expressed genes with FDR < 10% in our dataset ($P < 0.06$, Table S3b). In the external datasets, the number of pT2N0 tumors with later metastasis was too small for the identification of significantly differentially expressed genes in this comparison. Among those genes that were differentially expressed in patients who developed metastasis and died were PERP (a TP53 apoptosis factor), ATXN10 (an activator of the Ras-MAPK pathway) [24]; and PCM1 (associated with papillary thyroid cancer, chronic myeloid leukemia, and myeloproliferative disorders) [25].
Identification of a robust survival signature for MI tumors

Validation of published signatures. We then evaluated 49 published gene signatures (Table S2) alongside our classifiers described above for correlation with clinical features in our dataset and with the four independent gene expression datasets with available survival information (6, 7, 9, 13). Due to the smaller number of genes on the Lindgren et al. and Blaveri et al. cDNA microarrays, only fractions of these signatures’ genes could be mapped. We observed overfitting of published survival signatures so that most signatures achieved significance, in terms of survival information, only in the datasets used to identify them (Table 3). Applying previously published signatures to our cohort, we found no significant association between survival and any signatures. Nevertheless, when we analyzed all genes from all survival and progression signatures in a univariate fashion, several markers that achieved low p-values in multiple datasets emerged; for example BST2, HLA-G, ICAM1 and LIMCH1 from the Smith et al. signature (14); DGCR2, DSC3 and SCEL from Kim et al. (13); UBE2C, VCAM1 and WNT5B from Blaveri et al. (6); APOBEC3B from Lindgren et al. (9) and ATXN10, C17orf39 and INPP4B from Sanchez-Carbayo et al. (7) (Table S2). Especially striking are MAP kinases, with different members present in various signatures (7, 9, 11, 14, 29).

Validation of published signatures, stratified by molecular subtype. As reported in multiple other studies (6, 7, 9), bladder tumors cluster in two very distinct molecular subtypes. We reproduced this finding in all analyzed datasets (Figure S4) and further, found significant differences in survival between the two subtypes (Figure S5). The expression differences between the subtypes were remarkable and overlapped significantly across datasets (P < 0.00004, Table S5). This allowed a robust classification of tumors by subtype with a novel 19-gene signature (Table S7, Figure S6). The very heterogeneous landscape of MI bladder tumors motivated us to evaluate the performance of all signatures in the two subtypes separately. Two progression signatures, one developed for bladder (6) and one for breast cancer (30), were significantly associated with survival in our cohort when applied to one of the two subtypes (Figure S7). By analyzing primary and recurrent tumors from the Kim et al. cohort (13), we found evidence that tumors can progress from one molecular subtype to the other (Table S8).
Curation of a novel survival signature. We then curated a new 20-gene overall survival signature for MI tumors (Table 4) by stepwise optimization and tested its predictive accuracy in our dataset and the four external datasets with available survival information. A leave-one-out cross-validated multivariate Cox proportional hazards model based on this signature classified tumors into two equally sized risk groups with significantly different survival in all datasets (Figure 2a-e). As an example, with the model applied to the dataset by Kim et al. (13), there was a 5-fold increase in the risk of death in those classified as high-risk (HR 5.05, 95% CI 2.26 – 11.3, P < 0.008) compared to those classified as low-risk. See Figures S8-S9 for details of the signature. Neoadjuvant chemotherapy, administered to three patients in our cohort, had no impact on the prediction accuracies (Table S9). We next compared the prediction accuracies of the signature with pathologic stage and grade in all cohorts and found highly significant improvements of model concordance and likelihood in all datasets (mean ΔC-statistic: 0.14, 95% CI 0.01-0.27; likelihood-ratio test: P < 0.001; Figure 2f, Table S10).

Addition of survival signature to the nomogram increases its predictive accuracy
We then incorporated the new signature developed for MI tumors (Table 4) into the International Bladder Cancer Nomogram Consortium (IBCNC) nomogram (28), which was designed to predict 5-year recurrence-free survival after RC. Necessary data points for the nomogram were not available in the external datasets. When tested in our MI patient cohort, a multivariate Cox model predicting recurrence-free survival with both nomogram and gene signature score (C-statistic: 0.66, 95% CI 0.56 - 0.76) was a significant improvement (ΔC-statistic: 0.08, 95% CI -0.04-0.20, p-value < 0.005, likelihood-ratio test) over a model based on the nomogram score alone (C-statistic: 0.58, 95% CI 0.43 - 0.72). As the signature was designed for overall survival due to the lack of recurrence dates in the independent data, the prediction accuracies increased profoundly in a model predicting overall survival (ΔC-statistic: 0.15, 95% CI 0.07-0.23, Table S10).
DISCUSSION

This study demonstrates the utility of adding molecular information to an existing prognostic tool in bladder cancer patients who undergo RC. Gene expression data from a new cohort of bladder cancer patients was analyzed and compared to gene expression data from multiple published studies. A robust expression signature was identified that improved the predictive outcome of a well-accepted nomogram, and was independently predictive of survival in all cohorts where survival data was available. These findings are particularly striking given the relative homogeneity of the population analyzed, since our cohort, having all undergone the most aggressive surgical therapy, is highly selected for being at high risk of death from disease.

One important contribution of this study is the clinical validation of published gene signatures. To be clinically relevant, new survival markers must (i) stratify patients in groups with significantly different survival; (ii) deliver survival information that is not already included in established clinically used predictors such as grade and stage; and (iii) increase the accuracies of existing prediction models to an extent that warrants the cost and effort of obtaining biomarker status (31). Previously published gene signatures (Table S2), however, performed poorly in our patient cohort: no gene signature provided more survival information than would be expected by chance. This finding can partly be explained by cross-platform issues like incomplete mapping of genes to available probe sets or probe sets targeting different exons. Another reason is a likely overfitting of published signatures due to rather small training datasets for the very heterogeneous group of MI bladder cancers. To circumvent these limitations, we optimized the survival signatures for the subgroup of MI tumors, and showed that gene expression clearly stratifies patients into 2 groups, denoted as low-risk and high-risk.

The new signature further segregated by risk 305 patients with MI tumors in the external datasets containing gene expression and survival information. Interestingly, 4 genes in our MI survival signature, \textit{APOBEC3B}, \textit{DGCR2}, \textit{PRMT1}, \textit{SLC1A5}, are located on chromosomes 22 and 19, in neighboring chromosome bands of SNPs recently described as associated with an increased risk of developing bladder cancer (32). In addition to our signature, several other genes (Table S2) significantly correlated with survival in multiple independent cohorts. These
genes identified are potential therapeutic targets and worthy of further investigation to understand their role in bladder cancer. For example, inhibitors of MAP2K1 (i.e. MEK1), which is part of our overall survival signature (Table 4), are currently being evaluated in several clinical trials and our results provide a rationale for this in bladder cancer. Our screen for potential biomarkers in multiple datasets thus also has important applications in gene selection for technologies such as NanoString®, which can target only a limited number of genes.

To place our signature in the context of currently known clinical data, we applied the survival signature to a robust nomogram that predicts outcome after RC (28). This postoperative nomogram was shown to have a significantly better predictive accuracy (concordance index 0.75) than standard TNM (CI 0.68, P < 0.001) or pathologic (CI 0.62, P < 0.001) subgroupings. When this model was applied to our patient cohort with the addition of our newly developed survival signature, we observed statistically significant improvements in the prediction of risk of death as compared to the use of the nomogram only. As the signature reflects unique molecular features associated with risk, its addition to the nomogram thus provides more insights into the biologic behavior of a patient’s malignancy, and leads to improved prognostication as compared to the use of the nomogram alone. A more accurate prediction of progression may help select the higher-risk patients for additional treatment, such as adjuvant systemic therapy after surgery. External validation of these findings is ongoing.

In addition, our analysis led to the identification of 53 genes significantly associated with the development of metastasis and death of patients with pathologically organ-confined MI disease (pT2N0). This list of genes establishes a molecular signature of less favorable outcome, which may help stratify patients in this identically staged subgroup of bladder tumors to more or less intensive multimodal therapy. Patients with this stage of disease are typically followed expectantly, but evaluation of gene expression data or testing for downstream products of expression may identify those patients at higher risk for progression who may benefit from systemic therapy after radical surgery. The investigation of this list in a prospective collection
of similar cystectomy samples is essential. An interesting additional opportunity for evaluation would be among patients with MI disease at the time of TUR—by characterizing a greater risk of aggressive behavior, we may more carefully select patients for chemotherapy prior to radical surgery as well, to improve our survival outcomes using multimodal therapy.

We found that fibronectin 1 (FN1) was significantly overexpressed in MI tumors as compared to non-MI samples in all datasets we analyzed. The role of FN1 in cancer invasion and metastasis is still poorly understood, but its increased expression might be the result of an enhanced recruitment of cancer-associated fibroblasts (33, 34). The consistency across datasets suggests that the abundant differential expression of FN1 and other extracellular matrix (ECM) genes is unlikely due to stromal contamination. Our stage classifier (Table S6) could identify patients as having MI disease with a mean accuracy of 89%, which is within the range of previously published classifiers (6, 24). However, our stage classifier was validated in more datasets and consists of fewer genes than the published classifiers, which offer advantages for its potential clinical use. Furthermore, the accuracy rate of 89% does not imply a false classification rate of 11% as some tumors might have been incorrectly classified histologically. Urothelial cancer of the bladder is exceptionally difficult to stage accurately prior to cystectomy, given technical limitations of endoscopic resection and imaging. Significant rates of upstaging from time of TUR to RC have been reported, ranging from 31-61% (35-37). In our patient cohort, over 80% of the tumors were determined to have higher pathologic stage at the time of cystectomy, when compared to the preoperative TUR specimens, with many tumors found to be locally advanced (>pT2) when clinically staged as organ-confined (<pT2).

The unsupervised clustering produced two strikingly different groups of patients, who were not clearly distinguishable by many clinical variables such as the presence of lymph node metastases or recurrence status. Gene expression showed large differences between the two groups (Table S3c) with highly significant overlaps in all datasets (Table S5), allowing a robust identification of cluster membership by a 19-gene signature (Table S7). Lindgren et al. (9)
similarly identified two distinct molecular subtypes (denoted as MS1 and MS2) by analyzing 144 tumor samples obtained by transurethral bladder biopsy, representing a wide range of tumor stage and grade. Their investigation, which included whole genome and specific mutational analyses, identified a greater number of genomic alterations in the MS2 group, containing nearly all of the M1 samples. The genomic instability in the MS2 group was also associated with expression of genes involved in cell-cycle pathways and cellular transformation (9), all indicative of more aggressive behavior and worse prognosis. Consistent with their results, we found a progression from MS1 to MS2 to be more likely than a progression from MS2 to MS1 when analyzing 23 recurrent tissue samples from Kim et al. (13). While only stage had a significant association with subtype, other covariates not available in this study, for example response to treatments, could be tested for association in prospective data. Not surprisingly, survival signatures often showed very different prediction accuracies in the two subtypes. This heterogeneity of bladder cancer tumors thus points to the necessity of large training data to establish robust signatures. Interestingly, from all 49 considered previously published signatures, a breast cancer progression signature (30) showed the highest association with overall survival in the MS2 group of our patient cohort (Figure S7a). This finding suggests that optimizing thoroughly validated signatures designed for other solid tumors might yield more robust signatures than developing new signatures from scratch, in particular when only small training datasets are available. The identification of different molecular signatures for subtype suggests that these subtypes, which are strongly associated with invasion status, may be a building block for future work to non-invasively determine the aggressiveness of tumors; for example, these changes in expression may be detectable in urine.

Our dataset represents a high-risk population of bladder cancer patients who underwent RC at a high-volume academic center. The sample size is large for a single-center report of microarray analysis and comprises, to our knowledge, the largest number of high-risk bladder cancer patients of all published microarray datasets. One limitation of our findings is, however, a possible element of overfitting, resulting from the use of the same data in testing and
validation steps, although this concern is mitigated through the use of statistical methods such as cross-validation. This points to the imperative for validation in independent data of any gene signature associated with survival. Such a validation is necessary because signatures established from single, moderately sized datasets often perform poorly in external data due to the relatively high probability that genes correlate with survival by chance. However, extending the sample size with independent data, as performed here, can minimize this problem. For example, a gene found to be predictive in one dataset could arise due to chance, but if one gene is predictive with consistent expression in multiple datasets, it is more likely to be a robust survival marker.

A considerable strength of our investigation is its comprehensive effort to combine the majority of published data on gene expression in bladder cancer, all arrayed on different platforms, in a meaningful way. With this study, we have shown that meta-analyses across platforms are feasible. Especially in rare cancers or subtypes, the increase in sample size might often outweigh the disadvantages of meta-analyses, such as the heterogeneity of patient cohorts, the different treatments before and after surgery, and the decrease in potential markers when focusing on genes present on all platforms. Nonetheless, large independent patients cohorts are needed for further validation and optimization (i.e., for other endpoints such as recurrence-free survival or for particular microarray platforms) of the prognostic signatures, before warranting a clinical trial (38). The results of this study strongly motivate such sample collection for urothelial tumors.

CONCLUSIONS

Our gene expression analysis identifies novel signatures predictive of tumor stage, progression among patients with organ-confined disease, molecular subtypes of tumors, and overall survival. We further incorporated six published gene expression datasets for a large-scale comparison and validation of predictive signatures, and found a significant increase in predictive accuracy of a postoperative nomogram by addition of our gene signature. Multiple genes identified across several sets have emerged as highly promising candidates for further
investigation. The identification of patients at higher risk for progression or death will provide additional rationale for multimodality management, using chemotherapy and surgery, in an effort to improve survival.

ACKNOWLEDGMENTS

The authors acknowledge support from the NCI initiative to found Physical Science Oncology Centers (U54CA143798) (MR, FM) and an NRSA training grant T32-CA82088 from the National Institutes of Health (JMT). The authors would like to thank Parantu Shah, Dhruv Sharma, Levi Waldron, Marta Sanchez-Carbayo, Mithat Gönen, Samir Amin, Yi Li, and the Michor lab.

REFERENCES


FIGURE LEGENDS

**Figure 1. Overview of the study.** This study represents a massive meta-analysis of our cohort of high-risk bladder cancer patients (Table 1) and 6 independent studies (Table 2). All data was subject to comprehensive differential expression and survival analyses and to hierarchical clustering. Machine learning algorithms were used to find gene signatures predictive of stage, survival and molecular subtype. The survival gene signature was then included in a validated postoperative nomogram, developed by the International Bladder Cancer Nomogram Consortium (IBCNC) (28).

**Figure 2. Prediction accuracy of our novel gene signature.** (a-e) Kaplan-Meier plots for low- and high-risk groups in five datasets. Hazard ratios (HRs) indicate how well the signature separates tumors. A HR of 3.1, for example, corresponds to 3.1-fold increase in risk for the high-risk group. Some genes were not present on all microarray platforms. The novel signature is shown in Table 4. (a) Riester (all 20 genes); (b) Blaveri (18 genes); (c) Kim (20 genes); (d) Lindgren (10 genes); (e) Sanchez-Carbayo (20 genes). In all plots, patients were stratified in two equally sized groups by the cohort medians of the calculated risk scores. We further tested whether the gene signature improves the predictive utility of stage and grade, two established predictors available for all 5 datasets. This comparison was done by analyzing the concordance (C-statistic, (22)) of the models. A concordance of 0.5 corresponds to a random model, one of 1.0 of a perfect model. (f) C-statistics for all 5 datasets for a multivariate model with stage, grade and the gene signature risk score; for model with gene signature risk score alone; and for a model with only stage and grade. Error bars indicate 95% confidence intervals.
Table 1. Clinicopathologic characteristics of the tumor samples.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient age at RC, median (range), years</strong></td>
<td>69.1 (32.1-91.1)</td>
</tr>
<tr>
<td><strong>Gender, number (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>25 (27%)</td>
</tr>
<tr>
<td>Male</td>
<td>68 (73%)</td>
</tr>
<tr>
<td><strong>Histologic type</strong></td>
<td></td>
</tr>
<tr>
<td>Transitional Cell Carcinoma(TCC)</td>
<td>88 (95%)</td>
</tr>
<tr>
<td>TCC/Squamous</td>
<td>5 (5%)</td>
</tr>
<tr>
<td><strong>RC stage, number (%)</strong></td>
<td></td>
</tr>
<tr>
<td>pTa</td>
<td>5 (5%)</td>
</tr>
<tr>
<td>pT1</td>
<td>10 (11%)</td>
</tr>
<tr>
<td>pT2</td>
<td>17 (18%)</td>
</tr>
<tr>
<td>pT3</td>
<td>42 (45%)</td>
</tr>
<tr>
<td>pT4</td>
<td>19 (21%)</td>
</tr>
<tr>
<td><strong>Lymph node status</strong></td>
<td></td>
</tr>
<tr>
<td>pN0</td>
<td>49 (52%)</td>
</tr>
<tr>
<td>pN+</td>
<td>28 (30%)</td>
</tr>
<tr>
<td>pNx(^A)</td>
<td>16 (18%)</td>
</tr>
<tr>
<td><strong>Length of follow-up, median (range), mos</strong></td>
<td>32 (1-175)</td>
</tr>
<tr>
<td><strong>Last known status, number (%)</strong></td>
<td></td>
</tr>
<tr>
<td>NED</td>
<td>28 (30%)</td>
</tr>
<tr>
<td>Death from other cause</td>
<td>27 (29%)</td>
</tr>
<tr>
<td>Death from bladder cancer</td>
<td>38 (41%)</td>
</tr>
<tr>
<td><strong>Recurrence, number (%)(^B)</strong></td>
<td></td>
</tr>
<tr>
<td>No recurrence</td>
<td>30 (32%)</td>
</tr>
<tr>
<td>Urothelial</td>
<td>9 (10%)</td>
</tr>
<tr>
<td>Pelvic</td>
<td>20 (21%)</td>
</tr>
<tr>
<td>Distant site</td>
<td>34 (37%)</td>
</tr>
</tbody>
</table>

RC = radical cystectomy;  
A: Lymph node dissection not done or no lymph nodes analyzed  
B: numbers total more than 100%, as some patients developed recurrence in more than one site.
Table 2. Independent bladder cancer microarray datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accession Number</th>
<th>Platform</th>
<th># Probes</th>
<th># Samples (# invasive)</th>
<th>Median Age (range)</th>
<th>% Males</th>
<th>% RC</th>
<th>Median Follow-Up Time (max)</th>
<th>% Death (DOD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blaveri (6)</td>
<td>GSE1827</td>
<td>cDNA microarray</td>
<td>10,368</td>
<td>80 (53)</td>
<td>66 (28-113)</td>
<td>70.0%</td>
<td>62.5%</td>
<td>13 (145)</td>
<td>55.0% (n/a)</td>
</tr>
<tr>
<td>Sanchez-Carbajo (7)</td>
<td>-</td>
<td>Affymetrix U133A</td>
<td>22,283</td>
<td>90 (65)</td>
<td>69 (37-85)</td>
<td>68.9%</td>
<td>n/a</td>
<td>25 (76)</td>
<td>40.0% (n/a)</td>
</tr>
<tr>
<td>Lindgren (9)</td>
<td>GSE19915</td>
<td>cDNA microarray</td>
<td>2,506</td>
<td>144 (47)</td>
<td>n/a</td>
<td>n/a</td>
<td>32.6%</td>
<td>46 (180)</td>
<td>18.1% (n/a)</td>
</tr>
<tr>
<td>Dyrskjøt (10)</td>
<td>GSE89</td>
<td>Affymetrix Hu6800</td>
<td>7,129</td>
<td>40 (9)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Kim (13)</td>
<td>GSE13507</td>
<td>Illumina human-6 v2.0</td>
<td>43,148</td>
<td>165 (62)</td>
<td>66 (24-88)</td>
<td>81.8%</td>
<td>n/a</td>
<td>37 (137)</td>
<td>41.8% (19.4%)</td>
</tr>
<tr>
<td>Stransky (16)</td>
<td>E-TABM-147</td>
<td>Affymetrix U95A/U95A v2</td>
<td>12,626</td>
<td>59 (33)</td>
<td>n/a</td>
<td>83%</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

RC = radical cystectomy; DOD = death of disease; n/a not available
Table 3. Performance of published gene signatures. The table shows performance measures of all published bladder cancer survival gene signatures and three progression signatures. All signatures were tested in all public datasets with survival information (muscle-invasive tumors only). The p-values represent the probabilities that there is no association between expression and survival (27). The C-statistic ($C$) of a leave-one-out cross-validated Cox model with the gene intensities as covariates is reported as a second performance measure. A $C$ of 0.5 corresponds to a random model, one of 1.0 to a perfect model. All signatures were originally identified with the corresponding dataset, except for the Smith et al. signature (14), which was obtained by training with the Sanchez-Carbayo dataset (7) and with unpublished data. Our signature was obtained by using all datasets for training. Tested right endpoint was overall survival in all datasets, not necessarily the endpoint for which the signature was designed for. See Figure S6 for a detailed analysis of 41 other signatures.

<table>
<thead>
<tr>
<th>Signature</th>
<th>Endpoint</th>
<th>Stage</th>
<th>Datasets</th>
<th>Blaveri (6)</th>
<th>Kim (13)</th>
<th>Lindgren (9)</th>
<th>Sanchez-Carbayo (7)</th>
<th>Riester</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Blaveri (6)</td>
<td>P</td>
<td>C</td>
<td>P</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>Blaveri (6)</td>
<td>OS</td>
<td>Mi</td>
<td></td>
<td>0.01</td>
<td>0.66</td>
<td>0.66</td>
<td>0.57</td>
<td>0.15</td>
</tr>
<tr>
<td>Kim (13)</td>
<td>OS</td>
<td>Mi</td>
<td></td>
<td>0.41</td>
<td>0.60</td>
<td>0.00</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>Kim (13)</td>
<td>DSS</td>
<td>Mi</td>
<td></td>
<td>0.11</td>
<td>0.57</td>
<td>0.00</td>
<td>0.72</td>
<td>0.84</td>
</tr>
<tr>
<td>Kim (13)</td>
<td>Progression</td>
<td>non-Mi</td>
<td></td>
<td>0.70</td>
<td>0.50</td>
<td>0.09</td>
<td>0.56</td>
<td>0.53</td>
</tr>
<tr>
<td>Lindgren (9)</td>
<td>OS</td>
<td>all</td>
<td></td>
<td>0.29</td>
<td>0.55</td>
<td>0.17</td>
<td>0.55</td>
<td>0.10</td>
</tr>
<tr>
<td>Sanchez-Carbayo (7)</td>
<td>OS</td>
<td>Mi</td>
<td></td>
<td>0.15</td>
<td>0.56</td>
<td>0.56</td>
<td>0.54</td>
<td>0.33</td>
</tr>
<tr>
<td>Sanchez-Carbayo (7)</td>
<td>Progression</td>
<td>Mi</td>
<td></td>
<td>0.21</td>
<td>0.50</td>
<td>0.02</td>
<td>0.58</td>
<td>0.42</td>
</tr>
<tr>
<td>Smith (14)</td>
<td>Progression</td>
<td>Mi</td>
<td></td>
<td>0.42</td>
<td>0.54</td>
<td>0.17</td>
<td>0.54</td>
<td>0.15</td>
</tr>
<tr>
<td>Riester</td>
<td>OS</td>
<td>MI</td>
<td>0.04</td>
<td>0.76</td>
<td>0.04</td>
<td>0.75</td>
<td>0.002</td>
<td>0.87</td>
</tr>
</tbody>
</table>

$P = p$-value; $C = C$-statistic; OS = overall survival; DSS = disease-specific survival, MI=muscle-invasive

Downloaded from cancerres.aacrjournals.org on April 20, 2017. © 2012 American Association for Cancer Research.
Table 4. Gene signature predictive of overall for patients with MI tumors (stage T2-T4).

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Probeset</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOBEC3B</td>
<td>206632_s_at</td>
<td>apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B</td>
</tr>
<tr>
<td>ATF3</td>
<td>202672_s_at</td>
<td>activating transcription factor 3</td>
</tr>
<tr>
<td>CCL5</td>
<td>204655_at</td>
<td>chemokine (C-C motif) ligand 5</td>
</tr>
<tr>
<td>DGCR2</td>
<td>214198_s_at</td>
<td>DiGeorge syndrome critical region gene 2</td>
</tr>
<tr>
<td>ENDOD1</td>
<td>212573_at</td>
<td>endonuclease domain containing 1</td>
</tr>
<tr>
<td>FADD</td>
<td>202535_at</td>
<td>Fas (TNFRSF6)-associated via death domain</td>
</tr>
<tr>
<td>JUNB</td>
<td>203022_at</td>
<td>ribonuclease H2, subunit A</td>
</tr>
<tr>
<td>LMO7</td>
<td>242722_at</td>
<td>LIM domain 7</td>
</tr>
<tr>
<td>MAP2K1</td>
<td>202670_at</td>
<td>mitogen-activated protein kinase kinase 1</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>225927_at</td>
<td>mitogen-activated protein kinase kinase 1</td>
</tr>
<tr>
<td>PDGFC</td>
<td>218718_at</td>
<td>platelet derived growth factor C</td>
</tr>
<tr>
<td>PEA15</td>
<td>200787_s_at</td>
<td>phosphoprotein enriched in astrocytes 15</td>
</tr>
<tr>
<td>PFN1</td>
<td>214617_at</td>
<td>perforin 1 (pore forming protein)</td>
</tr>
<tr>
<td>PPP1R12A</td>
<td>201604_s_at</td>
<td>protein phosphatase 1, regulatory (inhibitor) subunit 12A</td>
</tr>
<tr>
<td>PRDX1</td>
<td>208680_at</td>
<td>peroxiredoxin 1</td>
</tr>
<tr>
<td>PRMT1</td>
<td>206445_s_at</td>
<td>protein arginine methyltransferase 1</td>
</tr>
<tr>
<td>SLC1A5</td>
<td>208916_at</td>
<td>solute carrier family 1 (neutral amino acid transporter), member 5</td>
</tr>
<tr>
<td>TNFAIP6</td>
<td>206025_s_at</td>
<td>tumor necrosis factor, alpha-induced protein 6</td>
</tr>
<tr>
<td>TSG101</td>
<td>201758_at</td>
<td>tumor susceptibility gene 101</td>
</tr>
<tr>
<td>TSPAN5</td>
<td>209890_at</td>
<td>tetraspanin 5</td>
</tr>
</tbody>
</table>
Figure 1

Gene expression from 7 cohorts

Differential Expression Analysis
Survival Analysis
Hierarchical Clustering

Machine Learning

Gene Signatures
Stage, Survival, Subtype

IBCNC Nomogram
HR 3.1; 95% CI, 1.7 to 5.66; P < 0.016

Overall Survival

Time (Months)

Low Risk 39 35 25 21 18 17
High Risk 39 18 10 8 8 7

Numbers at risk
HR 3.83; 95% CI, 1.82 to 8.1; P < 0.016

Overall Survival

Time (Months)

Low Risk

High Risk

Numbers at risk

26 19 15 10 8 4
27 8 3 3 3 2
Overall Survival

HR 8.47; 95% CI, 2.78 to 25.8; P < 0.008

Figure 2d
Figure 2e

HR 3.57; 95% CI, 1.59 to 7.98; P < 0.02

Overall Survival

Time (Months)

0.0 0.2 0.4 0.6 0.8 1.0

0 12 24 36 48 60

Low Risk: 32 25 18 11 9 4
High Risk: 33 20 12 7 6 4

Numbers at risk
Figure 2f

Concordance (C-Statistic)

Model
- Combined
- Signature
- Stage + Grade

Blaveri  Kim  Lindgren  Riester  Sanchez
Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer

Markus Riester, Jennifer Taylor, Andrew Feifer, et al.

Clin Cancer Res  Published OnlineFirst January 6, 2012.