Bladder Cancer Outcome and Subtype Classification by Gene Expression

Ekaterini Blaveri,[1] Jeff P. Simko,[2] James E. Korkola,[1] Jeremy L. Brewer,[1] Frederick Baehner,[2] Kshama Mehta,[1] Sandy DeVries,[1] Theresa Koppie,[3] Sunanda Pejavar,[1] Peter Carroll,[3] and Frederic M. Waldman[1,3,4]

Abstract Models of bladder tumor progression have suggested that genetic alterations may determine both phenotype and clinical course. We have applied expression microarray analysis to a divergent set of bladder tumors to further elucidate the course of disease progression and to classify tumors into more homogeneous and clinically relevant subgroups. cDNA microarrays containing 10,368 human gene elements were used to characterize the global gene expression patterns in 80 bladder tumors, 9 bladder cancer cell lines, and 3 normal bladder samples. Robust statistical approaches accounting for the multiple testing problem were used to identify differentially expressed genes. Unsupervised hierarchical clustering successfully separated the samples into two subgroups containing superficial (pT_a and pT_1) versus muscle-invasive (pT_2-pT_4) tumors. Supervised classification had a 90.5% success rate separating superficial from muscle-invasive tumors based on a limited subset of genes. Tumors could also be classified into transitional versus squamous subtypes (89% success rate) and good versus bad prognosis (78% success rate). The performance of our stage classifiers was confirmed *in silico* using data from an independent tumor set. Validation of differential expression was done using immunohistochemistry on tissue microarrays for cathepsin E, cyclin A2, and parathyroid hormone–related protein. Genes driving the separation between tumor subsets may prove to be important biomarkers for bladder cancer development and progression and eventually candidates for therapeutic targeting.

Bladder carcinoma is a heterogeneous neoplasm, presenting as either superficial disease (80%) or muscle-invasive disease (20%). This varied presentation results in widely divergent clinical outcomes, and it is thought that such outcomes reflect different genetic pathways (1–5). Only 10% to 30% of superficial tumors (pT_a and pT_1) will progress to invasive disease, whereas muscle-invasive tumors (pT_2-pT_4) have a much less favorable prognosis, with 50% of patients already harboring clinically detectable or occult metastatic disease at the time of initial presentation. Bladder cancers also may show divergent histopathologies, with 90% of cases in the United States and western Europe presenting as urothelial carcinoma [either as pure transitional cell carcinoma (TCC) or as TCC with focal squamous or glandular differentiation] and 3% to 6% as pure squamous cell carcinoma (SCC).

SCC seems to follow a distinct pathway of development, presumably originating from metaplastic squamous mucosa in an inflammatory field. It is the predominant pattern of bladder cancer seen in Africa and the Middle East and is thought to be due to schistosomiasis and the accompanying inflammation, which is prevalent in these regions. SCC in the United States generally has a poor prognosis, with a 10-year survival of only 5% to 10% (6), and successful treatment relies heavily on early detection (7).

At present, conventional diagnosis for bladder cancer is based on morphologic and pathologic criteria (histology, stage, and grade). These criteria provide essential prognostic information, but they have insufficient power to predict patient outcome precisely. The need for accurate predictive markers has led to the assessment of a variety of molecular markers in patients with bladder cancer. However, the sensitivity and specificity of these markers have been limited and different studies have produced conflicting results as to their usefulness in predicting disease outcome (8). Gene expression profiling has been reported for the molecular classification of many tumor types, identifying molecular subtypes with potential diagnostic and prognostic significance (9–14). This technology has the promise to define specific markers and therapeutic targets that may lead to individualized patient treatment. To date, the number of published studies on bladder cancer using microarray expression profiling has been rather limited and only a few publications have involved clinical material (12, 15, 16). These studies have identified gene signatures that were associated with tumor stage (16, 17) and superficial recurrence (16). However, the number of muscle-invasive tumors reported in these studies is small, and robust outcome classification models for the invasive tumors were not feasible.

Authors' Affiliations: [1]Comprehensive Cancer Center and [2]Departments of Pathology, [3]Urology, and [4]Laboratory Medicine, University of California at San Francisco, San Francisco, California

Received 11/23/04; revised 2/3/05; accepted 3/4/05.

Grant support: NIH grant CA089715.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Frederic M. Waldman, Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA 94143-0808. Phone: 415-476-3821; Fax: 415-476-8218; E-mail: waldman@cc.ucsf.edu.

The current study is the first to report a set of genes predictive of overall survival in a large cohort of patients with muscle-invasive disease. Studies like this show that the combination of several genes can be more powerful in predicting outcome than the use of individual markers alone. More studies are necessary in validating and fine-tuning the existing predictive gene sets to achieve personalized care in cancer.

Materials and Methods

Samples and RNA preparation. Freshly frozen bladder tumors were obtained from the tissue bank of the University of California at San Francisco Comprehensive Cancer Center in accordance with institutional guidelines on the use of human tissue. For each case, all blocks from the surgical specimen were reviewed to determine pathologic grade, stage, and histologic tumor type. Tumors were staged according to the American Joint Committee on Cancer (18) and graded according to the WHO and International Society for Urological Pathology classification system (19). Histologic type was according to WHO criteria, with pure features of squamous histology necessary for the diagnosis of SCC.

Before RNA extraction, an initial H&E-stained frozen section was reviewed to assess tumor quality and content. Normal and necrotic tissues were excluded by trimming of the frozen block. A tumor sample was considered suitable for study if the proportion of tumor cells was >70%. Every 20 sections, a section was stained by H&E to reevaluate the cellular composition. Twenty to 40 frozen 50 μm thick sections (depending on the size of the tumor) were placed directly in Trizol reagent (Invitrogen, Carlsbad, CA), and total RNA was extracted. RNA was further purified using the RNeasy Mini protocol (Qiagen, Valencia, CA). The quality of RNA from each sample was verified by A_{260}/A_{280} ratio and the integrity of 28S and 18S RNA by agarose gel electrophoresis.

A mixture of equal amounts of RNA from the following 11 cell lines, all originally obtained from American Type Culture Collection (Manassas, VA), was used as a common hybridization reference pool sample: SW872, WM115, NTERA2, MCF7, HEPG2, MOLT4, Hs578t, HL60, OVCAR3, COLO205, and RPMI8226. The same preparation of reference pool RNA was used for all hybridizations. Total RNA was extracted as described above. Three additional total RNA samples from normal bladder were obtained from commercial sources (Clontech, Palo Alto, CA; Chemicon, Temecula, CA; and Stratagene, La Jolla, CA). RNA from nine bladder cancer cell lines (RT4, J82, 5637, HT1376, HT1197, T24, UMUC-3, TCCSUP, and SCaBER) was extracted as described above.

Expression microarray analysis. Custom cDNA microarrays of 10,368 elements were prepared using standard procedures as described previously (20). cDNA clones were obtained from Research Genetics (Huntsville, AL) and were printed onto poly-L-lysine-coated glass slides. Total RNA (10 μg) was used as input for reverse transcription using random hexamers and oligo(dT) primers. Labeling of the amine-modified cDNA from tumors and reference cell line pool was done with Cy3 and Cy5 dyes, respectively. Hybridization was overnight, and washing and imaging were as described (20).

Data analysis. Links to all primary data are available at http://cc.ucsf.edu/people/waldman/Blaveri.html. Signal intensity values of each element were extracted using the GenePix version 3.06 software program (Axon Instruments, Union City, CA). Intensities were corrected by subarray (print tip) median centering followed by global locally weighed scatter plot smoothing correction (21) implemented by the Bioconductor R software package. Before clustering, the locally weighed scatter plot smoothing and median subarray corrected $\log_2$ ratios were filtered to include only these clones showing expression ratios in at least 80% of samples analyzed. Additionally, only clones that had a $\log_2$ ratio of less than -2 or more than 2 in at least one sample were used. Data were normalized and median centered for both genes and arrays using the software program Cluster. A two-way unsupervised hierarchical clustering was done using average linkage and an uncentered Pearson correlation metric. The results were visualized in the software program TreeView (22).

Significance analysis of microarrays (SAM; ref. 23) was used to examine differential expression patterns between subgroups. This method uses moderated t statistic as a gene score. False discovery rate (FDR) or expected proportion of false calls among genes identified as differentially expressed is estimated using a permutation approach based on a user defined threshold (23). In this study, we chose a threshold that would result in one (or less) false-positive gene among the genes selected as differentially expressed in each of the comparisons. A 10-nearest neighbor imputation engine was applied to estimate missing data, and 1,000 permutations were carried out to estimate FDR.

The Westfall and Young method, also called maxT, was used to identify differentially expressed genes as implemented by the mt.maxT function in the Multtest package of Bioconductor R software. For two-class unpaired comparisons, such as ours, a t statistic is used as a test statistic and its adjusted P is computed using step-down multiple testing procedures, thus providing strong control of the family-wise type I error rate. Briefly, the class labels are repeatedly permuted and the test statistic for each gene is calculated. The maximum test statistic over all is recorded for 10,000 random permutations and the P for each gene is estimated as the proportion of the maximum permutation based statistics exceeding the observed value.

Prediction analysis for microarrays (PAM) was used to identify subsets of genes that best characterized each bladder tumor class (24) and to test classification rates for these gene sets. PAM uses a modified version of the nearest centroids classification method, which "shrinks" the centroids by means of soft thresholding. Ten-fold cross-validation was used to choose an optimum gene number (threshold), which minimized classification errors. Missing values were estimated across the data set using a 10-nearest neighbor impute engine. All genes identified in the predictive sets were sequence verified. PAM is not a perfect cross-validation procedure because it uses the entire tumor set to define the optimum threshold before classification. However, it is established in the literature as a robust procedure for testing expression-derived gene sets for the classification strength.

Kaplan-Meier survival curves were generated using the survival package within Bioconductor R.

Bladder cancer data from an independent tumor set. Expression profiles of 40 bladder tumors (20 pT_a, 11 pT_1, and 9 pT_2-pT_4) were generated by Dyrskjot et al. (16) using an Affymetrix HuGeneFL platform. Microarray results were scaled to a global intensity of 150 units using the Affymetrix GeneChip software. Expression level ratios between tumors and a reference pool of 37 normal urothelial samples were calculated using the comparison analysis implemented in the Affymetrix GeneChip software as described previously (21). The log-transformed expression level ratios for this data set (accession no. GSE89) were retrieved from the Gene Expression Omnibus (National Center for Biotechnology Information).

Immunohistochemical analysis. Immunohistochemistry was done on a bladder-specific tissue microarray containing in total 135 cores of 0.6 mm diameter. These consisted of 89 bladder tumors of different stages (19 pT_a, 7 pT_1, and 63 pT_2-pT_4), grades (17 low grade and 72 high grade), and histologies (6 SCCs and 83 TCCs), 9 bladder cancer cell lines, 14 normal bladder tissues, 15 nonbladder control tissues, and 8 control cancer cell lines. The construction of the tissue microarrays was done using a tissue array (Beecher Instruments, Silver Spring, MD) as described by Kallioniemi et al. (25).

Tissue microarray sections were stained using standard avidin-biotin complex/peroxidase methods. The antibodies used were specific for cathepsin E (CTSE; 1:50 dilution, Santa Cruz Biotechnology, Santa Cruz, CA), cyclin A2 (CCNA2; 1:50 dilution, Santa Cruz Biotechnology), and parathyroid hormone–related protein (PTHrP; 1:250 dilution, Biogenesis, Kingston, NH). Slides were counterstained with hematoxylin before scoring. Percentage of tumor cells positive was scored in deciles, and intensity of staining was scored as 0 to 3+. For

CTSE and CCNA2, positive staining was defined as at least 2+ intensity in at least 10% of the tumor cells of each core specimen, whereas for PTHrP it was defined as at least 1+ intensity in 40% of tumor cells. Scores were recorded for tumor cell membranous, cytoplasmic, and nuclear compartments as well as for stromal staining. Normal urothelium was used as a negative control for all antibodies used. Positive controls were normal tonsil, colorectal adenocarcinoma, and pancreatic adenocarcinoma for CTSE, normal testis and invasive breast tumor for CCNA2, and normal human skin and MCF7 breast cancer cell line for PTHrP. Staining was cytoplasmic for CTSE and PTHrP and nuclear for CCNA2.

Results

Clinicopathologic characteristics. A total of 80 bladder tumors were analyzed, including 17 stage pT_a (12 low grade and 5 high grade), 10 pT_1 (all high grade), 15 pT_2, 25 pT_3, and 13 pT_4 tumors. Transurethral resection was used to sample all stage pT_a tumors and 8 of the 10 pT_1 tumors, whereas cystectomy was done for 48 of the 53 stage pT_2 or higher tumors (Supplementary Data[5]: Clinical.xls). Seventy-four tumors were TCC and 6 were pure SCC (Table 1). The median age for all patients was 66 years. Outcome information was available for the 47 patients with muscle-invasive disease (stage pT_2 or higher). Fourteen of these patients survived with a median follow-up of 46.5 months, whereas 33 patients died with a median survival of 8.8 months. A listing of clinicopathologic information can be found in Supplementary Data (Clinical.xls).

Unsupervised hierarchical clustering of tumors and genes. Gene expression profiles were obtained by hybridization to cDNA arrays containing 10,368 human gene elements; 2,675 clones remained after data filtering (Supplementary Data: log_2ratios.xls). Unsupervised hierarchical clustering was used to group tumors with similar repertoires of expressed genes without knowledge of tumor class. Two main clusters of tumors resulted from this process, showing a strong association with stage (Fig. 1). One cluster contained 26 of the 27 superficial tumors (17 of 17 pT_a and 9 of 10 pT_1) and the second cluster contained 48 of 53 muscle-invasive tumors (11 of 15 pT_2, 24 of 25 pT_3, and 13 of 13 pT_4), resulting in a 92.5% correct assignment. There were six SCCs included in the analysis, all muscle invasive, and the hierarchical clustering placed these within the muscle-invasive cluster. All six SCC tumors were within the same subcluster.

Unsupervised clustering of the 27 superficial tumors alone resulted in their separation into two clusters (Supplementary Data: Unsupervised.superficial). One group contained mainly low-grade pT_a tumors (8 low grade and 1 high grade), whereas the other contained all the pT_1 tumors (10 of 10; all pT_1 tumors were high grade) as well as 8 pT_a tumors (4 high grade and 4 low grade). Similarly, when only the muscle-invasive tumors were tested by hierarchical clustering, three main groups of tumors were formed, one of which contained all the SCC cases (Supplementary Data: Unsupervised.muscle-invasive).

The expression profiles of nine bladder cancer cell lines were also determined (Supplementary Data: Unsupervised.celllines.-tumors). When these 9 cell lines were included in the hierarchical clustering with the 80 primary tumors, 8 of them formed a subcluster together with 5 muscle-invasive tumors but distinct from the remaining superficial and muscle-invasive tumors. Cell line RT4, which is derived from a low-grade papillary tumor, clustered separately from the other cell lines, within the main cluster of superficial tumors. The three normal bladder samples clustered adjacent to each other and formed a subgroup with four tumors of mixed stage (Supplementary Data: Unsupervised.normals.tumors).

In addition to tumor clusters, unsupervised hierarchical clustering simultaneously defines gene clusters whose expression levels vary in a similar fashion across tumors. Several characteristic gene clusters were identified which contained several functionally related genes (Fig. 2). The extracellular matrix cluster (cluster F) contained genes encoding proteins with roles in the synthesis or modification of the extracellular matrix as well as genes involved in matrix remodeling and angiogenesis. The immune-cell cluster (cluster E) included immune response genes, such as genes encoding members of the class I and II MHC, complement components, immunoglobulins, and chemokines. The genes in these clusters (F and E) had high expression levels in muscle-invasive tumors. The squamous cluster (cluster D) contained genes involved in inflammation and keratinization and showed elevated expression in the SCC tumors. Cell cycle–related genes formed the proliferation cluster (cluster C); they were highly expressed in the cell lines and were under expressed in the low-grade pT_a tumors. Finally, there were two related gene clusters that were highly expressed in the superficial tumors, the up in superficials (cluster B) and the ribosomal cluster (cluster A), containing transcription factors and nuclear proteins and containing ribosomal genes, respectively.

Table 1. Clinicopathologic characteristics of tumor samples

	n (%)
Histologic type	
Transitional	74 (93)
Squamous*	6 (7)
Stage	
pT_a	17 (21)
pT_1	10 (13)
pT_2	15 (19)
pT_3	25 (31)
pT_4	13 (16)
Gender	
Male	56 (70)
Female	24 (30)
Lymph node status	
Negative	29 (36)
Positive	17 (21)
Unknown	34 (43)
Clinical outcome†	
Alive (≥118 mo)	13 (32.5)
Dead (<18 mo)	27 (67.5)

*All squamous cell carcinomas were muscle invasive (pT_2-pT_4).
†Muscle-invasive tumors only.

[5] Supplementary Data can be found at http://cc.ucsf.edu/people/waldman/Blaveri.html.

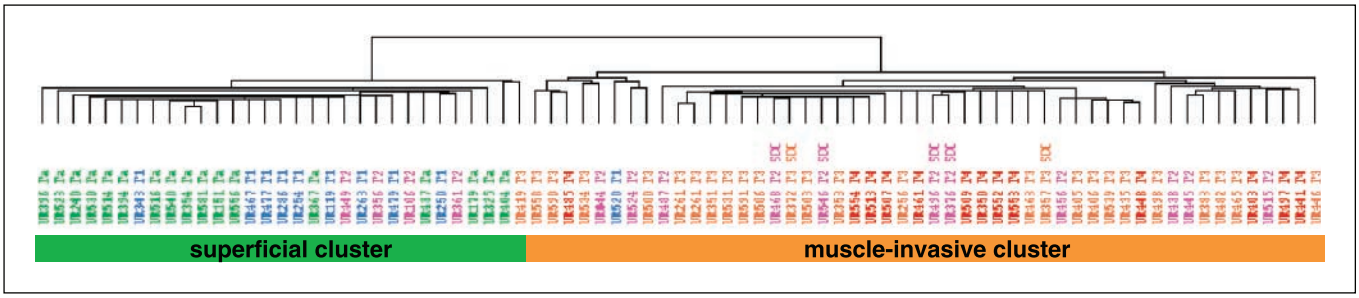


Fig. 1. Unsupervised clustering of 80 bladder tumors. The 2,675 cDNA clones remaining after filtering were used as input. Tumors are colored by stage: pT_a (*green*), pT_1 (*blue*), pT_2 (*pink*), pT_3 (*orange*), and pT_4 (*red*). Note the clear separation of superficial versus muscle-invasive tumors. SCC tumors are indicated, whereas the rest of the tumors are TCCs.

Supervised classification according to bladder tumor stage. The PAM method of classification was used to test whether expression patterns were able to classify the 74 TCCs into superficial versus muscle-invasive subgroups. Twenty-seven superficial tumors (17 pT_a and 10 pT_1) were compared with 47 muscle-invasive tumors (pT_2-pT_4). PAM classified 25 of 27 superficial cases and 42 of 47 muscle-invasive cases correctly for an overall success rate of 90.5%. The classifier was composed of 25 unique genes, with the top predictive gene being *sulfatase 1* (*SULF1*; Table 2).

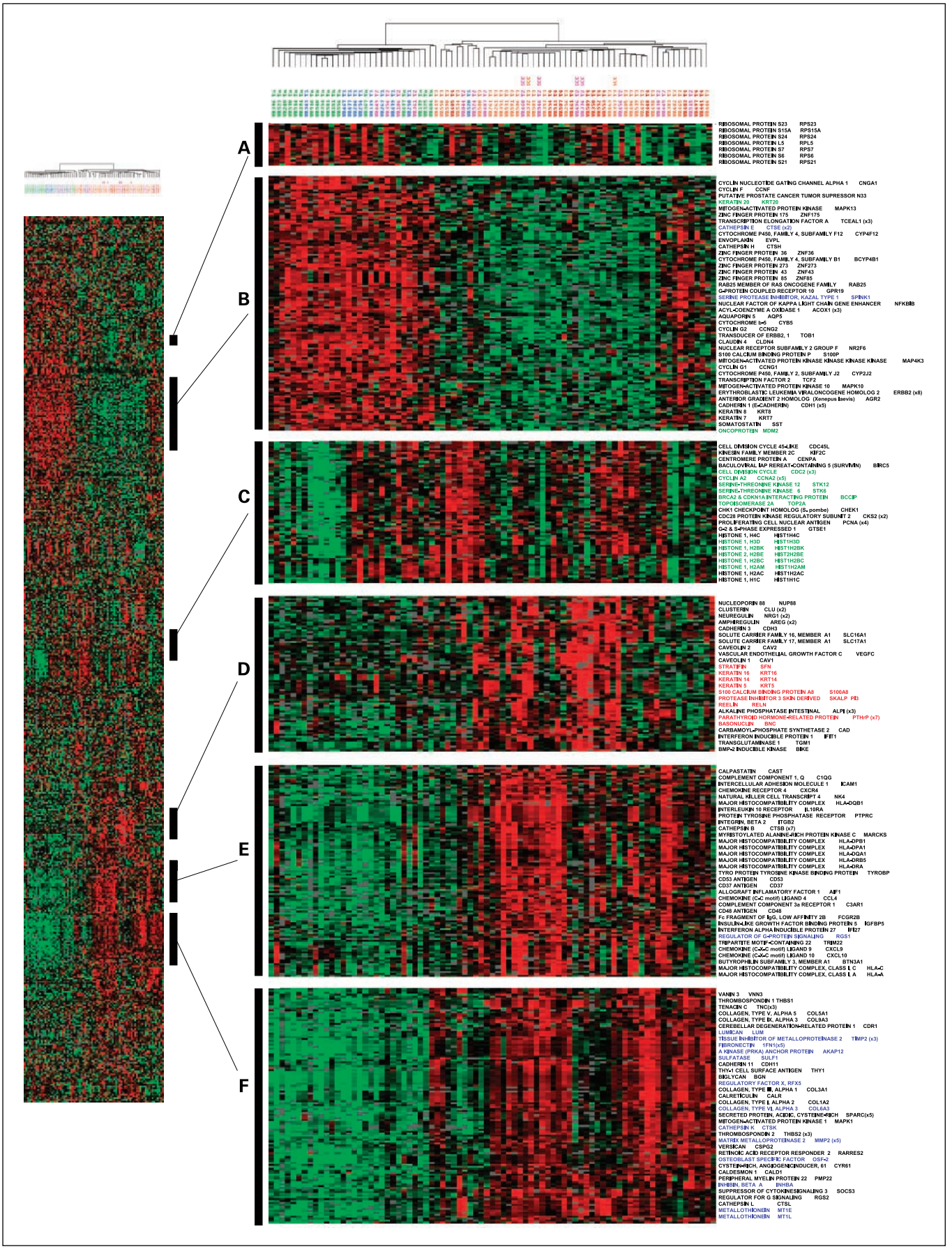
Because PAM analysis does not test statistical significance of differentially expressed genes but rather identifies genes that best classify two groups of samples, each gene was further tested for differential expression using maxT and SAM as described in Materials and Methods. Both methods provide adjusted *P*s corrected for multiple comparisons using permutation approaches. There were 272 cDNA clones differentially expressed between superficial and muscle-invasive tumors with a maxT adjusted *P* of 0.001. SAM identified 378 cDNA clones highly expressed in muscle-invasive tumors and 248 cDNA clones highly expressed in the superficial tumors, with a FDR of 0.05% (i.e., we expected 0.3 false positives among the 626 cDNAs). Many genes were identified by both analyses, including 174 cDNA clones expressed at higher levels in the muscle-invasive tumors and 94 cDNA clones expressed at higher levels in the superficial tumors (Supplementary Data: CommoninSAMmaxT.xls). All of the 25 genes used in the PAM classifier were identified by both adjustments as highly significant. *CTSE* was the most highly expressed gene in the superficial tumors, and *regulator of G-protein signaling 1* (*RGS1*) showed the highest significant level of expression in muscle-invasive tumors.

PAM was also used to classify superficial tumors into stage pT_a versus stage pT_1 subgroups. A classification success rate of 70% was obtained using a classifier composed of 54 unique genes (Supplementary Data: ClassifierTavsT1.xls). SAM identified 19 significant clones differentially expressed between pT_a and pT_1 tumors with a FDR of 4.5% (i.e., 0.8 false positive among the 19 clones). Genes expressed at higher levels in the pT_a tumors included *IFN-γ-inducible protein 16* (*IFI16*), *caspase-1* (*CASP1*), and hypothetical gene *BC008967*, whereas genes showing higher levels in the pT_1 tumors included those involved in the control of cell cycle regulation, such as *CCNA2*, *CDC6*, *CDC2*, and *ubiquitin-conjugating enzyme E2C* (*UBE2C*). maxT analysis identified only one transcript (hypothetical gene *BC008967*) that was differentially expressed between pT_a and pT_1 tumors after multiple test correction, with a *P* of 0.05.

Supervised analysis of transitional cell carcinomas versus squamous cell carcinomas. The PAM method of supervised classification was used to subgroup tumors as transitional versus squamous types. The 6 SCC cases were all muscle invasive and thus were analyzed with the muscle-invasive TCC cases (47 total) to have a comparable stage distribution. PAM classified 5 of 6 SCC and 42 of 47 TCC samples correctly for an overall success rate of 89%. The optimum classifier consisted of 40 cDNA clones (30 unique genes; Table 3). The top gene that separated the histologic types was *PTHrP* and was represented by seven replicate clones in the classifier. SAM analysis identified 28 genes as significantly differentially expressed with a FDR of 2.8% (i.e., 0.8 false positive among the 28 genes), all of which also overlapped with the PAM classifier. maxT permutation analysis identified only three genes as differentially expressed, with an adjusted *P* of <0.05: *PTHrP*, *calcitonin receptor-like* (*CALCRL*), and *carbonic anhydrase II* (*CA2*). Again, the gene that was most associated with histologic type was *PTHrP* and was identified by all three supervised methods of analysis.

TCC tumors with features of squamous differentiation are generally not considered a separate diagnostic category. They are considered TCCs, with the degree of squamous differentiation reported. In this study, there were 22 tumors with varying levels of squamous differentiation of the 47 TCC muscle-invasive tumors. PAM analysis was used to predict whether these cases classified more closely as SCC versus TCC. A classifier was generated using as training set the 6 "pure" SCC and the 25 "pure" TCC tumors (with no squamous differentiation present). This classifier showed an overall success rate of 90% and contained 15 cDNA clones (9 unique genes). This gene set was then used to predict the best class (SCC versus TCC) for the 22 tumors, which were histologically defined as TCCs with varying degrees of squamous differentiation. The gene set classified 7 of these 22 tumors as SCC and the remaining 15 as TCC. The percentage of squamous differentiation was not associated with the classification into squamous versus transitional groups.

Outcome prediction. For outcome prediction, subsets of TCC muscle-invasive patients were separated into good prognosis and bad prognosis groups. Good prognosis was defined as survival with follow-up of ≥18 months (n = 13). Bad prognosis was defined as death occurring in <18 months (n = 27). The remaining seven muscle-invasive patients included one who was alive for <18 months and six who died >18 months after diagnosis. In this way, generation of our classification pattern was based on the extreme survival groups. There were no significant differences in the distribution of stage or nodal



status between good and bad prognosis groups. Seven muscle-invasive patients did not fall into either category.

PAM analysis had a 78% success rate for classification of tumors into good and bad prognosis groups based on 25 clones (24 unique genes; Table 4). Kaplan-Meier survival analyses based on this classification showed strong significance ($P < 0.006$; Supplementary Data: OutcomeSupervised). SAM analysis identified seven genes as significantly differentially expressed between the two outcome groups (FDR, 13.5%; i.e., 0.9 false positive among the seven genes).

As described above, unsupervised hierarchical clustering of the 47 muscle-invasive tumors resulted in their separation into three distinct tumor subclusters (Supplementary Data: Unsupervised. muscle-invasive and OutcomeUnsupervised). One of these subclusters contained tumors from 11 patients who all died with very short overall survival (median, 5.4 months), whereas the remaining two subclusters contained a mixture of both short and long survival tumors (median, 16.7 and 19.3 months). Kaplan-Meier survival analysis showed significant association with outcome when done based on these tumor groups ($P < 0.0001$).

Assessment of the classifiers on an independent data set. To evaluate the performance of our stage classifiers, we analyzed the only publicly available independent expression data set of 40 bladder tumors (20 pT_a, 11 pT_1, and 9 pT_2-pT_4) generated by Dyrskjot et al. (16). We first tested our 25-gene classifier that separated our tumors into superficial versus muscle-invasive subgroups (90% success rate). Twenty-four of these 25 genes were present on the Affymetrix HuGeneFL platform used in the Dyrskjot et al. study. Using this 24-gene set, a 10-fold cross-validation was done by PAM. An 82.5% success rate was obtained in classifying the tumors in this independent set into superficial and muscle-invasive groups.

We also evaluated the performance of our 54-gene classifier that separated our superficial tumors into stage pT_a versus pT_1 (70% success rate). In this case, 42 of these 54 genes were present on the Affymetrix HuGeneFL arrays. We therefore assessed this 42-gene set on the 31 superficial tumors available from the Dyrskjot et al. study. An 87% classification success was obtained for separating these tumors into stage pT_a versus pT_1 using PAM. Classifiers for predicting histology (SCC versus TCC) and outcome could not be assessed on additional independent data sets because such data sets are not available from the literature.

Finally, Dyrskjot et al. identified a 32-gene classifier that showed the largest possible separation among pT_a, pT_1, and muscle-invasive tumors. Interestingly, there was no overlap between this 32-gene classifier and our 25-gene stage classifier. Twenty of their 32 genes were present on our cDNA array and were used for stage prediction using our 74 TCC tumor set. An 88% classification success was obtained for separating these tumors into superficial versus muscle-invasive groups.

Protein expression by immunohistochemistry. Immunohistochemistry was carried out with antibodies against three proteins (CTSE, CCNA2, and PTHrP) on a tissue microarray containing 89 bladder tumor cores of different histology, stage, and grade (Fig. 3). These three genes were among the top predictors in three of the PAM classifiers, and specific antibodies were commercially available. A total of 49 of these tumors were also analyzed by the cDNA arrays and were used to compare our findings from RNA expression analysis with protein expression data.

CTSE was part of our classifier that successfully separated superficial from muscle-invasive tumors and was also identified as the gene with the most significantly high RNA expression

Table 2. PAM classifier of muscle-invasive (pT_2-pT_4) versus superficial (pT_a and pT_1) tumors

Name*,†	Symbol	Genbank ID	Fold change‡
Sulfatase 1	*SULF1*	AI653116	5.5
Osteoblast-specific factor 2	*OSF2*	AA598653	7.6
Fibronectin 1	*FN1*	R62612	7.2
Metallothionein 1L	*MT1L*	N80129	7.8
Nuclear transport factor 2	*NUTF2*	AI289110	6.6
Regulator of G-protein signaling 1	*RGS1*	AA017544	7.9
Metallothionein 1E	*MT1E*	AA872383	7.5
Milk fat globule-epidermal growth factor 8 protein	*MFGE8*	AA449667	7.5
Inhibin, β A inhibin, β A	*INHBA*	N27159	6.2
Fibrinogen B, β polypeptide	*FGB*	NM005141	4.4
Stress-induced phosphoprotein 1	*ST1P1*	M86752	6.2
A kinase anchor protein (gravin) 12	*AKAP12*	AA478543	5.2
Tissue inhibitor of metalloproteinase 2	*TIMP2*	AA486280	3.5
Complement component 1S	*C1S*	T62048	6.1
Serine protease inhibitor, Kazal type 1	*SPINK1*	AA845156	0.2
Acyl-CoA dehydrogenase A	*CADSB*	NM001609	5.4
Regulatory factor X, 5	*RFX5*	AA418045	5.1
Proteoglycan 1, secretory granule	*PRG1*	A278759	4.5
Collagen, type I, α2	*COL1A2*	AA490172	5.3
Metallothionein 1F	*MT1F*	N55459	5.0
Metallothionein 1H	*MT1H*	H77597	7.0
Lumican	*LUM*	AA447781	4.8
Collagen, type VI, α3	*COL6A3*	R62603	3.5
Cathepsin K	*CTSK*	R00859	5.3
Cathepsin E	*CTSE*	H94487	0.1

*Genes are sorted according to the best prediction between classes.
†All genes have been sequence verified.
‡The fold change of muscle-invasive to superficial expression.

Fig. 2. Two-way hierarchical cluster of 80 bladder tumors. Thumbnail overview using 2,675 cDNA clones remaining after filtering (*left*). Rows, individual genes; columns, experimental samples. High expression levels are in red and low expression levels in green after centering across genes and samples; gray, missing values. Enlarged views of the selected gene expression clusters (*right*): ribosomal cluster (*A*), up in superficials cluster (*B*), proliferation cluster (*C*), squamous cluster (*D*), immune-cell cluster (*E*), and extracellular matrix cluster (*F*). Because of space limitations, only selected genes of these clusters are indicated. Genes belonging to the PAM classifiers are indicated in blue (superficial versus muscle-invasive), green (pT_a versus pT_1), and red (SCC versus TCC).

Table 3. PAM classifier of SCC versus TCC

Name*,†	Symbol	Genbank ID	Fold change‡
Parathyroid hormone – related protein§	*PTHrP*	AA845432	7.4
Matrix metalloproteinase 3§	*MMP3*	W51794	6.7
Galectin 7	*LGALS7*	AA011057	4.0
Carbonic anhydrase II	*CA2*	H23187	3.3
Claudin 4	*CLDN4*	AA430665	0.2
S100 calcium binding protein A8§	*S100A8*	AA086471	4.6
Keratin 16§	*KRT16*	AA928454	3.7
Tumor-associated calcium signal transducer 1	*TACSTD1*	AI340883	0.3
ELL-related RNA polymerase II, elongation factor	*ELL2*	AA884897	2.4
Keratin 14§	*KRT14*	H44051	4.3
Keratin 5§	*KRT5*	AA160507	3.7
Basonuclin	*BCN*	R26526	3.6
Desmocollin 2§	*DSC2*	AA075299	2.5
v-maf musculoaponeurotic fibrosarcoma oncogene	*MAF*	AA043501	2.6
Stratifin	*SFN*	AI356363	4.1
Carbamoyl phosphate synthetase	*CAD*	R85414	2.3
GATA 3 binding protein	*GATA3*	H72875	0.2
S100 calcium binding protein A7	*S100A7*	AA583574	3.4
Protease inhibitor 3, skin derived	*PI3 (SKALP)*	AI582329	6.3
Amyloid β (A4) precursor protein	*APBA1*	H19687	2.4
LIM domain binding 3	*LDB*	AI383171	2.2
MAX dimerization protein 1	*MAD*	H86558	2.1
Solute carrier family 9	*SLC9A1*	AA455369	0.3
S100 calcium binding protein P	*S100P*	R32848	0.2
KIAA0089 protein	*KIAA0089*	AA485401	0.3
Interleukin-13 receptor, α2	*IL13RA2*	R52796	3.6
v-erb-b2 oncogene homologue 2	*ERBB2*	AA443351	0.3
Diphtheria toxin receptor	*DTR*	R45640	2.0
Collagen, type XVII, α1§	*COL17A1*	H87536	2.7
Purine-rich element binding protein A	*PURA*	H46425	2.9

*Genes are sorted according to the best prediction between classes.
†All genes have been sequence verified.
‡The fold change of SCC to TCC expression.
§Gene belonging in the classifier when considering only SCC and TCC cases with "pure" histologic features.

among the superficial tumors. Immunohistochemical analysis showed that 65% of the superficial tumors had positive staining (≥2 in ≥10% of cells) compared with only 12% of the muscle-invasive tumors ($P < 0.001$, $n = 81$). Interestingly, $\log_2$ ratios of RNA expression correlated with immunohistochemistry scores for *CTSE* ($P < 0.001$; $n = 44$).

CCNA2 was identified as differentially expressed between high-grade versus low-grade superficial tumors. Immunohistochemistry showed a similar pattern in protein expression, with 43% of the low-grade tumors and 89% of the high-grade tumors showing positive staining for CCNA2 ($P < 0.05$; $n = 13$). However, in contrast to CTSE, $\log_2$ ratios of RNA expression were not significantly correlated with immunohistochemistry scores for CCNA2 ($P > 0.5$; $n = 13$). *CCNA2* was also part of the classifier for stage pT_a versus pT_1 tumors. However, CCNA2 protein expression levels did not show a

Table 4. PAM classifier for outcome prediction of good versus bad prognosis

Name*,†	Symbol	Genbank ID	Fold change‡
Protein kinase, AMP activated, γ1	*PRKAG1*	AA070495	4.1
v-myc homologue 1	*MYCL1*	R62813	5.4
Serine protease inhibitor, Kazal type 1	*SPINK1*	AA845156	10.1
mal, T-cell differentiation protein	*MAL*	AA227885	3.0
Transmembrane protease, serine 2	*TMPRSS2*	AI261741	4.3
Granulysin	*GNLY*	AA701652	2.4
Hypothetical protein FLJ20425	*LYAR*	AA063624	0.5
Ribosomal protein L3	*RPL3*	T67053	1.8
N-acylsphingosine amidohydrolase 1	*ASAH1*	AA664155	2.1
Retinoic acid receptor responder 3	*RARRES3*	AA449667	2.3
Regulator of G-protein signaling 16	*RGS16*	AA453774	2.4
Amphiregulin	*AREG*	AA857163	0.3
TONDU	*TONDU*	AA700322	2.8
ALL1-fused gene from chromosome 1q	*AF1Q*	AA456008	0.3
Glycine amidinotransferase	*GATM*	R61229	2.7
Vascular cell adhesion molecule 1	*VCAM1*	AU121762	1.6
Stearoyl-CoA desaturase	*SCD*	AA457700	1.8
CASP8 and FADD-like apoptosis regulator	*CFLAR*	T62048	1.8
Dual-specificity phosphatase 6	*DUSP6*	AA455254	0.5
Fibulin 1	*FBLN1*	AA134871	1.8
Acyl-CoA dehydrogenase A	*ACOX1*	NM001609	2.1
Integrin, β_5	*ITGB5*	AA434397	0.6
msh homeobox homologue 2	*MSX2*	AA195636	1.8
Chemokine binding protein 2	*CCBP2*	R12708	1.8

*Genes are sorted according to the best prediction between classes.
†All genes have been sequence verified.
‡The fold change of expression in good prognosis versus bad prognosis tumors.

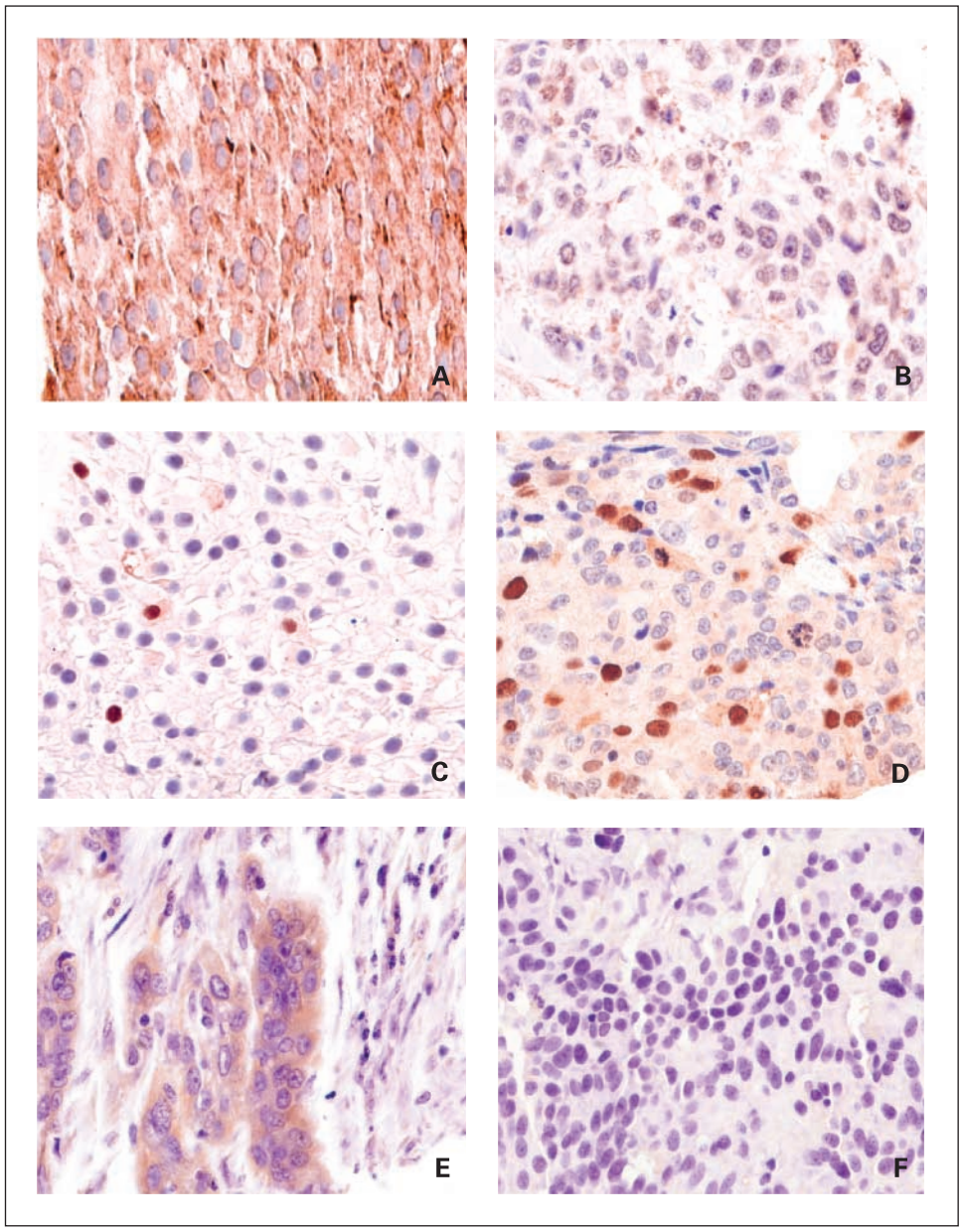


Fig. 3. Immunohistochemical staining of tissue cores from a bladder cancer tissue microarray for CTSE, CCNA2, and PTHrP. CTSE shows strong staining in a superficial tumor (*A*) and very weak expression in a muscle-invasive tumor (*B*). CCNA2 shows low-frequency nuclear expression in a low-grade superficial tumor (*C*) and high expression in a high-grade superficial tumor (*D*). PTHrP shows moderate staining in SCC (*E*) and absent staining in TCC (*F*). Original magnification, ×400.

significant difference between these two groups of tumors by immunohistochemistry on our tissue array.

Finally, *PTHrP* was the top predictive gene in the classifier differentiating between SCC and TCC. Although the number of samples with squamous histology on our tissue array was small, 4 of the total of 6 cores (67%) stained positively for the PTHrP protein, whereas only 12 of 54 (22%) muscle-invasive tumors of transitional histology stained positively ($P < 0.05$; $n = 60$). Log_2 ratios of RNA expression again did not show significant correlation with immunohistochemistry scores for this protein ($P > 0.4$; $n = 38$).

Discussion

Bladder cancer consists of a heterogeneous group of tumors that follow different pathways of development and progression due to discrete genetic alterations. Further elucidation of the course of tumor progression and the separation of tumors into distinct, clinically relevant groups should have a significant effect on the management of this disease. To this end, we have studied expression patterns in tumors of different stages, grades, and histologic types.

Both unsupervised (hierarchical clustering) and supervised (PAM) analyses were used, because different methods often highlight different patterns in the data. Hierarchical clustering separated tumors into two clearly distinct groups, superficial and muscle invasive, similar to previous reports (12, 13, 16, 17). When a set of nine bladder tumor cell lines was included in the unsupervised clustering, eight of nine (all derived from advanced tumors) were grouped in a separate subcluster with five muscle-invasive tumors. This subcluster showed lower expression of stromal genes than the other invasive tumors and higher expression of proliferation-associated genes. In contrast, the RT4 cell line, derived from

a superficial tumor, clustered with the superficial tumors. These results suggest that the cell lines maintain patterns of gene expression, which represent the stage of progression of their tumors of origin.

The PAM method of supervised classification correctly predicted the stage of the TCC with a 90.5% success rate based on the expression signature of only 25 genes. Four of the tumors misclassified by PAM analysis were also "misplaced" in the hierarchical tumor tree, providing further support for the robustness of both analytic approaches. It is somewhat surprising that supervised analysis, in which the tumor classes are preassigned, did not outperform unsupervised hierarchical clustering in classification accuracy. This may have been because the superficial and muscle-invasive tumor groups are so distinct from each other. In such cases, where tumors comprise very distinct biological and molecular entities, both methods may be equally powerful in distinguishing tumor subtypes. Four of the five misclassified muscle-invasive cases by supervised analysis were stage pT_2 tumors and thus may have shared expression pattern with the high-grade superficial tumors. For the pT_1 misclassified case, the transurethral biopsy specimen did not contain muscularis, so the extent of invasion could not be fully assessed, raising the possibility that this was a muscle-invasive tumor.

Supervised PAM analysis of superficial versus muscle-invasive tumors identified a panel of 25 genes, all of which were statistically significant by both maxT and SAM analyses. The majority of these genes belonged to the extracellular matrix cluster. This may be due to differences in the type or quantity of normal cell elements in the tumor area. However, this finding highlights the dynamic nature of the tumor-stroma interaction, with epithelial tumor cells differentially modulating stromal components and stromal cells supporting tumor growth and invasion. In the current study, the expression of stromal-related genes was seen in higher-stage tumors. Although it is not clear for many of these genes whether they are produced by stromal or epithelial components, their expression patterns across these tumors are central to the classification of our tumor set into either superficial or muscle-invasive subgroups. Such differences may be related to the degree of stromal response in the different tumor stages and may also provide valuable information about previously unrecognized relationships among tumor cells and their microenvironment and may be developed as markers for monitoring disease progression.

The most predictive gene in the superficial versus muscle-invasive classifier was *SULF1*, which was highly expressed in muscle-invasive tumors. This gene is an extracellular endosulfatase that diminishes sulfation of cell surface heparan sulfate proteoglycans (HSPG). The signaling functions of cell surface HSPGs are dependent on their sulfation states; therefore, SULF1 may provide a mechanism to regulate such signaling responses (26). The avian orthologue of *SULF1* is thought to initiate Wnt signaling by this process (27). Similarly, increased *SULF1* in bladder tumors may also lead to activation of the Wnt signaling pathway (28). Similar to Wnt, the modulation of the binding affinity of other growth factors, cytokines, and differentiation factors to HSPGs by increased levels of SULF1 may be initiating signaling cascades promoting cellular proliferation and repression of differentiation in the muscle-invasive tumors.

Muscle-invasive tumors also expressed high levels of the lysosomal cysteine proteinases cathepsins B, K, and L. Cathepsin B has been implicated in bladder tumor invasion (29) and correlated strongly to both tumor grade and stage (30–32). In addition to directly degrading extracellular matrix components, cathepsin B is thought to play an active role in the initiation of the proteolytic cascade that involves urokinase-type plasminogen activator, matrix metalloproteinase (MMP) 2, 3, and 9, and plasmin. Both urokinase-type plasminogen activator and MMP2 have been associated with bladder cancer invasion (33, 34). *MMP2* showed increased expression in the muscle-invasive tumors in our study similar to *tenascin C* (*TNC*), which is also downstream of the cathepsin B proteolytic cascade and has been reported as highly up-regulated during tumor invasion in bladder cancer (35).

In contrast to *cathepsins B, K, and L*, *CTSE* showed higher expression levels in superficial tumors, especially pT_a tumors. CTSE is a nonsecretory intracellular, but nonlysosomal, aspartic proteinase. Expression of this protein has been reported in pancreatic cancer (36) as well as in preneoplastic lesions of the pancreas (37) and stomach (38). Cathepsins may play diverse roles in bladder cancer, with some cathepsins (B, K, and L) promoting muscle invasion, whereas other cathepsins (E and H) may be active in superficial tumors.

Genes involved in the promotion of angiogenesis were expressed at high levels in the muscle-invasive tumors. Two such genes, *RGS1* and *RGS2*, have been reported to play a role in the development and function of heart and blood vessels (39). They showed increased expression in muscle-invasive tumors, and *RGS1* also belonged to the classifier panel for superficial versus muscle-invasive tumors. Other genes involved in angiogenesis with increased levels in muscle-invasive tumors were *thrombospondin 1 and 2* (*THBS1* and *THBS2*), *vascular endothelial growth factor C* (*VEGFC*), and *neuropilin 2* (*NRP2*).

Although both pT_a and pT_1 tumors are considered superficial, they represent a heterogeneous group of tumors that may differ in their molecular pathogenesis and pathways of progression. Unsupervised hierarchical clustering separated these tumors into two groups. One cluster contained only pT_a tumors (mainly low grade), whereas the other contained all the pT_1 tumors (all high grade) and a subset of the pT_a tumors (both high and low grades). This second group of tumors expressed high levels of genes involved in promoting cell cycle progression (e.g., *CCNA2*, *CDC2*, *CDC6*, and *TOP2A*), suggesting a potential for greater proliferative activity. The pT_a cases in this second group potentially represent more aggressive pT_a tumors. Recently, a meta-analysis study of undifferentiated versus differentiated cancers of various types identified a 69-gene metasignature of neoplastic transformation (40). Eight of these 69 genes were predominantly associated with proliferation and overlapped with the genes in our stage pT_a and pT_1 classifier. Therefore, it may be possible that a more general expression signature may exist containing key proliferation genes that can distinguish between undifferentiated and differentiated tumor types or may be associated with aggressive behavior and poor patient outcome. Supervised PAM classification of pT_a versus pT_1 tumors had a 70% success rate based on 54 unique genes. The gene with the most power to discriminate between pT_a and pT_1 tumors was *IFI16*. Several studies have proposed a role for this protein in the regulation of cell growth and differentiation (41, 42). It has been proposed that IFI16 inhibits cell growth of epithelial cells in

part by up-regulating $p21^{WAF1}$ and also through its binding to the E2F1 transcription factor (43, 44). This is especially interesting because the cell cycle progression pathway is frequently altered during bladder tumor progression by *p16* deletion, *Rb* mutation, and *E2F* amplification.

Despite small numbers of samples, we report the successful classification of TCCs versus SCCs (89% correct) based on the expression patterns of 30 unique genes. Only cases with pure squamous features were designated as SCCs. Five of the TCC tumors were classified as SCC by the supervised analysis. Four of these tumors were considered to have a component of squamous differentiation, which may have contributed to their misclassification. Only one of six SCC tumors was classified incorrectly. Twenty-two of the 47 muscle-invasive transitional tumors had varying degrees of squamous differentiation. We attempted to predict the class (SCC versus TCC) of these samples based on the gene expression patterns of only nine genes (Table 3). This gene set was identified as the optimum classifier (overall success rate of 90%) when considering only SCC and TCC with "pure" histologic features. When the 22 transitional tumors with partial squamous differentiation were tested with this classifier, 32% were classified as SCC and 68% as TCC. The degree of squamous differentiation did not correlate with the classification of these samples to either category. The requirement for "pure" squamous features may be too stringent a criterion to identify squamous tumors, and it may prove useful to consider some of the TCC tumors with partial squamous differentiation as squamous type. The development of a molecular classifier based on RNA expression patterns may establish the histologic group of such tumors with greater precision and may better define this group for therapeutic trials.

The top predictive gene separating SCC from TCC tumors in all classifiers was *PTHrP*. This gene was also identified as significantly differentially expressed by both SAM and maxT, and differential expression on the protein level was validated by immunohistochemistry. Elevated expression of PTHrP has been reported previously in a spinal cord injury patient with SCC of the bladder (45) and in other reports of bladder SCC (46, 47). Neoplastic production and secretion of PTHrP may potentially result in hypercalcemia of malignancy, a well-known paraneoplastic syndrome. PTHrP has been implicated in differentiation, epithelial-mesenchymal interaction, proliferation, and apoptosis (48). Not surprisingly, many of the other genes also expressed at higher levels in the SCCs than the TCCs are also induced during inflammation and wound repair. Some have been reported previously as increased in bladder SCC, including *psoriasin* (*S100A7*; ref. 49), *S100A8* (*MRP-8*), and *S100A9* (*MRP-14*; ref. 50). Finally, *skin-derived protease inhibitor 3* (*SKALP*), which is an epithelial host defense protein highly induced in inflamed skin, also had elevated expression in SCC tumors. The expression of this gene has been associated with abnormal epithelial differentiation in the context of hyperproliferation. Other genes with the same function expressed in the SCC tumors were *basonuclin* (*BCN*) and *keratin 16* (*KTR16*). *KTR16* has been reported previously with high protein expression in bladder SCC (together with *KRT5* and *KRT14*).

Previous reports studying a diverse set of cancers have shown the utility of expression profiling in identifying gene sets predictive of outcome (9, 51–54). In bladder cancer, expression patterns have identified gene signatures that are associated with tumor stage (16, 17) and superficial recurrence (16). However, the number of muscle-invasive tumors reported in those studies was small, and robust outcome classification models for the invasive tumors were not possible. The current study is the first to report a predictive gene set for overall survival in a large cohort of patients with muscle-invasive disease. The classifier had a 78% classification success rate and was more accurate in predicting patients with bad prognosis (sensitivity 93%, specificity 65%). Such a classifier might help to identify those patients at highest risk, perhaps the best candidates for more aggressive therapies.

Genes in this classifier belonged to several functional groups. In the good prognosis group of muscle-invasive patients, there was high expression of genes involved in energy-related pathways (*PRKAG1*, *GAMT*, *ACOX1*, *ASAH1*, and *SCD*). In the bad prognosis group of patients, there was overexpression of genes involved in the regulation of cell growth and maintenance (*AF1Q*, *AREG*, *DUSP6*, and *LYAR*) and underexpression of tumor suppressor genes (*MAL* and *RARRES*). These predictive genes may point to novel therapeutic targets for bladder cancer patients with advanced disease.

All predictive genes identified by our supervised PAM analyses were sequence verified. This was necessary given the 10% to 20% error rate reported for such user-generated cDNA arrays. Further validation was secured by spotting several replicate clones several times on the array. This allowed for multiple expression measurements and showed that replicate clones always clustered together by unsupervised hierarchical clustering. In addition, as expected, several of the genes were represented by more than one clone in the classifier panels.

It is important to evaluate the performance of any classification rule on independent data sets. In the absence of an additional data set, approaches, such as sample splitting (train and test sets) or cross-validation approaches, have been used. In this study, we chose the method of cross-validation, as we could not afford to lose power by splitting our data set into a training set and a test set. Two of the classifiers were further assessed on the only publicly available independent data set of bladder tumors (Dyrskjot et al.). The first was the 25-gene classifier that separated our tumors into superficial versus muscle-invasive groups. This classifier had similar success in classifying our data set and the Dyrskjot et al. data set (90% versus 82.5%, respectively). Similarly, the gene classifier that separated our stage pT_a from stage pT_1 tumors was also successful when applied to the tumor set of Dyrskjot et al. (70% versus 87% success rate). Interestingly, there were no genes in common between our 25-gene classifier for superficial versus invasive tumors and the 32-gene classifier identified by Dyrskjot et al. Other studies have also reported minimal overlap for the genes identified for clinically similar tumors (9, 51, 52, 55). These differences might be related to the many genes whose expression is coordinately expressed in these tumor samples, thus allowing separate nonintersecting sets to perform similarly for classification.

The expression levels of the corresponding protein products for three genes (*CTSE*, *CCNA2*, and *PTHrP*) representing three of our tumor classifiers were evaluated by bladder tumor tissue microarrays. A similar differential pattern was seen for protein expression as for RNA expression, showing that these genes represent true biological differences and that protein

and RNA expression are directly linked for these genes. Interestingly, the $\log_2$ ratio values for RNA expression were significantly correlated to protein immunohistochemistry on a case-by-case basis for *CTSE* but not for *CCNA2* or *PTHrP*. This lack of association may be due to limited power due to small number of overlapping cases, to tumor heterogeneity if the tissue core did not represent the entire tumor, and to the variability quantitating immunohistochemistry in tissue arrays.

In this study, we have identified gene expression classifiers that separate bladder tumors by stage, histologic type, and outcome, and we have shown their applicability on an independent data set. These predictive gene sets should be validated on larger cohorts of patients with extended clinical follow-up. It would be especially useful to use samples from clinical trials with homogeneous treatment. Genes driving the separation between tumor subsets may prove to be important biomarkers for bladder cancer development and progression and eventually candidates for therapeutic targeting.

Acknowledgments

We thank Karen Chew, Clarissa Bush, and the University of California at San Francisco Comprehensive Cancer Center Tissue Core for the tissue microarray construction and immunohistochemistry and Drs. Dan Moore and Jane Fridlyand for careful review of the article. Microarray printing was done at the University of California at San Francisco Core Facility for Genomics and Proteomics, with the support of Dr. Ron Jensen. Sequencing was done by the University of California at San Francisco Comprehensive Cancer Center Genome Core.

References

1. Knowles MA. The genetics of transitional cell carcinoma: progress and potential clinical application. BJU Int 1999;84:412–27.
2. Spruck CH III, Ohneseit PF, Gonzalez-Zulueta M, et al. Two molecular pathways to transitional cell carcinoma of the bladder. Cancer Res 1994;54:784–8.
3. Simoneau AR, Jones PA. Bladder cancer: the molecular progression to invasive disease. World J Urol 1994; 12:89–95.
4. Dalbagni G, Presti J, Reuter V, Fair WR, Cordon-Cardo C. Genetic alterations in bladder cancer. Lancet 1993;342:469–71.
5. Cordon-Cardo C, Sheinfeld J, Dalbagni G. Genetic studies and molecular markers of bladder cancer. Semin Surg Oncol 1997;13:319–27.
6. Johnson DE, Schoenwald MB, Ayala AG, et al. Squamous cell carcinoma of the bladder. J Urol 1976;115: 542–4.
7. Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1999. CA Cancer J Clin 1999;49:8–31, 1.
8. Al-Sukhun S, Hussain M. Molecular biology of transitional cell carcinoma. Crit Rev Oncol Hematol 2003; 47:181–93.
9. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
10. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
11. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.
12. Thykjaer T, Workman C, Kruhoffer M, et al. Identification of gene expression patterns in superficial and invasive human bladder cancer. Cancer Res 2001;61: 2492–9.
13. Sanchez-Carbayo M, Socci ND, Charytonowicz E, et al. Molecular profiling of bladder cancer using cDNA microarrays: defining histogenesis and biological phenotypes. Cancer Res 2002;62:6973–80.
14. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531–7.
15. Sanchez-Carbayo M, Socci ND, Lozano JJ, et al. Gene discovery in bladder cancer progression using cDNA microarrays. Am J Pathol 2003;163:505–16.
16. Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet 2003;33:90–6.
17. Sanchez-Carbayo M. Use of high-throughput DNA microarrays to identify biomarkers for bladder cancer. Clin Chem 2003;49:23–31.
18. Frederick L, Greene DLP, Irvin D, et al. editors. AJCC cancer staging manual. 6th ed. New York: Springer-Verlag; 2002. p. 367.
19. Lopez-Beltran ASG, Gasser T, Hartmann A, et al. Infiltratingurothelial carcinoma. In: Eble JNSG, Epstein JI, Sesterhenn IA, editors. World Health Organization pathology and genetics tumors of the urinary system and male genital organs. 2004. p. 93.
20. Korkola JE, DeVries S, Fridlyand J, et al. Differentiation of lobular versus ductal breast carcinomas by expressionmicroarray analysis. Cancer Res 2003;63: 7167–75.
21. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation [online]. Nucleic Acids Res 2002;30:e15.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95: 14863–8.
23. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116–21.
24. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002; 99:6567–2.
25. Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. Hum Mol Genet 2001; 10:657–62.
26. Sasisekharan R, Shriver Z, Venkataraman G, Narayanasami U. Roles of heparan-sulphate glycosaminoglycans in cancer. Nat Rev Cancer 2002;2:521–8.
27. Ai X, Do AT, Lozynska O, Kusche-Gullberg M, Lindahl U, Emerson CP Jr. QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. J Cell Biol 2003; 162:341–51.
28. Stoehr R, Krieg RC, Knuechel R, et al. No evidence for involvement of β-catenin and APC in urothelial carcinomas. Int J Oncol 2002;20:905–11.
29. Hirano T, Manabe T, Takeuchi S. Serum cathepsin B levels and urinary excretion of cathepsin B in the cancer patients with remote metastasis. Cancer Lett 1993;70:41–4.
30. Visscher DW, Sloane BF, Sameni M, Babiarz JW, Jacobson J, Crissman JD. Clinicopathologic significance of cathepsin B immunostaining in transitional neoplasia. Mod Pathol 1994;7:76–81.
31. Eijan AM, Sandes E, Puricelli L, Bal De Kier Joffe E, Casabe AR. Cathepsin B levels in urine from bladder cancer patients. Oncol Rep 2000;7:1395–9.
32. Eijan AM, Sandes EO, Riveros MD, et al. High expression of cathepsin B in transitional bladder carcinoma correlates with tumor invasion. Cancer 2003; 98:262–8.
33. Sier CF, Blasi F, Sidenius N. Urinary levels of urokinase-type plasminogen activator and its receptor in the detection of bladder carcinoma. Cancer 2003; 98:1995; author reply 1996.
34. Vasala K, Paakko P, Turpeenniemi-Hujanen T. Matrix metalloproteinase-2 immunoreactive protein as a prognostic marker in bladder cancer. Urology 2003; 62:952–7.
35. Booth C, Harnden P, Selby PJ, Southgate J. Towards defining roles and relationships for tenascin-C and TGFβ-1 in the normal and neoplastic urinary bladder. J Pathol 2002;198:359–68.
36. Uno K, Azuma T, Nakajima M, et al. Clinical significance of cathepsin E in pancreatic juice in the diagnosis of pancreatic ductal adenocarcinoma. J Gastroenterol Hepatol 2000;15:1333–8.
37. Terris B, Blaveri E, Crnogorac-Jurcevic T, et al. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol 2002;160:1745–54.
38. Kamai T, Arai G, Takagi K. Granulocyte colony-stimulating factor and parathyroid hormone related protein producing bladder cancer. J Urol 1999;161: 1565–6.
39. Cho H, Harrison K, Schwartz O, Kehrl JH. The aorta and heart differentially express RGS (regulators of G-protein signalling) proteins that selectively regulate sphingosine 1-phosphate, angiotensin II and endothelin-1 signalling. Biochem J 2003;371:973–80.
40. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 2004;101:9309–14.
41. Johnstone RW, Trapani JA. Transcription and growth regulatory functions of the HIN-200 family of proteins. Mol Cell Biol 1999;19:5833–8.
42. Lembo M, Sacchi C, Zappador C, et al. Inhibition of cell proliferation by the interferon-inducible 204 gene, a member of the Ifi 200 cluster. Oncogene 1998;16: 1543–51.
43. Hertel L, Rolle S, De Andrea M, et al. The retinoblastoma protein is an essential mediator that links the interferon-inducible 204 gene to cell-cycle regulation. Oncogene 2000;19:3598–608.
44. Xin H, Curry J, Johnstone RW, Nickoloff BJ, Choubey D. Role of IFI 16, a member of the interferon-inducible p200-protein family, in prostate epithelial cellular senescence. Oncogene 2003;22: 4831–40.
45. Vaidyanathan S, Mansour P, Ueno M, et al. Problems in early diagnosis of bladder cancer in a spinal cord injury patient: report of a case of simultaneous production of granulocyte colony stimulating factor and parathyroid hormone-related protein by squamous cell carcinoma of urinary bladder. BMC Urol 2002;2:8.
46. Wolchok JD, Herr HW, Kelly WK. Localized squamous cell carcinoma of the bladder causing hypercalcemia and inhibition of PTH secretion. Urology 1998; 51:489–91.
47. Hirasawa K, Kitamura T, Oka T, Matsushita H.

Bladder tumor producing granulocyte colony-stimulating factor and parathyroid hormone related protein. J Urol 2002;167:2130.

48. Wysolmerski JJ, Stewart AF. The physiology of parathyroid hormone-related protein: an emerging role as a developmental factor. Annu Rev Physiol 1998;60:431–60.

49. Ostergaard M, Wolf H, Orntoft TF, Celis JE. Psoriasin (S100A7): a putative urinary marker for the follow-up of patients with bladder squamous cell carcinomas. Electrophoresis 1999;20:349–54.

50. Celis JE, Wolf H, Ostergaard M. Bladder squamous cell carcinoma biomarkers derived from proteomics. Electrophoresis 2000;21:2115–21.

51. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100:10393–8.

52. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002;346:1937–47.

53. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.

54. Lossos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004;350:1828–37.

55. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8:68–74.

Clinical Cancer Research

Bladder Cancer Outcome and Subtype Classification by Gene Expression

Ekaterini Blaveri, Jeff P. Simko, James E. Korkola, et al.

Clin Cancer Res 2005;11:4044-4055.