

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2009 February ; 18(2): 444–453. doi:10.1158/1055-9965.EPI-08-1002.

Bladder Cancer Associated Gene Expression Signatures Identified by Profiling of Exfoliated Urothelia

Charles J. Rosser^{*,1}, Li Liu[#], Yijun Sun[#], Patrick Villicana^{*}, Molly McCullers^{*}, Stacy Porvasnik^{*}, Paul R. Young[^], Alexander S. Parker[^], and Steve Goodison[~]

^{*} Department of Urology, University of Florida, Gainesville, Florida

[#] ICBR Bioinformatics, University of Florida, Gainesville, Florida

[~] Department of Surgery, University of Florida, Gainesville, Florida

[^] Department of Urology, Mayo Clinic Jacksonville, Jacksonville, Florida

Abstract

Bladder cancer is the fifth most commonly diagnosed malignancy in the United States and one of the most prevalent worldwide. It harbors a probability of recurrence of >50%, thus rigorous, long-term surveillance of patients is advocated. Flexible cystoscopy coupled with voided urine cytology (VUC) is the primary diagnostic approach, but cystoscopy is an uncomfortable, invasive procedure and the sensitivity of VUC is poor in all but high-grade tumors. Thus, improvements in non-invasive urinalysis assessment strategies would benefit patients. We applied gene expression microarray analysis to exfoliated urothelia recovered from bladder washes obtained prospectively from 46 patients with subsequently confirmed presence or absence of bladder cancer. Data from microarrays containing 56,000 targets was subjected to a panel of statistical analyses to identify bladder cancer-associated gene signatures. Hierarchical clustering and supervised learning algorithms were used to classify samples on the basis of tumor burden. A differentially expressed geneset of 319 gene probes was associated with the presence of bladder cancer ($P < 0.01$), and visualization of protein interaction networks revealed VEGF and AGT as pivotal factors in tumor cells. Supervised machine learning and a cross-validation approach were used to build a 14-gene molecular classifier that was able to classify patients with and without bladder cancer with an overall accuracy of 76%. Our results show that it is possible to achieve the detection of bladder cancer using molecular signatures present in exfoliated tumor urothelia. Further investigation and validation of the cancer-associated profiles may reveal important biomarkers for the non-invasive detection and surveillance of bladder cancer.

Keywords

Urinary bladder cancer; diagnostic signature; microarray profiling; urinalysis; VEGF

INTRODUCTION

Cancer of the urinary bladder is among the five most common malignancies world-wide and one of the most prevalent (1). Early detection remains one of the most urgent issues in bladder cancer research. When detected early, the 5-year survival rate is approximately 94%, thus timely intervention dramatically increases patient survival rate. At presentation, more than 80% of bladder tumors are non-invasive papillary tumors, but the remaining 20% exhibit muscle

1Correspondence: University of Florida College of Medicine Department of Urology Suite N215 PO Box 100247 Gainesville, FL 3210. Phone (352) 273-7658; Fax (352) 392-8846; Email: charles.rosser@urology.ufl.edu.

invasion at the time of diagnosis and have a much less favorable prognosis. While non-invasive lesions are treated by transurethral resection, more than 70% of patients with these lesions have disease recurrence during the first two years. If left untreated these initially non-invasive lesions can progress to being muscle-invasive (2). The recurrence phenomenon of non-invasive bladder tumors makes bladder cancer one of the more prevalent cancers; specifically 500,000 Americans are currently under treatment for bladder cancer (3). Furthermore, once bladder tumors are identified and removed, patients will routinely require strict surveillance cystoscopy every 3 months for 2 years, then every 6 months for 2 years, then yearly thereafter. As with the early detection of a first carcinoma event, the timely diagnosis and treatment of disease recurrence can dramatically improve the patients' quality of life.

The detection and surveillance of bladder cancer involves the visual inspection of the bladder for lesions using cystoscopic examination of the bladder. Cystoscopy is an uncomfortable, invasive procedure associated with significant cost and possible infection and trauma. Thus, the development of non-invasive assessment strategies is desirable for both patients and healthcare providers. Voided urine cytology (VUC) remains the method of choice for the non-invasive detection of bladder cancer lesions, with its major application being to recognize disease recurrence and early progression in tumor stage and grade. VUC can be used to diagnose new malignancy, yet whilst it has a specificity of >93%, its sensitivity is only 25–40%, especially for low-grade and low-stage tumors (4,5). Furthermore, this analysis is prone to inter-observer variation, results are not available rapidly, and it is relatively expensive. Accordingly, a good deal of research has focused on identifying potential urine tumor markers with higher sensitivity than provided by urine cytology alone. Diagnostic protein markers for urinalysis have been developed commercially, but these tests also suffer from high false-positive rates (6). Other promising diagnostics include telomerase detection (7) or activity assays (8) microsatellite instability assays and fluorescent *in situ* hybridization methods (9). However, these assays may have insufficient predictive power to be applied to the management of individual patients, and importantly, these techniques are complex, and require skillful interpretation. Thus, the identification of alternative biomarkers for the early detection and surveillance of bladder cancer in non-invasively obtained material remains important for the management of patients with this disease.

The advent of high-throughput microarray gene expression technology has greatly enabled the search for clinically important disease biomarkers. Numerous exploratory studies have demonstrated the potential value of gene expression signatures in tumor classification (10), diagnosis (11), and in assessing the risk of post-surgical disease recurrence (12,14) in many tissue types including bladder cancer (15–23). To date, gene expression profiling studies of urological clinical material have focused on the analysis of excised solid tumor tissue (15–23). These studies have identified gene signatures that are associated with tumor stage (15, 17,18,23), disease recurrence and outcome prediction (15,16,19,20,23), and subtype classification (16,17). The fact that follow-up studies have validated some of the biomarkers in independent tissue collections shows the potential utility of microarray profiling of bladder source materials (24,25), however, the molecular analysis of solid tissue is most applicable to the development of assays that will aid the histological evaluation of biopsy or excised tumor material. In order to progress towards the development of novel molecular assays for non-invasively obtained material, the more clinically appropriate material for profiling is the urine, and/or the surface transitional urothelia that are naturally shed into the urine. We and others are performing proteomic profiling of soluble factors in urine (26,27), but to our knowledge no one has yet performed gene expression profiling on shed urothelial cells obtained from patients with bladder cancer.

In this study, we investigated the feasibility of gene expression profiling of exfoliated urothelia in order to identify differentially expressed gene signatures associated with the presence or

absence of bladder cancer. Urothelial cells were obtained from bladder washings from 20 patients with confirmed bladder cancer and 26 patients with no evidence of bladder cancer. Two rounds of linear amplification of the urothelial mRNA enabled us to obtain enough material for hybridization to Affymetrix U133 Plus 2.0 GeneChips, and the application of robust statistical analyses identified a set of differentially expressed genes associated with bladder cancer. Furthermore, the application of a recently developed feature selection algorithm (14,28) revealed the optimal gene signature for discriminating between the two groups. Some of the genes identified in this study have been implicated in bladder cancer previously, but many have not. Further analysis of the data implicated a role for specific signaling pathways in the neoplastic urothelia. The ability to perform global gene expression profiling on the minimal material present in shed urothelial will greatly facilitate the identification and development of potential biomarkers for the detection of bladder cancer in non-invasively obtained patient samples.

MATERIALS and METHODS

Clinical sampling and processing

Under IRB approval and informed consent, urothelial samples and associated clinical information were prospectively collected from 46 consecutive individuals with no previous history of urothelia carcinoma. This cohort served as our phase I (feasibility study) according to the International Consensus Panel on Bladder Tumor Markers (29). Patients were undergoing complete hematuria workup including office cystoscopy and upper tract imaging by computed tomography of the abdomen and pelvis (without and with intravenous contrast). Two different clinical cohorts were analyzed in this study. The first group (control) consisted of 26 subjects with a negative evaluation (i.e., normal imaging of upper tract and normal cystoscopy). The second group (experimental) consisted of 20 subjects with a visible bladder tumor detected by upper tract imaging and/or cystoscopy, and which was later proven by evaluation of a biopsy to be urothelial carcinoma. Initially 100 mL of voided urine was obtained and processed for RNA. However, normal subjects were noted to have RNA concentrations < 2 ng/ μ L. It was felt that this concentration was too low to adequately perform microarray analysis. Thus the decision was made to utilize bladder washings to generate the initial genomic profile.

Sampling of exfoliated urothelia for both cohorts was obtained by injection of 50ml of saline into the bladder during the time of their office cystoscopy (barbotage). The saline solution was immediately aspirated and collected for subsequent analysis. Pertinent information on presentation, histologic grading and staging, therapy, and outcome were recorded (see Table 1). Each barbotage sample (50ml) was assigned a unique identifying number before immediate delivery and laboratory processing. Urothelial cells were pelleted by centrifugation ($600 \times g$, 4°C, 5 min), rinsed in PBS, pelleted again and lysed by direct application of RNeasy lysis buffer (Qiagen, Valencia, CA). RNA samples were evaluated quantitatively and qualitatively using an Agilent Bioanalyzer 2000, prior to storage at -80°C.

Gene expression profiling

RNA processing and hybridization—Gene expression profiling was performed on Affymetrix Human Genome arrays according to standard protocols (Affymetrix, Santa Clara, CA). However, due to the paucity of RNA recovered from the urothelial samples (50–200ng total RNA), a double amplification protocol was employed (30). Preparation of labeled cRNA was performed according to the Affymetrix Two-Cycle Target Labeling Assay (Affymetrix). Fragmented, biotinylated cRNA was hybridized to Affymetrix Human Genome U133 Plus 2.0 microarrays.

Generation of Expression Values—Microarray Suite version 5 (Affymetrix) was used to generate signal values, but not detection calls due to the practice of double amplifications. Rather, a mixture Gaussian model was built to determine the “Absent/Present” calls (31). Quality control (QC) of each GeneChip experiment included the assessment of the 5':3' ratio. This index reflects not only the original level of RNA integrity but also the accuracy of sample processing (32). Any samples that had a 5':3' index <1 were removed from analysis.

Tests of Significance—The two-sample Welch t statistics that allows unequal variances was used to identify genes that were differentially expressed between normal and tumor samples. The p-value was used to assess the statistical significance for each gene. To correct for multi-testing errors, the family-wise error rate (FWER) was employed following a permutation based bootstrap step-down minP procedure. In this procedure, no specific parametric form was assumed for the distribution of the test statistics. The class labels of the samples were permuted 10,000 times, and for each permutation, two-sample Welch t statistics were computed for each gene. The permutation p value for a particular gene is the proportion of the permutations (out of 10,000) in which the permuted test statistic exceeds the observed test statistic in absolute values. The above analyses were conducted using Bioconductor and AnalyzeIt Tools (<http://genomics3.biotech.ufl.edu/AnalyzeIt/AnalyzeIt.html>).

Gene Ontology and Pathway analysis—Gene Ontology annotations were obtained from Affymetrix. Biological network relationships among significantly regulated genes were explored using the Pathway Studio and ResNet mammalian database (Ariadne Genomics, Inc., Rockville, MD) (33). All microarray data obtained in the course of this study are available at <http://www.biotech.ufl.edu/people/rosser/ancillaries.html>.

Derivation of a diagnostic gene signature

In order to extract an accurate diagnostic signature from the microarray data, we applied a feature selection algorithm that we previously derived (13,14,28). The algorithm performs multivariate data analyses on high-dimensional data using well-established machine learning and numerical analysis techniques, but without making any assumptions about the underlying data distribution. The algorithm can perform feature selection and classification simultaneously. We have previously applied this algorithm, and an earlier version of it, to the derivation of optimal prognostic classifiers in breast and prostate cancer (14,28). To avoid possible overfitting of a computational model to training data, we used a rigorous experimental protocol with the leave-one-out cross validation (LOOCV) method to estimate classifier parameters and prediction performance (34). A receiver operating characteristic (ROC) curve obtained by varying a decision threshold is used to provide a direct view on how a prediction approach performs at the different sensitivity and specificity levels. Here, specificity is defined as the probability that a patient who did not have bladder cancer was assigned to the normal group, and the sensitivity is the probability that a patient who had bladder cancer was assigned to the disease group. For details of the computational algorithm see Sun et al. (13,14,28). The Matlab implementation of the algorithm is available upon request for validating the reported results and academic research.

Validation of profiles in an independent data set

Expression profiles of a panel of bladder tumors were generated by Dyrskjot et al. (17) in a study designed to identify gene expression in superficial tumors with respect to the presence of associated carcinoma in situ (CIS) lesions. The study included 28 tumor biopsies from superficial UC (stages Ta to T2) and 9 biopsies of normal bladder mucosa from patients with no history of bladder cancer. The study also contained 13 tumor biopsies from muscle-invasive UC, but as we had no such cases in our cohort, we did not include these in the validation study. All of the samples were obtained directly from surgery after removal of tissue for routine

pathological examination. RNA from these samples was hybridized to Affymetrix U133A GeneChips. Preparation of cRNA, hybridization and image acquisition were performed according to standard Affymetrix protocols. The microarray data for this study (accession # GSE3167) were retrieved from the Gene Expression Omnibus (National Center for Biotechnology Information). The mapping of the probes between the U133A and U133 Plus 2.0 GeneChip platforms was done using NetAffy Analysis Center (<http://www.affymetrix.com>).

RESULTS

Urothelial cell samples obtained from a total of 46 patients were obtained for this study. After complete evaluation, 26 subjects had no evidence of bladder tumors, and 20 subjects had biopsy confirmed urothelial carcinoma (UC). The median age for all patients was 65 years. Urine cytology was performed in all cases and was reported as non-suspicious/no neoplasia in all cases in the control group, and reported as suspicious/positive in 7 (35%) of the tumor-bearing group. All tumor-bearing samples were obtained from early stage disease (stages Ta, T1 or T2). A listing of the patient cohort characteristics are summarized in Table 1.

Gene expression profiles of the 46 urothelial samples were obtained by hybridization to Affymetrix U133 Plus 2.0 arrays containing 54,613 targets (covering 47,000 transcripts). Due to the paucity of RNA recovered from the urothelial cell samples (ranging from 50 ng to 200 ng total RNA), a two-cycle amplification strategy was required in order to generate enough labeled cRNA for array hybridization (30). As evaluation of minimal starting RNA material is awkward, we added a post-array hybridization quality control strategy to remove samples of insufficient RNA quality from further analysis. A high post chip 5':3' ratio > 1 corresponds to high quality material and processing (32).

After appropriate normalization, two-way hierarchical cluster analysis resulted in a distinct separation of the samples into two groups and identified a geneset that had expression patterns associated with disease status. The dendrogram (Figure 1) shows that one cluster contained 19 of the 20 tumor-bearing cases, and the second cluster contained the 26 non-cancer cases and one tumor case. A total of 319 differentially expressed genes (*p-value* < 0.01) associated with these clusters were ranked by *p-value* (see Table S1 in Supplemental data). The predominant compartmental class of genes in the set encoded integral membrane proteins (110 genes), and 53 genes encoded secreted proteins. These classes are of particular potential for development as biomarkers for urinalysis. The statistical significance of the gene expression correlations with disease status was further refined by calculation of the Family-wise Error Rate (FWER). Results derived from 10,000 permutations of the class labels (tumor or normal) revealed that the top 45 genes had a FWER < 0.05 (Table 2 and Table S2 in Supplemental data). Differentially expressed gene information was imported into 'Pathway Studio' software (Ariadne Genomics Inc. Rockville, MD). This analysis can reveal common regulators and associated pathway components within the dataset, based on multiple citations (33). The analysis revealed connectivity between many of the genes associated with bladder disease status, but mapping these relationships showed that this connectivity is mediated through a few key factors which act as signaling hubs (Figure 2). The two major hubs, vascular endothelial growth factor (VEGF) and angiotensinogen (AGT), both upregulated in tumor cases, are linked directly biologically, and indirectly through three minor hubs FLT1, ANG and ERBB2.

Gene expression differences between tumor and normal urothelial cell samples may implicate specific genes in malignant processes, and thus reveal insights into tumor biology. However, the feasibility of using such differences to classify unlabeled testing samples requires a different computational approach known as supervised machine learning. Here, we used our previously developed feature selection/classification algorithm to identify the gene signature that could

most accurately diagnose the 46 cases with respect to the presence of bladder cancer. With this modeling classification approach, a 14-gene model reached 76% overall accuracy in predicting class label during leave-one-out cross validation (Table 3). This level of accuracy supports the feasibility that gene expression differences can potentially be used to predict the identity of urothelial cell samples from patients of unknown clinical status. A receiver operator characteristic (ROC) plot was used to illustrate how the accuracy of the classifier varies at different sensitivity and specificity levels (Figure 3). For example, at a sensitivity of 90%, the molecular classifier had a specificity of 65%. The results compared very favorably with cytological evaluation of this cohort, where only 35% of tumor cases were correctly diagnosed. The genetic signature correctly classified 35 of the 46 samples (76%), including 17 of 26 normal cases and 18 of 20 tumor cases. The mean expression of each of the 14 classifier signature genes in the 46 samples obtained from patients with, and without, disease was visualized by creating individual gene scatter plots (Figure 4). The observed expression pattern for each of the 14 genes (p -value <0.01) is also listed in Table 3. P -values, computed using a standard Student's t -test, quantify the up- or down-regulation of individual genes between patients with and without disease.

Wherever possible, it is important to evaluate the performance of a disease classifier on independent data sets, however, given that our study is the first to profile exfoliated urothelia for bladder cancer detection, there are no directly comparable, independent datasets available. Furthermore, the majority of bladder cancer studies have used exclusively tumor tissue because their goal was to reveal genes associated with tumor subtype or outcome (14–23). The most appropriate publicly available microarray dataset we could find was a study by Dyrskjot et al that was designed to identify genes associated with carcinoma in situ (CIS), and included analysis of material from 41 tumor tissues and biopsies of nine normal bladder mucosa from patients with no history of bladder cancer (17). The gene expression levels of these tissue samples were measured using Affymetrix HG-U133A arrays. This is an earlier microarray format with some 20,000 targets and so less than half of the probes on the HG-U133 Plus 2.0 array used in our study are present on the HG-U133A array. Hence, it was not possible to validate the 14-gene prediction model directly on the solid tissue sample data. Instead, we checked the data distributions of genes associated with bladder cancer in our study that were present on both platforms. We first examined the distribution of our 45 discriminatory genes ($\text{FWER} < 0.05$) in the solid tissue sample dataset. Nineteen of the 45 genes were on both platforms, and 8 of those 19 (PART1, ZFAND6, SPATA2, DMBT1, KLF10, UBXN7, chr9orf42, WDR42A) were found to be significantly differentially expressed (p -value < 0.05) in the same direction in both urothelia and solid tissue profiles with respect to tumor vs. normal cases (see Table S3 in Supplemental Data). We then examined the distribution of the 14-gene tumor classifier on the solid tissue dataset. In this case, only 4 genes from the tumor classifier were present on both platforms. Among the four genes, DMBT1 expression was significantly (p -value 0.006) reduced in the solid tissue tumor samples in line with the urothelial cell data. Figure S4 in supplemental data shows a scatter plot depicting the distribution of DMBT1 expression in the 37 cases in the solid tissue dataset (16).

DISCUSSION

Detecting bladder cancer using diagnostic markers still remains a challenge. The inadequate power of single markers may partly explain this. The concept that the presence or absence of one molecular marker will aid diagnostic or prognostic evaluation has not proved to be the case which makes sense when one analyses the complex interactions between various molecules within a single pathway, the cross-talk between molecular pathways, the redundancy of some pathways and the oligoclonality of many tumors. There needs to be a paradigm shift from single-marker/single pathway research to a more global assessment of bladder cancer. To look

for such a profile in bladder cancer, it requires not only high-throughput molecular profiling, but also sophisticated bioinformatics tools for complex data analysis and pattern recognition.

The impetus for our search for bladder cancer biomarkers comes from the idea that an accurate biomarker can reduce the number of cystoscopies performed each year, and thus, cut down the frequency of this invasive and costly procedure. The way to achieve this is to identify tumor-associated molecules that are available in non-invasively obtained urine samples, either as soluble factors, or within the genome and/or transcriptome of urothelial cells that are naturally shed from the bladder lining and can be readily recovered from urine samples. The proteomic component of urine is an excellent biomarker discovery source material, and indeed we are developing techniques to maximize the analysis of urinary proteins (26), but proteomics does currently lag a little behind genomics in terms of target coverage and high-throughput technologies applicable to complex biological samples. Genomic profiling has been successfully applied to excised bladder tumor tissue specimens, and a panel of promising markers are undergoing validation in larger cohorts (15,20,34). Profiles gleaned from solid tissue specimens are confounded to some extent by cellular heterogeneity, and it is not clear whether candidate biomarkers present in solid tissue will necessarily translate to utility in non-invasively obtained urine specimens. Thus, solid tissue profiling data is perhaps more likely to augment histological evaluation of excised tissue for tumor subtype classification, treatment options, and prognostication.

Global gene expression analysis of 46 urothelial specimens revealed that gene expression differences were sufficiently robust to distinguish bladder cancer cases from non-cancer conditions. The complete listing and ranking of statistically significant genes is available in the supplementary data (Table S1). Protein interaction analysis of the data from exfoliated urothelia revealed connectivity between many of the genes associated with bladder disease status, but mapping these relationships showed that this connectivity is mediated through a few key factors which act as signaling hubs. The two major hubs, VEGF and angiotensinogen (AGT), both upregulated in tumor cases, are biologically linked directly, and indirectly through the three minor hubs FLT1, ANG and ERBB2. The network shown in Figure 2 places VEGF at the center of multiple interactions. Beyond the genes in the differentially expressed geneset derived in this study, VEGF regulates the expression of several extracellular factors, for example MMPs and uPA, which are believed to play a pivotal role in tumor growth by degrading extracellular matrix. The other major hub in the network, AGT, has multiple roles including being part of the tissue renin-angiotensin systems (tRAS) which may be a local source of angiotensin II that has specific paracrine functions. Any shift in the balance of the tRAS will have multiple effects on cell proliferation and angiogenesis in a tumor (35).

In order to extract an accurate diagnostic signature from the urothelial cell microarray data, we applied an improved feature selection algorithm that we previously derived (13,14,28). The algorithm addresses the major issues with prior work, including problems with computational complexity, solution accuracy, and capability to handle problems with extremely large data dimensionality (13). The key idea is to decompose an arbitrary complex model into a set of locally linear ones through local learning, and then estimate feature relevance globally within a large margin framework. We have experimentally demonstrated that our algorithm is capable of handling problems with extremely large input data dimensionality, to a point far beyond that needed for gene expression data analysis, and we have successfully applied the algorithm to the derivation of optimal prognostic classifiers in breast and prostate cancer (14,28). In this study, attempts to build a gene expression-based classifier of bladder cancer presence led to a 14-gene model that correctly predicted the status of 18 of the 20 cancer patients. A major advantage of deriving an accurate classifier signature with relatively few genes is that it facilitates downstream validation studies and the development of potential multiplex detection assays.

The performance of the disease status classifier was evaluated *in silico* using data from an independent bladder tumor study (17). Nineteen of the 45 top ranked genes in our study were present on the platform used in the validation dataset, and eight of these genes were significantly associated with bladder cancer in both urothelial and solid tissue datasets. Furthermore, expression of the DMBT1 gene, of the four genes of our 14-gene classifier that were included in the solid tissue dataset, was significantly reduced in cancer cases in both our study and the solid tissue study. As shown in the study by Dyrskjot et al (16), the 'normal' biopsy material does not include only epithelial cells, but an undefined amount of supporting stromal tissue also, such that the gene expression profile will be an average obtained from a complex tissue containing multiple specialized cell types. This highlights the advantage of using samples of exfoliated urothelia over solid tissues. In excised or biopsied tumor tissue samples the epithelial component will likely be the majority of cellular material, whereas, in normal healthy tissue samples, the epithelia may very well be in the minority. This discrepancy is not a problem when the sample is being used for morphological or immunohistochemical evaluation, but this can be a problem when samples are being compared using global molecular profiling techniques, when the data is an average of gene expression obtained from a complex tissue sample. The use of exfoliated urothelia samples overcomes this discrepancy in that the vast majority of cells will be of epithelial origin, whether obtained from tumor-bearing or healthy individuals. Given the differences in tissue sample, the different platforms used in the two studies and the fact that nonintersecting sets can perform similarly for classification due to coordinate expression, the observed overlap is encouraging for the use of urothelial cell sampling for bladder cancer detection and surveillance.

The above study has several limitations. First, though with similar number of subjects to previously reported bladder cancer microarray studies (17–20,22), the present data is from a small phase I study designed to illustrate the feasibility of profiling urine. Second, since sample RNA concentrations were low in normal individuals (< 2 ng/μl), bladder washings were used to generate the genomic profile. Subsequently, we assayed carefully obtained first morning voids and 24-hour urines which produced similar quantity and quality of RNA as we obtained from bladder washes (*data not shown*), thus if needed, the above data can be generated in a non-invasive manner. Thirdly, though we validated our data with that of published microarray databases, we did not validate our specific 14-gene model. In accordance to the recommendation of the International Consensus Panel on Bladder Tumor Markers on the development of biomarkers, next we will perform a phase II study assessing the clinical utility of our 14-gene model in a diverse cohort (e.g., gross hematuria, voiding symptoms, urinary tract infection and urolithiasis) and subsequently validate the results in a phase III study (29). Therefore, we did not see the need to validate our profile at this time if it is not robust enough in the phase II study. Note, RNA concentration from 100 ml of voided urine would be sufficient to validate our genomic profile utilizing quantitative PCR.

To our knowledge, this is the first report describing the global expression profiling of urothelia obtained from patients visiting the urology clinic. In this study, we have identified genes and diagnostic classifiers that can separate urothelial samples by disease status, and we have shown their association with solid tissue profiles on an independent data set. Currently, larger, confirmatory studies are underway prior to the development of clinically applicable tests, but the ability to profile the minimal urothelial component of urine, and the application of appropriate data analysis approaches, will greatly facilitate the development of non-invasive methods for the diagnosis and monitoring of both primary and recurrent bladder lesions. The data presented here suggest that it may be possible to detect and characterize bladder cancer based upon gene expression analysis of urothelia. Such strategies, if generalizable, would allow for the reduction of invasive procedures, improve surveillance and provide asymptomatic screening of high-risk populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the Florida Biomedical Research through a James and Esther King Award (CJR), Flight Attendant Medical Research Institute (CJR) and in part by the National Cancer Institute under grant RO1CA116161 (SG).

Abbreviation list

TCC	transitional cell carcinoma
SCC	squamous cell carcinoma
CIS	carcinoma in situ

References

1. Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. *Int J Cancer* 1999;83:18–29. [PubMed: 10449602]
2. Millan-Rodriguez F, Chechile-Toniolo G, Salvador-Bayarri J, Palou J, Algaba F, Vicente-Rodriguez J. Primary superficial bladder cancer risk groups according to progression, mortality and recurrence. *J Urol* 2000;164:680–4. [PubMed: 10954628]
3. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics. *CA Cancer J Clin* 2007;57(1):43–66. [PubMed: 17237035]
4. Cajulis RS, Haines GK 3rd, Frias-Hidvegi D, McVary K, Bacus JW. Cytology, flow cytometry, image analysis, and interphase cytogenetics by fluorescence in situ hybridization in the diagnosis of transitional cell carcinoma in bladder washes: a comparative study. *Diagn Cytopathol* 1995;13:214–23. [PubMed: 8575280]
5. Rife CC, Farrow GM, Utz DC. Urine cytology of transitional cell neoplasms. *Urol Clin North Am* 1979;6:599–612. [PubMed: 505675]
6. Hautmann S, Toma M, Lorenzo Gomez MF, et al. Immunocyt and the HA-HAase urine tests for the detection of bladder cancer: a side-by-side comparison. *Eur Urol* 2004;46:466–71. [PubMed: 15363562]
7. Khalbuss W, Goodison S. Immunohistochemical detection of hTERT in urothelial lesions: a potential adjunct to urine cytology. *Cytojournal* 2006;3:18. [PubMed: 16901344]
8. Yoshida K, Sugino T, Tahara H, et al. Telomerase activity in bladder carcinoma and its implication for noninvasive diagnosis by detection of exfoliated cancer cells in urine. *Cancer* 1997;79:362–9. [PubMed: 9010110]
9. Lotan Y, Bensalah K, Ruddell T, Shariat SF, Sagalowsky AI, Ashfaq R. Prospective evaluation of the clinical usefulness of reflex fluorescence in situ hybridization assay in patients with atypical cytology for the detection of urothelial carcinoma of the bladder. *J Urol* 2008;179:2164–9. [PubMed: 18423745]
10. 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. [PubMed: 10521349]
11. Stuart RO, Wachsman W, Berry CC, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc Natl Acad Sci U S A* 2004;101:615–20. [PubMed: 14722351]
12. LaTulippe E, Satagopan J, Smith A, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res* 2002;62:4499–506. [PubMed: 12154061]

13. Sun, Y.; Todorovic, S.; Goodison, S. A feature selection algorithm capable of handling extremely large data dimensionality. *Proc. 8th SIAM International Conference on Data Mining*; 2008. p. 530-540.
14. Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* 2007;23:30–7. [PubMed: 17130137]
15. Blaveri E, Brewer JL, Roydasgupta R, et al. Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin Cancer Res* 2005;11:7012–22. [PubMed: 16203795]
16. Blaveri E, Simko JP, Korkola JE, et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res* 2005;11:4044–55. [PubMed: 15930339]
17. Dyrskjot L, Kruhoffer M, Thykjaer T, et al. Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res* 2004;64:4040–8. [PubMed: 15173019]
18. Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* 2003;33:90–6. [PubMed: 12469123]
19. Dyrskjot L, Zieger K, Kruhoffer M, et al. A molecular signature in superficial bladder carcinoma predicts clinical outcome. *Clin Cancer Res* 2005;11:4029–36. [PubMed: 15930337]
20. Dyrskjot L, Zieger K, Real FX, et al. Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study. *Clin Cancer Res* 2007;13:3545–51. [PubMed: 17575217]
21. Frohlich C, Albrechtsen R, Dyrskjot L, Rudkjaer L, Orntoft TF, Wewer UM. Molecular profiling of ADAM12 in human bladder cancer. *Clin Cancer Res* 2006;12:7359–68. [PubMed: 17189408]
22. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol* 2006;24:778–89. [PubMed: 16432078]
23. Sanchez-Carbayo M, Socci ND, Lozano JJ, Haab BB, Cordon-Cardo C. Profiling bladder cancer using targeted antibody arrays. *Am J Pathol* 2006;168:93–103. [PubMed: 16400012]
24. Als AB, Dyrskjot L, von der Maase H, et al. Emmpin and survivin predict response and survival following cisplatin-containing chemotherapy in patients with advanced bladder cancer. *Clin Cancer Res* 2007;13:4407–14. [PubMed: 17671123]
25. Zieger K, Dyrskjot L, Wiuf C, et al. Role of activating fibroblast growth factor receptor 3 mutations in the development of bladder tumors. *Clin Cancer Res* 2005;11:7709–19. [PubMed: 16278391]
26. Kreunin P, Zhao J, Rosser C, Urquidi V, Lubman DM, Goodison S. Bladder cancer associated glycoprotein signatures revealed by urinary proteomic profiling. *J Proteome Res* 2007;6:2631–9. [PubMed: 17518487]
27. Wu TF, Ku WL, Tsay YG. Proteome-based diagnostics and prognosis of bladder transitional cell carcinoma. *Expert Rev Proteomics* 2007;4:639–47. [PubMed: 17941819]
28. Sun, Y.; Cai, Y.; Goodison, S. Combining nomogram and microarray data for predicting prostate cancer recurrence. *Proc. 8th IEEE International Conference on Bioinformatics and Bioengineering*; 2008.
29. Lokeshwar VB, Habuchi T, Grossman HB, Murphy WM, Hautmann SH, Hemstreet GP 3rd, Bono AV, Getzenberg RH, Goebell P, Schmitz-Dräger BJ, Schalken JA, Fradet Y, Marberger M, Messing E, Droller MJ. Bladder tumor markers beyond cytology: International Consensus Panel on bladder tumor markers. *Urology* 2005 Dec;66(6 Suppl 1):35–63. [PubMed: 16399415]
30. Wagner F, Radelof U. Performance of different small sample RNA amplification techniques for hybridization on Affymetrix GeneChips. *J Biotechnol* 2007;129:628–34. [PubMed: 17408796]
31. Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A* 2000;97:9834–9. [PubMed: 10963655]
32. Copois V, Bibeau F, Bascoul-Mollevi C, et al. Impact of RNA degradation on gene expression profiles: assessment of different methods to reliably determine RNA quality. *J Biotechnol* 2007;127:549–59. [PubMed: 16945445]
33. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* 2003;19:2155–7. [PubMed: 14594725]

34. Ruschhaupt M, Huber W, Poustka A, Mansmann U. A compendium to ensure computational reproducibility in high-dimensional classification tasks. *Stat Appl Genet Mol Biol* 2004;3:Article 37
35. Tahmasebi M, Barker S, Puddefoot JR, Vinson GP. Localisation of renin-angiotensin system (RAS) components in breast. *Br J Cancer* 2006;95:67–74. [PubMed: 16755291]

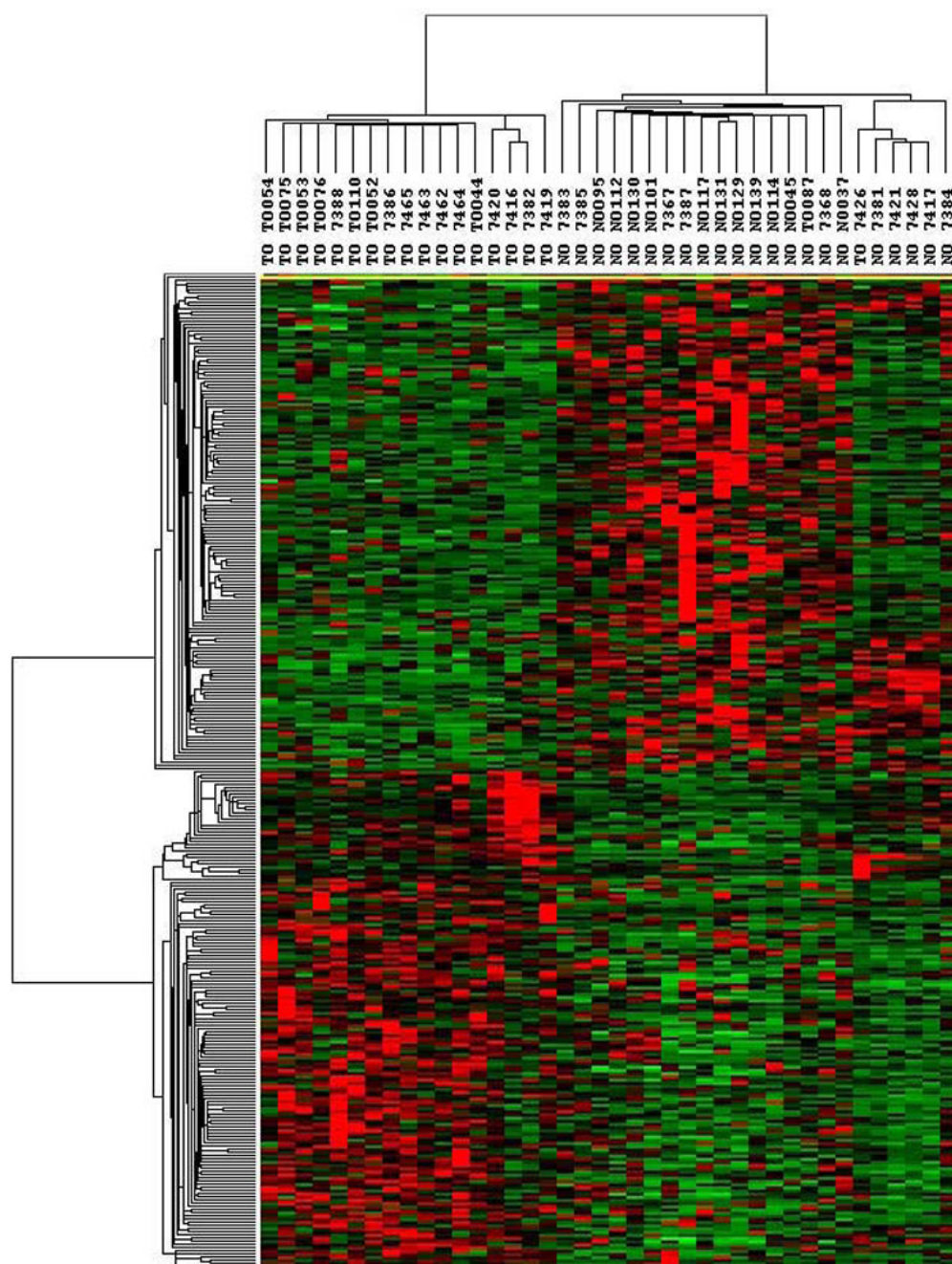


Figure 1.

Hierarchical cluster analysis of microarray data profiling urothelial cells obtained from 46 patients of known bladder disease status. Each row represents a sample and each column a gene. The scale represents standard deviations from the mean after a Z-transformation of signal values of a gene across all samples. *Red* represents a higher level of gene expression and *green* a lower level, relative to the mean across all samples for each gene. TO, tumor case; NO, non-tumor case.

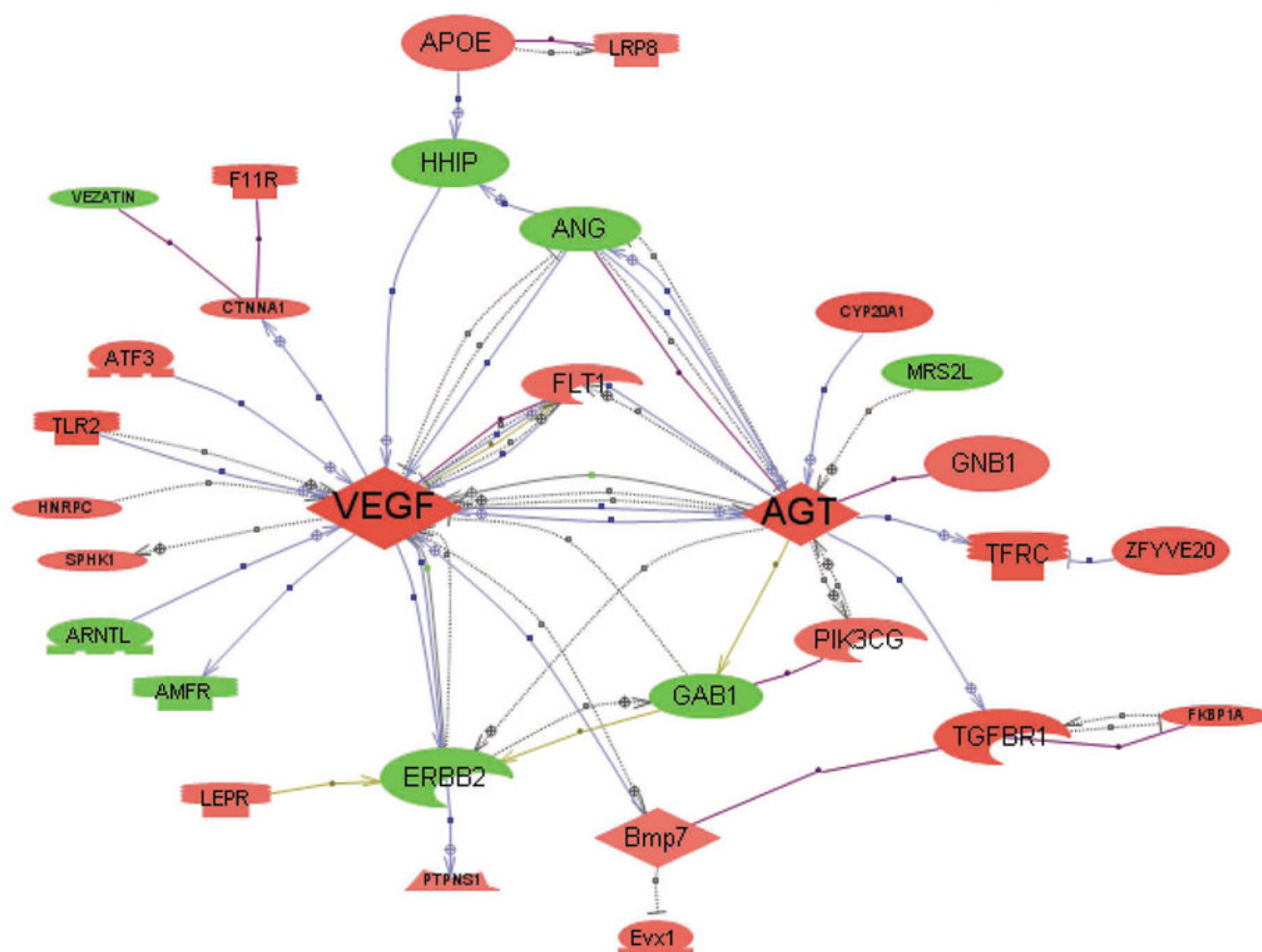


Figure 2.

Map of protein interactivity (binding, regulation, modification) between genes revealed to be significantly associated with bladder cancer in urothelial cell samples. Map was created by Pathway Studio (Ariadne Genomics, Inc., Rockville, MD). *Red* icons depict genes upregulated, *Green* icons depict downregulated genes. Lines depict reported interactions between proteins.

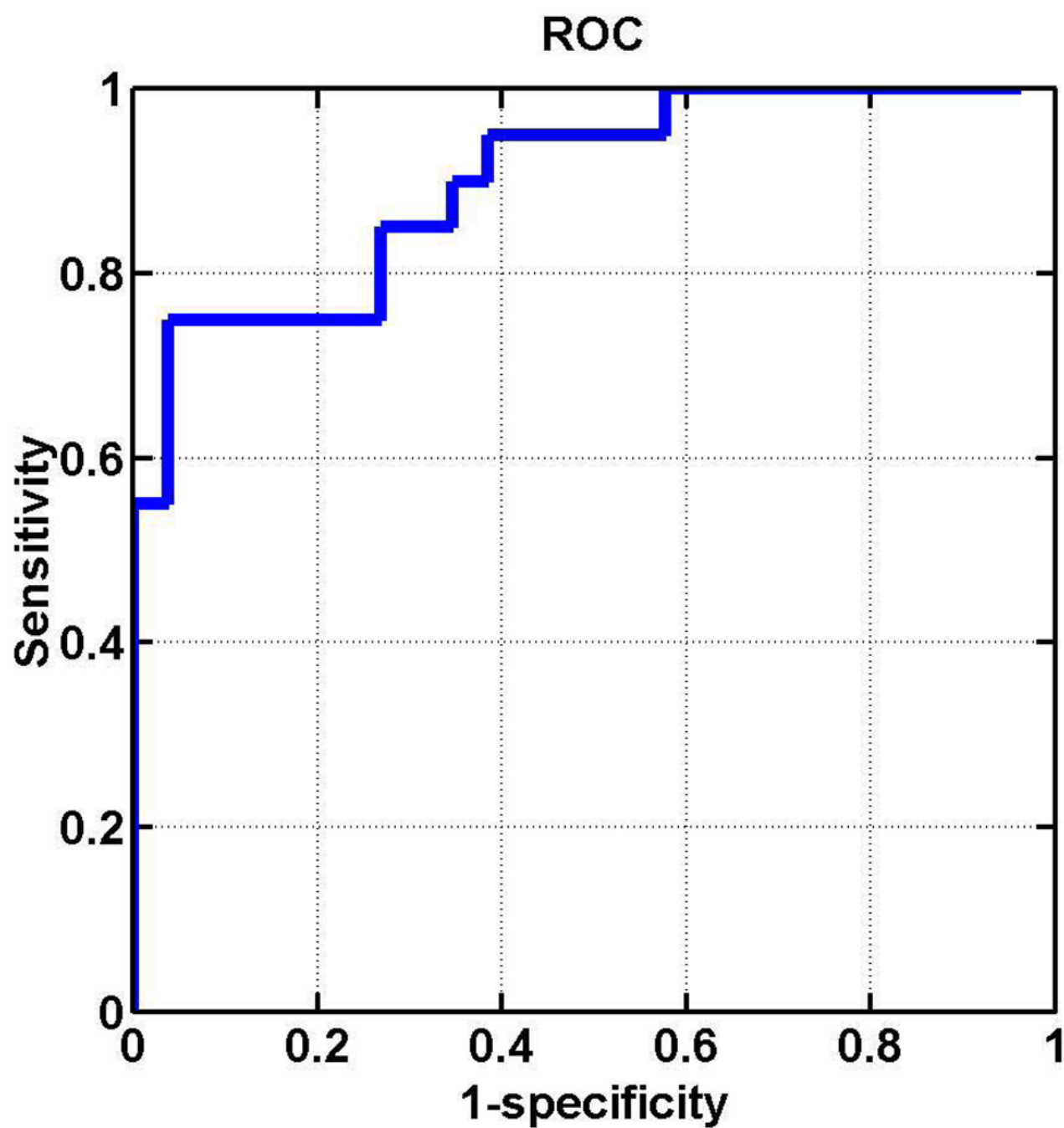


Figure 3.
ROC curve revealing the diagnostic accuracy of a 14-gene classifier for bladder cancer identified using a LOOCV validation.

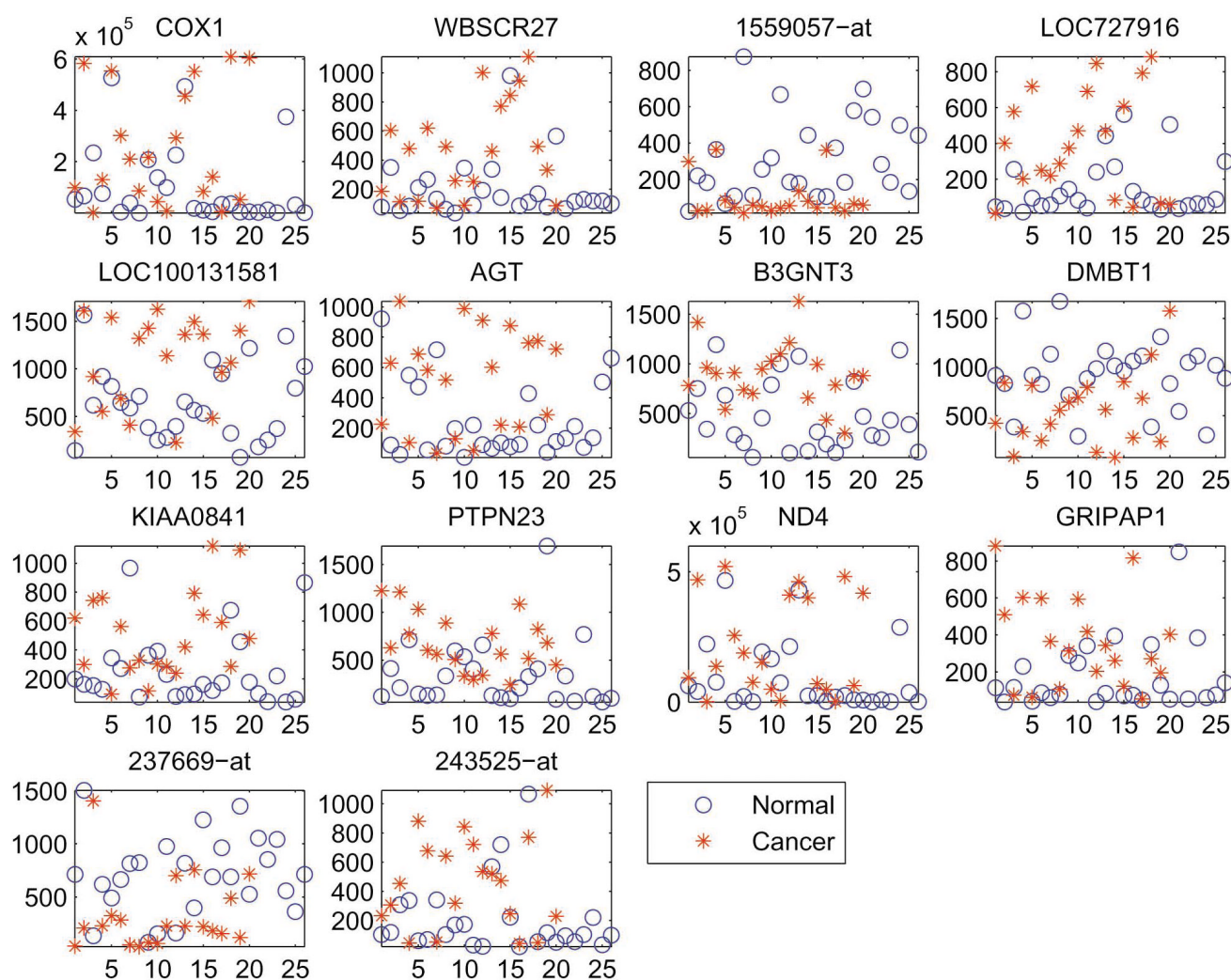


Figure 4.

Scatter plots of the fourteen gene markers in the diagnostic classifier demonstrating expression between patients with and without bladder cancer. The x-axis is the sample index, and the y-axis represents the gene expression values for that gene in each sample.

Table 1
Demographic and clinicopathologic characteristics of study cohort

	Non-cancer (%) [*] (n=26)	Cancer (%) (n=20)
Median Age (range, yrs)	65 (30–83)	65 (36–82)
Male: Female ratio	16:10	17:3
Race		
White	17 (65)	17 (85)
African American	6 (23)	3 (15)
Other	3 (12)	0 (0)
Tobacco use	10 (46)	14 (70)
Reported gross hematuria	18 (69)	14 (70)
Suspicious/Positive urine cytology	0 (0)	7 (35)
Clinical stage		
Ta	n/a	9 (45)
T1	n/a	4 (20)
T [^]	n/a	7 (35)
Grade		
1/2	n/a	9 (45)
3	n/a	11(55)
Median tumor size (cm)		3.1
Median no. of tumors		1.7

^{*}, negative hematuria evaluation; n/a, not applicable.

[^] one patient was noted to have carcinoma in situ (CIS) on pathologic evaluation

Table 2

Annotated differentially expressed genes identified in exfoliated urothelia from 46 cases. Expression patterns are described as either *upregulated*, or *downregulated* in the 20 tumor cases (Welch's t-statistic $p < 0.01$ and FWER < 0.05)

Gene ID	Gene Symbol	Gene Name	Ascribed Function
<i>Upregulated</i>			
1495	CTNNA1	Catenin (cadherin-associated protein), alpha 1	cell adhesion/negative regulation of apoptosis
7097	TLR2	toll-like receptor 2	induction of apoptosis
388121	TNFAIP8L3	tumor necrosis factor, alpha-induced protein 8- like 3	<i>unknown</i>
25930	PTPN23	protein tyrosine phosphatase, non-receptor type 23	tyrosine phosphatase
10331	B3GNT3	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3	protein glycosylation
9570	GOSR2	golgi SNAP receptor complex member 2	intracellular protein transport
9630	GNA14	Guanine nucleotide binding protein (G protein), alpha 14	ADP-ribosylation
2782	GNB1	guanine nucleotide binding protein (G protein), beta polypeptide 1	GTPase activity
26043	UBXN7	UBX domain protein 7	<i>unknown</i>
112399	EGLN3	egl nine homolog 3	Apoptosis/metal ion binding
6515	SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	glucose transmembrane transporter
54469	ZFAND6	zinc finger, AN1-type domain 6	DNA binding
7071	KLF10	Kruppel-like factor 10	transcription factor
4299	AFF1	AF4/FMR2 family, member 1	transcription factor
<i>Downregulated</i>			
2064	ERBB2	erb B2; Her2/neu	receptor signaling protein tyrosine kinase
1755	DMBT1	deleted in malignant brain tumors 1	cell differentiation, cell cycle regulation
25859	PART1	prostate androgen-regulated transcript 1	<i>unknown</i>
23762	OSBP2	oxysterol binding protein 2	lipid transport
138639	PTPDC1	Protein tyrosine phosphatase domain containing 1	protein dephosphorylation
4245	MGAT1	mannosyl (alpha-1,3-)-glycoprotein beta	carbohydrate metabolism
4809	NHP2L1	NHP2 non-histone chromosome protein 2-like 1	nuclear mRNA splicing
7627	ZNF75A	zinc finger protein 75a	transcription factor
2330	FMO5	Flavin containing monooxygenase 5	electron transport
9825	SPATA2	spermatogenesis associated 2	<i>unknown</i>

Table 3

The genes used in the 14-gene diagnostic model. Expression patterns are described as either *upregulated*, or *downregulated* in the 20 bladder cancer cases

Gene symbol	p-value	Function
<i>Upregulated</i>		
COX1	<0.01	oxidation/reduction, aerobic respiration
AGT	<0.005	serpin peptidase inhibitor
WBSCR27	<0.005	methyltransferase
B3GNT3	<0.001	glycosyltransferase
PTPN23	<0.005	protein tyrosine phosphatase
ND4	<0.05	NADH dehydrogenase (ubiquinone) activity
GRIPAP1	<0.005	guanine nucleotide exchange factor for the Ras family
KIAA0841	<0.005	hypothetical protein
LOC727916	<0.001	n/a
LOC100131581	<0.005	n/a
243525_at*	<0.005	n/a
<i>Downregulated</i>		
DMBT1	<0.005	scavenger receptor, immune response
237668_at*	<0.001	transmembrane glycoprotein
1559057_at*	<0.001	n/a

n/a, not available; distribution of each gene across all samples is depicted in Figure 4;

* Affymetrix probe id, Gene symbol unavailable.