

## Novel Blood Biomarkers of Human Urinary Bladder Cancer

Iman Osman,<sup>1</sup> Dean F. Bajorin,<sup>2</sup> Tung-Tien Sun,<sup>1</sup> Hong Zhong,<sup>1</sup> Diah Douglas,<sup>1</sup> Joseph Scattergood,<sup>2</sup> Run Zheng,<sup>3</sup> Mark Han,<sup>3</sup> K. Wayne Marshall,<sup>3</sup> and Choong-Chin Liew<sup>3,4</sup>

**Abstract Purpose:** Recent data indicate that cDNA microarray gene expression profile of blood cells can reflect disease states and thus have diagnostic value. We tested the hypothesis that blood cell gene expression can differentiate between bladder cancer and other genitourinary cancers as well as between bladder cancer and healthy controls.

**Experimental Design:** We used Affymetrix U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA) to profile circulating blood total RNA from 35 patients diagnosed with one of three types of genitourinary cancer [bladder cancer ( $n = 16$ ), testicular cancer ( $n = 10$ ), and renal cell carcinoma ( $n = 9$ )] and compared their cDNA profiles with those of 10 healthy subjects. We then verified the expression levels of selected genes from the Affymetrix results in a larger number of bladder cancer patients ( $n = 40$ ) and healthy controls ( $n = 27$ ).

**Results:** Blood gene expression profiles distinguished bladder cancer patients from healthy controls and from testicular and renal cancer patients. Differential expression of a combined set of seven gene transcripts (*insulin-like growth factor-binding protein 7*, *sorting nexin 16*, *chondroitin sulfate proteoglycan 6*, and *cathepsin D*, *chromodomain helicase DNA-binding protein 2*, *nell-like 2*, and *tumor necrosis factor receptor superfamily member 7*) was able to discriminate bladder cancer from control samples with a sensitivity of 83% (95% confidence interval, 67-93%) and a specificity of 93% (95% confidence interval, 76-99%).

**Conclusion:** We have shown that the gene expression profile of circulating blood cells can distinguish bladder cancer from other types of genitourinary cancer and healthy controls and can be used to identify novel blood markers for bladder cancer.

Blood cells communicate with cells and extracellular matrixes in almost all tissues and organs in the body. Such interactions can affect gene expression of the blood cells. Thus, it has been suggested that the gene expression profiles of circulating blood cells may reflect the presence of disease in the body (1). Microarray studies of disease have been based on RNA derived from biopsy/tissue samples. However, blood-based microarray studies have several major advantages over tissue-based assessments. Blood samples are less invasive, allow for a larger sample size, and make feasible repeated sampling to monitor disease progression.

Several recent studies have shown that total RNA derived from circulating blood can distinguish between control subjects and patients with various disease types (2-7). For example,

analyses of blood-derived total RNA were able to differentiate between patients with cardiovascular disease and healthy controls (2). Blood-based microarrays also distinguish patients with chronic fatigue syndrome from healthy controls (4). This latter example is particularly noteworthy because this syndrome is not detectable by any existing laboratory tests, and it has been unclear whether it represents a unique disease.

In this study, we test the hypothesis that analyses of blood cell-derived RNA may provide diagnostic clues to different types of malignant tumors. We used bladder cancer as our model because of the compelling need for novel markers for the diagnosis and monitoring of this disease. First, the diagnosis of bladder cancer is problematic; the most common symptom, hematuria, occurs in only 4% to 10% of bladder cancer cases (8). Second, the recurrence rate of treated bladder cancer is >70%, and tumor progression to a higher stage or grade develops in 40% of these patients; thus, bladder cancer patients require regular surveillance (9). Cystoscopy, the gold standard for detecting and monitoring bladder tumors, has a sensitivity of ~70% (10). However, cystoscopy is invasive and costly and is therefore not an ideal screening tool. Urine cytology, an alternative screening method, has a specificity of >90% (11) but a sensitivity of only 20% to 40% for low-grade tumors and 80% for high-grade tumors (12). Due to its low sensitivity for low-grade bladder cancer, urine cytology can be used only as an adjunct to cystoscopy. Thus, a simple, noninvasive screening test with adequate sensitivity and specificity is needed for the diagnosis and surveillance of bladder cancer after surgery.

**Authors' Affiliations:** <sup>1</sup>New York University School of Medicine; <sup>2</sup>Memorial Sloan-Kettering Cancer Center, New York; <sup>3</sup>ChondroGene, Inc., Toronto, Ontario, Canada; and <sup>4</sup>Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

Received 9/23/05; revised 3/15/06; accepted 4/4/06.

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.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

**Requests for reprints:** Iman Osman, New York University School of Medicine, Room H-100, 550 First Avenue, New York, NY 10016. Phone: 212-686-7500, ext. 3522; Fax: 212-951-3214; E-mail: Iman.Osman@med.nyu.edu.

©2006 American Association for Cancer Research.  
doi:10.1158/1078-0432.CCR-05-2081

In this study, we show that the gene expression profile of circulating blood cells can be used as a "sensor" to detect bladder cancer and to distinguish bladder cancer from other genitourinary cancers and healthy controls. We show that the expression signature, which is typically composed of several hundred genes, can be used as a platform to select a small set of, in our present study only seven, candidate genes for verification based on their relative discriminatory power between disease and control subjects.

## Patients and Methods

**Patient recruitment and clinical variables.** Ninety-two blood samples were collected for this study (44 bladder cancer, 10 testicular cancer, 9 kidney cancer, and 29 from healthy controls). As we are primarily interested in bladder cancer, we included both patients with superficial ( $n = 18$ ) and patients with locally advanced and/or metastatic disease ( $n = 26$ ). All bladder cancer patients had transitional cell carcinoma, except one patient who had clear cell carcinoma. Of these 92 blood samples, 45 samples were used in microarray experiments, comprising 16 subjects with bladder cancer (3 female, 13 male; mean age, 67 years; age range, 42-82 years), 9 subjects with metastatic kidney cancer (3 female, 6 male; mean age, 34 years; age range, 48-70 years), 10 subjects with testicular cancer under follow-up after completion of their treatment (mean age, 33 years; age range, 23-44 years), and 10 controls (5 female, 5 males; mean age, 30 years; age range, 19-42 years). Forty-seven samples were used in reverse transcription-PCR (RT-PCR) only (28 bladder cancer and 19 control), and 20 samples were used in both microarray and RT-PCR experiments (12 bladder cancer and 8 control).

Patients diagnosed with bladder, renal, and testicular cancers were recruited from Sidney Kimmel Center for Prostate and Urological Cancers at Memorial Sloan-Kettering Cancer Center under an institutional review board-approved protocol; all subjects provided written, informed consent before entry in the study. Clinical evaluation included a complete history and physical examination, an automated blood cell count, and imaging studies, including chest, abdominal, and/or pelvic computerized topography scans when clinically indicated. Tumor stage at the time of the initial blood sample collection was defined according to the American Joint Committee on Cancer guidelines (13). Controls were volunteers with no known medical conditions recruited from the Veterans Administration Medical Center; blood samples were provided on protocols for normal subjects approved by the Veterans Administration Medical Center institutional review board.

**Blood RNA isolation.** Blood samples (8-12 mL) were drawn into EDTA Vacutainer tubes (Becton Dickinson, Franklin Lakes, NJ) by routine venipuncture, labeled with a unique identification code (with no indication of patient identity), put immediately on ice, and transferred to the laboratory within 4 hours for blood processing. Total RNA from nucleated blood cells was isolated following lysis of erythrocytes and removal of cell debris. RNA was then purified with RNeasy Mini kit (Qiagen, Mississauga, Ontario, Canada). The quality of purified total RNA was assessed by Agilent 2100 Bioanalyzer RNA 6000 NanoChip (Agilent, Palo Alto, CA), and quantity was assessed by absorbance at 260 nm in Beckman Coulter (Fullerton, CA) DU640 spectrophotometer.

**Microarray hybridization and statistical analysis.** Purified total RNA (5  $\mu$ g) was labeled and hybridized onto Affymetrix U133 Plus 2.0 GeneChip oligonucleotide arrays (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Briefly, hybridization signals were scaled in the Affymetrix GCOS software (version 1.1.1) using a scaling factor determined by adjusting the global trimmed mean signal intensity value to 500 for each array and imported into GeneSpring version 6.2 (Silicon Genetics, Redwood City, CA). Signal intensities were then centered to the 50th percentile of each chip and, for each individual gene, to the median intensity of each specific subset first to minimize the possible technical bias and then to the whole sample set. The intensities of any replicate hybridizations were averaged subsequent to further analysis. Only genes labeled by the GCOS software as "present" or "marginal" in all samples were used for further analysis. Differentially expressed genes were identified using the Wilcoxon-Mann-Whitney nonparametric test ( $P < 0.05$ ). The Benjamini-Hochberg false discovery rate multiple test correction was applied whenever applicable. Hierarchical cluster analysis was done on each comparison to assess correlations among samples for each identified gene set.

**Quantitative real-time RT-PCR.** The forward and reverse primers for the target genes were designed by PrimerQuest (<http://biotools.idtdna.com/primerquest>; Integrated DNA Technologies, Coralville, IA). Sequences of the primer sets used for RT-PCR verification are listed in Table 1.

Amplification efficiency and specificity of the primer pairs were determined using serial dilution of reference cDNA generated from a normal blood RNA pool with confirmation on agarose gel to ensure that the values were within linear range and the amplification efficiency was approximately equal for each of the target gene tested. For real-time RT-PCR assay, first-strand cDNA was synthesized from 2  $\mu$ g total RNA using the ABI High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA) on a Perkin-Elmer (Wellesley, MA) DNA Thermal Cycler according to the manufacturer's protocol, and further amplification and quantitation of specific cDNA were done by Qiagen

**Table 1.** Primer sequences for genes used for RT-PCR

| Symbol         | Description  | 5' Primer sequence     | Position | 3' Primer sequence      | Position | Amplicon (bp) |
|----------------|--|------------------------|----------|-------------------------|----------|---------------|
| <i>IGFBP7</i>  | <i>Insulin-like growth factor – binding protein 7</i>      | TGCGAGCAAGGTCCTTCATA   | 488      | ATGAGGACAGGTGTCGGGATTC  | 597      | 110           |
| <i>SNX16</i>   | <i>Sorting nexin 16</i>                                    | AGTCTCTGCACTTGAGGTTGA  | 1,065    | CAGCATCATATGCCACTTCTGCT | 1,199    | 135           |
| <i>CSPG6</i>   | <i>Chondroitin sulfate proteoglycan 6 (bamacan)</i>        | ACTCGTGCCAACTTGATGAGC  | 794      | ACTTGGCGTTCGATATCCTCCA  | 912      | 119           |
| <i>CTSD</i>    | <i>Cathepsin D (lysosomal aspartyl peptidase)</i>          | TGTGGAGGACCTGATTGCCAAA | 259      | ACTGGGCGTCCATGTAGTTCTT  | 368      | 110           |
| <i>CHD2</i>    | <i>Chromodomain helicase DNA-binding protein 2</i>         | TGACAAGAAGCCAAAGCGCA   | 4,037    | TATGCACTCCAGCCGTTCAAGA  | 4,175    | 139           |
| <i>NELL2</i>   | <i>Nell-like 2 (chicken)</i>                               | CCAGCTGTGAAACGGACATTGA | 1,742    | TCATGGTAGCCATCTCTGCACT  | 1,859    | 118           |
| <i>TNFRSF7</i> | <i>Tumor necrosis factor receptor superfamily member 7</i> | AACTCTGGTCTTCTCGTTCGCA | 458      | TTGGAAGAGGATCACACTCGGT  | 569      | 112           |

Quantitect SYBR Green PCR kit using an ABI Prism 7500 real-time PCR system (Applied Biosystems). The confirmation of specific amplification and lack of primer dimer formation were determined by calculated melting dissociation curve. In each sample, the expression level of a target gene was quantified by its threshold cycle (Ct) value, which is the concentration-dependent PCR cycle number at which the amplicon becomes distinguishable over background. Ratio  $\Delta Ct$  (changes in Ct) value between two candidate genes (ratio  $\Delta Ct = Ct$  of target gene 1 - Ct of target gene 2) was used to assess the difference between the tested experimental groups on specific gene combinations. The significance in change of ratio  $\Delta Ct$  was evaluated by the nonparametric Mann-Whitney test (SigmaStat3.1, SPSS, Inc., Chicago, IL).

To ensure reproducibility of results, all genes were tested in triplicate and the averaged Ct value obtained was used for calculation. Total RNA isolated from a set of 40 bladder cancer and 27 healthy control peripheral blood samples was used for the assays.

**Gene ratio, combination, and receiver operating characteristic curve analysis.** Receiver operating characteristic (ROC) curve analysis using MedCalc (Mariakerke, Belgium), XLSTAT (Addin-Soft, Paris, France), and also our home developed software, as described previously (7), was used to evaluate the diagnostic discrimination power of the selected differentially expressed genes that best functioned as classifiers for bladder cancer. Initially identified candidate genes from microarray analysis were assayed by quantitative RT-PCR. The ratio  $\Delta Ct$  of two candidate genes from the result of quantitative RT-PCR was used as analysis input to assay of ROC curve. The top 10 ratio  $\Delta Ct$  pairs evaluated by ROC curve were then selected for multiple combination analysis using logistic regression to generate a best combination ratio  $\Delta Ct$ . Logistic regression generates the coefficients of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:  $\text{Logit}(P) = b_0 + b_1 \text{ ratio } \Delta Ct_1 + b_2 \text{ ratio } \Delta Ct_2 + \dots + b_n \text{ ratio } \Delta Ct_n$ , where  $P$  is the probability of presence of the characteristic of interest and  $b_0, b_1, b_2, \dots, b_n$  are the coefficient of regression equation as described previously by Pampel (14). In this study, logistic regression was done to generate a best combination of biomarker panel.

## Results

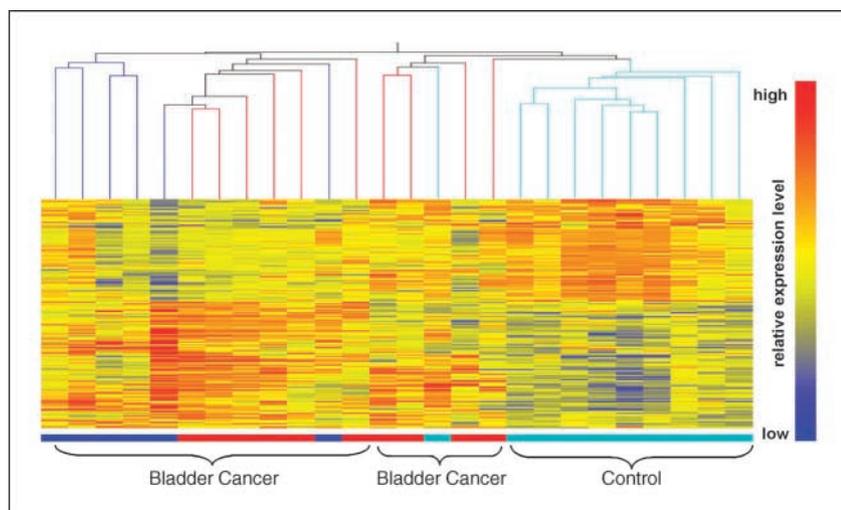
**Identification of differentially expressed genes through microarray.** Patients were recruited from among those who presented with urological, oncological diseases to the Sidney Kimmel Center of the Memorial Sloan-Kettering Cancer Institute.

A comparison of the blood-cell gene expression profiles of bladder cancer patients and healthy individuals using the Affymetrix GeneChip system showed significant differences ( $P < 0.05$ ) in the expression levels of 1,088 genes. To avoid potential bias because age and gender were not tightly controlled in our patient population, 34 genes known to be highly correlated with age or gender (data not shown) were removed from further consideration to avoid potential bias. Of the remaining 1,054 transcripts, 514 were overexpressed in the blood cells of bladder cancer patients and 540 were underexpressed (see Supplementary Table S1). Of these 1,054 transcripts, only 842 encoded proteins that have been defined previously or that have predicted functions as determined using the Database for Annotation, Visualization, and Integrated Discovery 2.0 (<http://david.niaid.nih.gov/david/version2/index.htm>) software. Proteins with predicted functions included proteins known to be involved in signaling pathways, such as cell cycle and apoptosis, mitogen-activated protein kinase signaling, Janus-activated kinase-signal transducers and activators of transcription signaling, transforming growth factor- $\beta$  signaling, and p53 signaling pathways.

We also did cluster analysis on the 1,054 gene transcripts to arrange the samples according to similarities in gene expression patterns as described previously (15). Hierarchical cluster analysis of the above-mentioned, differentially expressed genes resulted in a clear separation of the bladder cancer group from the healthy controls (Fig. 1).

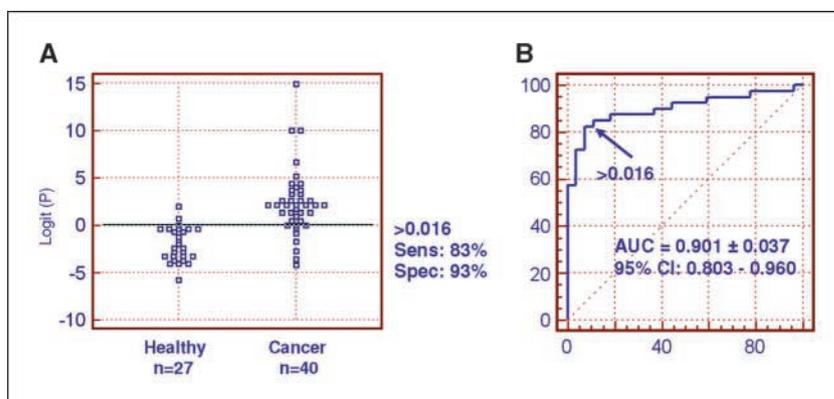
**Select candidate genes for RT-PCR verification.** From the 514 blood genes overexpressed and the 540 blood genes underexpressed in bladder cancer patients, we selected a set of 20 genes as initial candidates to be evaluated by quantitative real-time RT-PCR. We used the following criteria to make our selection:  $P$ , fold change, discrimination power from ROC curve analysis between bladder cancer patients and the control group, and function of the gene.

The initially identified candidate genes were assayed by quantitative RT-PCR on a sample set of 20 bladder cancers and 14 controls. The ratio  $\Delta Ct$  of two candidate genes was used as analysis input to assay if good discrimination power based on the level of ROC curve was observed. Then, the top 10 ratio  $\Delta Ct$  combinations were further investigated on a larger population of quantitative RT-PCR testing using 40 bladder cancer and



**Fig. 1.** Hierarchical cluster analysis of differentially expressed genes in blood from patients with bladder cancer and healthy controls. Hierarchical cluster analysis of 157 genes ( $P < 0.005$ ) differentially expressed in blood from patients with bladder cancer and healthy control subjects by using Spearman correlation. *Cyan*, control ( $n = 10$ ); *red*, superficial bladder cancer ( $n = 10$ ); *blue*, metastatic bladder cancer ( $n = 6$ ).

**Fig. 2.** Discrimination power of a combination of six gene ratios for bladder cancer. The level of gene expression ratio was evaluated using controls ( $n = 27$ ) and bladder cancer ( $n = 40$ ) by SYBR Green real-time RT-PCR. The best combination of six two-gene ratios  $\Delta Ct$  was generated using logistic regression analysis. **A**, interactive dot diagram presents the outcome for each sample in each class ( $\square$ ) as calculated. 0.016 is used as the cutoff to classify these samples as cancer ( $>0.016$ ) or control ( $\leq 0.016$ ). *Sens* and *Spec*, sensitivity and specificity for the complete training set, respectively. **B**, ROC curve analysis for bladder cancer versus healthy control.



27 controls. The top 10 two-gene ratio  $\Delta Ct$  combinations included the following seven genes (Table 1): *insulin-like growth factor-binding protein 7 (IGFBP7)*, *sorting nexin 16 (SNX16)*, *chondroitin sulfate proteoglycan 6 (CSPG6; bamacan)*, *cathepsin D (CTSD; lysosomal aspartyl peptidase)*, *chromodomain helicase DNA-binding protein 2 (CHD2)*, *nell-like 2 (NELL2)*, and *tumor necrosis factor receptor superfamily member 7 (TNFRSF7)*.

The logistic regression and ROC curve analysis showed that combinations of multiple ratio  $\Delta Ct$  values yield stronger diagnostic discrimination power than any of the single two-gene ratio  $\Delta Ct$ . The best combination equation used all seven genes, in the format of six two-gene ratios formulated as follows:  $\text{Logit}(P) = -6.3439 + 4.5490(\text{CHD2}/\text{CSPG6}) + 1.8878(\text{CHD2}/\text{CTSD}) - 0.7096(\text{CHD2}/\text{IGFBP7}) + 0.9320(\text{CHD2}/\text{SNX16}) - 1.0460(\text{IGFBP7}/\text{NELL2}) - 0.5698(\text{IGFBP7}/\text{TNFRSF7})$ .

**RT-PCR verification of seven genes identified from microarray.** To control for technical variation, we usually used a change in Ct value between a candidate gene and a housekeeping gene ( $\Delta Ct = Ct$  of target gene  $- Ct$  of housekeeping gene) to assess differences between the tested experimental groups. In this study, we found statistically significant differences in the expression of both  $\beta$ -actin and glyceraldehyde-3-phosphate dehydrogenase between control and bladder cancer groups at the RT-PCR level; glyceraldehyde-3-phosphate dehydrogenase also showed a statistically significant ( $P = 0.0127$ ) 1.27-fold increase in expression in bladder cancer samples compared with the control samples in the array experiment. Therefore, neither glyceraldehyde-3-phosphate dehydrogenase nor  $\beta$ -actin could be considered a "housekeeping gene," making it impossible to use them as a control to verify the expression of other genes by RT-PCR. A strategy of directly comparing the differential expression of two expressed genes was therefore implemented and proven more beneficial than using a housekeeping gene. The use of pair-wise gene ratios for disease group discrimination using cDNA array data from tissues (rather than from blood) has been described (16, 17). The technique reduces interindividual variation caused by biological and technical factors and is also independent of the expression measuring platform for data acquisition. Other advantages of the pair-wise gene method include the fact that the method is relatively independent of the input of sample amount, making it unnecessary to use a housekeeping gene as loading control, and the method requires only small amounts of RNA. Finally, the use of the ratio of an up-regulated gene to a

down-regulated gene amplifies the signal, thus making the assay more sensitive.

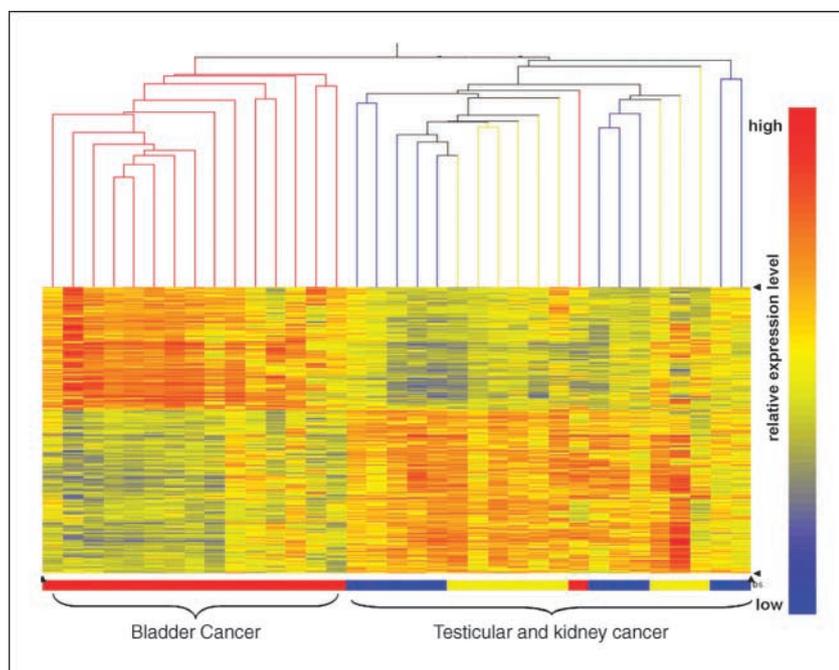
The sensitivity of the assay in this cohort is 83%, as 34 of 40 bladder cancer cases were predicted correctly. Of note, 3 of the 6 bladder cancer patients predicted wrongly by the assay to have no bladder cancer were patients with resected superficial bladder cancer. This suggests that the assay might be useful in determining patient prognosis. The specificity of the assay is 93%, as only 2 of the 27 healthy controls were predicted wrongly to have bladder cancer (Fig. 2). We stress here that the true specificity of the assay will require the inclusion of patients with bladder pathology other than cancer.

**Comparison of the three different genitourinary cancers.** We considered the possibility that the divergent hierarchical clustering of genes observed in the patients with bladder cancer was not bladder cancer specific but rather a general phenomenon associated with the development of any cancer. We therefore investigated gene expression profiles in patients with two nonbladder genitourinary tract cancer types: testicular and renal cell cancer. Unlike bladder cancer, which derives from urothelium, testicular cancer arises from germ cell elements and renal cell carcinoma arises from the convoluted tubules of the renal parenchyma. Interestingly, we found that blood cell gene expression clearly distinguished bladder cancer patients from cured testicular cancer patients and from advanced renal cell carcinoma patients (Fig. 3).

We identified 3,245 genes that were differentially expressed between bladder cancer patients and testicular and renal carcinoma patients. Of these genes, 346 overlapped with the set of 1,054 genes differentially expressed in bladder cancer and control groups (Supplementary Table S2). These 346 genes were distinct bladder cancer-specific genes showing significant differences in their level of transcription between bladder cancer patients and controls as well as between bladder cancer patients and testicular and renal cell carcinoma patients. Three of these genes, *CSPG6*, *SNX16*, and *IGFBP7*, were also among the seven quantitative RT-PCR verified genes that, when used as a combination of six gene ratios, showed rigorous classification power for discriminating bladder cancer and control subject samples.

## Discussion

We report here the feasibility of using gene expression profiling of circulating blood cells to identify novel blood



**Fig. 3.** Hierarchical cluster analysis of differentially expressed genes in blood from patients with bladder cancer compared with patients with testicular and kidney cancer. Hierarchical cluster analysis of 663 genes ( $P < 0.001$ , with multiple test correction) differentially expressed in blood from patients with bladder cancer (red;  $n = 16$ ), testicular cancer (blue;  $n = 10$ ), and kidney cancer (yellow;  $n = 9$ ) by using Spearman correlation.

biomarkers in cancer patients. Our study makes several points. First, blood gene expression signatures can differentiate between patients with bladder cancer and healthy controls and among patients with other types of genitourinary tumors. Second, the expression signature, which is typically composed of several hundred genes, can be used as a platform to select a small set of candidate genes for verification based on their relative discriminatory power between diseases and controls. Although our primary interest is bladder cancer, we believe that the approach shown in our study can be used to identify biomarkers in other types of cancers.

We identified a set of seven genes (*IGFBP7*, *SNX16*, *CSPG6*, *CTSD*, *CHD2*, *NELL2*, and *TNFRSF7*) that, when combined as six two-gene ratios, had the power to discriminate between bladder cancer patients and healthy control subjects. Four genes were overexpressed in bladder cancer patients compared with healthy controls. The first of these, *IGFBP7*, is one of several IGFBP isoforms that, through high-affinity binding interactions, acts to modulate the biological functions of IGFs, which are potent mitogens and antiapoptotic factors (18). The IGF system has been implicated in bladder cancer progression: exogenous IGF-I has been shown to activate antiapoptotic events in human bladder cancer T24 cells; suppression of IGF-I receptor expression has been shown to inhibit growth and survival of bladder cancer cells (19). However, the significance of IGFBP expression in bladder cancer tissues has not been well characterized. One recent report suggests that bladder cancer patients who have higher expression levels of IGFBP and lower levels of IGFBP3 in their bladder cancer tissues have a higher rate of disease recurrence compared with those patients who do not show this expression pattern in their tumor (20). The reasons for the differential expression of IGFBP2 and IGFBP3 were, however, not clear. *IGFBP7* (mac25) was overexpressed in the blood of bladder cancer patients in our study and has been found by other groups to

be overexpressed in inflammatory breast cancer cells and mesothelioma (21, 22). Further studies are needed to understand better the biological and clinical relevance of *IGFBP7* in both circulating blood and bladder cancer tissues. The second gene shown to be overexpressed in bladder cancer patients compared with healthy controls is *SNX16*. Sorting nexins are implicated in the regulation of membrane trafficking and in sorting processes of epithelial growth factor receptor, a protein known to play a significant role in bladder cancer progression. The third gene, *CSPG6*, has been shown to induce cell transformation (23). The last gene, *CTSD*, has been shown to have a role in promoting cancer cell proliferation and invasion (24). Studies have shown that *CTSD* facilitates early phases of tumor progression, such as cell proliferation and local dissemination (25).

Three genes were underexpressed in bladder cancer patients compared with healthy controls. *CHD2* belongs to a protein family characterized by the presence of the chromo (chromatin organization modifier) domain and the SNF2-related helicase/ATPase domain. CHD genes may modulate gene expression by modifying chromatin structure, thereby affecting access of the transcriptional apparatus to its chromosomal DNA template (26). It is unclear why this gene would be underexpressed in the blood of bladder cancer patients. The second gene, *NELL2*, encodes a cytoplasmic protein that contains epidermal growth factor-like repeats. The recently described *NELL1* and *NELL2* (neural epidermal growth factor-like 1 and 2) are members of the epidermal growth factor gene family that have been shown previously to be expressed almost exclusively in brain tissue (27). *TNFRSF7*, the third gene found in our study to be underexpressed in bladder cancer, is a member of the TNF receptor superfamily. This receptor is required for generation and long-term maintenance of T-cell immunity. It binds to ligand CD70 and plays a key role in regulating B-cell activation and immunoglobulin synthesis (28).

Several interesting findings emerged from our comparison between bladder cancer patients and patients with testicular and renal cancer. First, we identified 346 genes that overlap with the 1,054 genes differentially expressed between bladder cancer and controls (Supplementary Table S2). We postulate that these overlapping genes are bladder-specific markers. Second, three genes (*IGFBP7*, *SNX16*, and *CSPG6*) are among the four genes (*IGFBP7*, *SNX16*, *CSPG6*, and *CTSD*) that are overexpressed in bladder cancer patients compared with control subjects. The fact that *CTSD* cannot distinguish between bladder cancer and the two genitourinary cancers indicates that overexpression of this gene represents a generalized phenomenon in patients with cancer.

Our study has several limitations. The first possible limitation is that the expression profiles identified in this study may represent the activation of specific immunologic response to the presence of bladder tumors, and the profiles identified in this study may be intrinsic to the cohort of patients evaluated in this study. Our choice of seven genes for our assay is based on comparison between bladder cancer and healthy control subjects, and it is possible that a larger, independent cohort of patients could identify different genes or more genes that have better discriminatory power between bladder cancer and control subjects that include patients with bladder pathology other than cancer (e.g., urinary tract infection or stone). To address this limitation, we have embarked on a prospective study of a larger number of patients that includes three cohorts of test subjects: (a) bladder cancer patients with superficial, noninvasive bladder cancer; (b) patients with advanced bladder cancer that has metastasized to lymph nodes or visceral organs; and (c) control subjects with noncancer bladder pathology.

Second, we do not know the nature of the mechanisms driving the different gene expression signatures in circulating blood cells of the patients and control subjects studied. One possibility is that circulating tumor cells might be contributing, at least in part, to the distinct gene expression patterns, although we do not believe this to be the case. There are data from our group and other groups that both support and refute this hypothesis. Supporting this interpretation is our recent finding of uroplakin messages in the circulating blood cells of bladder cancer patients (29). However, uroplakins were not among the overexpressed genes in the circulating blood of the bladder cancer patients in this study. As well, a recent study of gene expression profiles of lymphocytes in patients with renal cell carcinoma (30) did not detect circulating renal cancer-specific markers.

In conclusion, we show that the gene expression profile of circulating blood cells can distinguish bladder cancer from other genitourinary cancers and from healthy controls. We also show that the expression signature can be used to select a small set of candidate genes for verification based on their relative discriminatory power between disease and control subjects. This potentially promising methodology needs to be evaluated more extensively to determine its clinical utility. Lastly, the use of gene transcript biomarkers from circulating blood cells provides a new paradigm for the identification of novel markers for various cancers.

## Acknowledgments

We thank Drs. Tom Yager, Hongwei Zhang, and Adam Dempsey (ChondroGene) for many thoughtful discussions and expert technical assistance on this project, and Joanna Spira for editorial and administrative assistance.

## References

- Liew CC, inventor; GeneNews, Inc., assignee. Method for the detection of gene transcripts in blood and uses thereof. United States patent US 20040014059. 2004 Jan 22 (priority date January 1999).
- Ma J, Liew CC. Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 2003;35:993–8.
- Bull TM, Coldren CD, Moore M, et al. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004;170:911–9.
- Whistler T, Unger ER, Nisenbaum R, Vernon SD. Integration of gene expression, clinical, and epidemiologic data to characterize chronic fatigue syndrome. *J Transl Med* 2003;1:10.
- Alcorta D, Preston G, Munger W, et al. Microarray studies of gene expression in circulating leukocytes in kidney diseases. *Exp Nephrol* 2002;10:139–49.
- Lampe JW, Stepaniants SB, Mao M, et al. Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol Biomarkers Prev* 2004;13:445–53.
- Tsuang MT, Nossova N, Yager T, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005;133:1–5.
- Grossfeld GD, Carroll PR. Evaluation of asymptomatic microscopic hematuria. *Urol Clin North Am* 1998;25:661–76.
- Herr HW. Natural history of superficial bladder tumors: 10- to 20-year follow-up of treated patients. *World J Urol* 1997;15:84–8.
- Kriegmair M, Baumgartner R, Knuchel R, Stepp H, Hofstadter F, Hofstetter A. Detection of early bladder cancer by 5-aminolevulinic acid induced porphyrin fluorescence. *J Urol* 1996;155:105–10.
- Badalament RA, Hermansen DK, Kimmel M, et al. The sensitivity of bladder wash flow cytometry, bladder wash cytology, and voided cytology in the detection of bladder carcinoma. *Cancer* 1987;60:1423–7.
- Burchardt M, Burchardt T, Shabsigh A, De La Taille A, Benson MC, Sawczuk I. Current concepts in biomarker technology for bladder cancers. *Clin Chem* 2000;46:595–605.
- Herr HW, Bajorin DF. Cancer of the bladder. In: Rosenberg S, editor. *Cancer: principles and practice of oncology*. Philadelphia: Lippincott, Williams, and Wilkins; 2001. p. 1396–18.
- Pampel FC. Logistic regression: a primer. Sage quantitative applications in the social sciences series 132. Thousand Oaks: Sage Publications; 2000.
- 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.
- Yap Y, Zhang X, Ling MT, Wang X, Wong YC, Danchin A. Classification between normal and tumor tissues based on the pair-wise gene expression ratio. *BMC Cancer* 2004;4:72.
- Hofmann HS, Hansen G, Burdach S, Bartling B, Silber RE, Simm A. Discrimination of human lung neoplasm from normal lung by two target genes. *Am J Respir Crit Care Med* 2004;170:516–9.
- Mutaguchi K, Yasumoto H, Mita K, et al. Restoration of insulin-like growth factor binding protein-related protein 1 has a tumor-suppressive activity through induction of apoptosis in human prostate cancer. *Cancer Res* 2003;63:7717–23.
- Sun HZ, Wu SF, Tu ZH. Blockage of IGF-1R signaling sensitizes urinary bladder cancer cells to mitomycin-mediated cytotoxicity. *Cell Res* 2001;11:107–15.
- Miyake H, Hara I, Yamanaka K, Muramaki M, Eto H. Prognostic significance of insulin-like growth factor (IGF) binding protein-2 to IGF-binding protein-3 ratio in patients undergoing radical cystectomy for invasive transitional cell carcinoma of the bladder. *BJU Int* 2005;95:987–91.
- Bieche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. *Clin Cancer Res* 2004;10:6789–95.
- Mohr S, Bottin MC, Lannes B, et al. Microdissection, mRNA amplification and microarray: a study of pleural mesothelial and malignant mesothelioma cells. *Biochimie* 2004;86:13–9.
- Ghiselli G, Siracusa LD, Iozzo RV. Complete cDNA cloning, genomic organization, chromosomal assignment, functional characterization of the promoter, and expression of the murine Bamacan gene. *J Biol Chem* 1999;274:17384–93.
- Iacobuzio-Donahue C, Shuja S, Cai J, Peng P,

- Willett J, Murnane MJ. Cathepsin D protein levels in colorectal tumors: divergent expression patterns suggest complex regulation and function. *Int J Oncol* 2004;24:473–85.
25. Leto G, Tumminello FM, Crescimanno M, Flandina C, Gebbia N. Cathepsin D expression levels in nongynecological solid tumors: clinical and therapeutic implications. *Clin Exp Metastasis* 2004;21:91–106.
26. Klaassens M, van Dooren M, Eussen HJ, et al. Congenital diaphragmatic hernia and chromosome 15q26: determination of a candidate region by use of fluorescent *in situ* hybridization and array-based comparative genomic hybridization. *Am J Hum Genet* 2005;76:877–82.
27. Luce MJ, Burrows PD. The neuronal EGF-related genes NELL1 and NELL2 are expressed in hemopoietic cells and developmentally regulated in the B lineage. *Gene* 1999;231:121–6.
28. Abraham RS, Ballman KV, Dispenzieri A, et al. Functional gene expression analysis of clonal plasma cells identifies a unique molecular profile for light chain amyloidosis. *Blood* 2005;105:794–803.
29. Osman I, Kang M, Lee A, et al. Detection of circulating cancer cells expressing uroplakins and epidermal growth factor receptor in bladder cancer patients. *Int J Cancer* 2004;111:934–9.
30. Twine NC, Stover JA, Marshall B, et al. Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. *Cancer Res* 2003;63:6069–75.