RESEARCH ARTICLE

Comprehensive proteomics and platform validation of urinary biomarkers for bladder cancer diagnosis and staging

Kamala Vanarsa[1†], Jessica Castillo[1†], Long Wang[2†], Kyung Hyun Lee[3], Claudia Pedroza[3], Yair Lotan[4†] and Chandra Mohan[1*†]

Abstract

Background Bladder cancer (BC) is among the most common cancers diagnosed in men in the USA. The current gold standards for the diagnosis of BC are invasive or lack the sensitivity to correctly identify the disease.

Methods An aptamer-based screen analyzed the expression of 1317 proteins in BC compared to urology clinic controls. The top hits were subjected to systems biology analyses. Next, 30 urine proteins were ELISA-validated in an independent cohort of 68 subjects. Three of these proteins were next validated in an independent BC cohort of differing ethnicity.

Results Systems biology analysis implicated molecular functions related to the extracellular matrix, collagen, integrin, heparin, and transmembrane tyrosine kinase signaling in BC susceptibility, with HNF4A and NFKB1 emerging as key molecular regulators. STEM analysis of the dysregulated pathways implicated a functional role for the immune system, complement, and interleukins in BC disease progression. Of 21 urine proteins that discriminated BC from urology clinic controls (UC), urine D-dimer displayed the highest accuracy (0.96) and sensitivity of 97%. Furthermore, 8 urine proteins significantly discriminated MIBC from NMIBC (AUC = 0.75–0.99), with IL-8 and IgA being the best performers. Urine IgA and fibronectin exhibited the highest specificity of 80% at fixed sensitivity for identifying advanced BC.

Conclusions Given the high sensitivity (97%) of urine D-dimer for BC, it may have a role in the initial diagnosis or detection of cancer recurrence. On the other hand, urine IL-8 and IgA may have the potential in identifying disease progression during patient follow-up. The use of these biomarkers for initial triage could have a significant impact as the current cystoscopy-based diagnostic and surveillance approach is costly and invasive when compared to a simple urine test.

Keywords Bladder cancer, Proteomics, Biomarkers, Protein biomarkers, Diagnostic biomarkers, Surveillance biomarkers

†Kamala Vanarsa, Jessica Castillo, and Long Wang have contributed equally and are co-first authors.

†Yair Lotan and Chandra Mohan are co-senior authors.

*Correspondence:
Chandra Mohan
cmohan@central.uh.edu
Full list of author information is available at the end of the article

Vanarsa *et al. BMC Medicine* (2023) 21:133

Page 2 of 17

Background

Bladder cancer (BC) is the fourth most common cancer diagnosed in men in the USA [1]. The incidence rate of the disease is four times higher in men than in women and approximately twice as high in White men compared to Black men [1]. It is estimated that 6% of all new cancer diagnoses in men in the year 2022 will be BC [1]. Overall, 81,180 people are expected to be newly diagnosed with BC in 2022, of which 61,700 being male and 19,480 being female [1]. It is also estimated that 17,100 people will die from the disease in 2022 [1]. The risk of developing BC increases with age with the highest risk being in 80-year-old males and females [2]. The 5-year relative survival rate for those with BC is 77% [1]. If the tumor is non-invasive, the 5-year survival increases to 96% [1]. However, 51% of all cases are diagnosed after this occurrence [1].

The current gold standard for the diagnosis of BC is cystoscopy. However, cystoscopy is often associated with complications including pain, urinary tract infection, and hematuria. Urine cytology is also commonly used for the diagnosis and surveillance of BC. This non-invasive method involves the examination of cells collected from a urine specimen. Research has indicated a high specificity of 86%, but it is constrained by a low sensitivity of 48% [3]. There is also subjectiveness when grading urothelial carcinoma on urine samples thus resulting in poor inter-observer variability [3].

The United States Food and Drug Administration (FDA) has approved 6 urinary assays to use in conjunction with cystoscopy for the surveillance and diagnosis of BC. These include BTA stat, BTA TRAK, NMP22 BladderChek Test, NMP22 ELISA, UroVysion, and uCyt [4]. A meta-analysis of NMP-22 BladderChek shows a pooled specificity of 88% and a sensitivity of 56% for BC detection from 19 research studies [5]. The sensitivity of the test was found to steadily increase with higher stages and grades of the disease. An additional meta-analysis of BTA stat identified a pooled specificity of 67% and a sensitivity of 75% in 13 research studies [6]. Similar to NMP22, BTA stat's sensitivity positively correlates with an increasing grade of BC. Due to false positives and lower specificity values, these tests cannot be used as the sole measure of diagnosis and surveillance. The American Urologic Association guidelines for the evaluation of hematuria and surveillance of bladder cancer do not currently recommend the routine use of urine markers [7, 8].

Given these metrics, there is a need for better biomarkers for BC. Urine biomarkers are promising as a non-invasive test for BC. Urine can be obtained non-invasively, is a readily available biological fluid, and is close to the site of pathology. This allows for repeated tests as deemed necessary for both diagnosis and potential monitoring of disease progression. Urine is also advantageous for potential cost-effective point-of-care tests. Emerging urine point-of-care tests may empower individuals to monitor their health from the comfort of their own homes [9].

As opposed to previous studies in the field examining a handful of proteins selected based on their known properties, here, we report the first and largest use of a comprehensive aptamer-based proteomic screen of urine samples from 42 subjects. This platform has been successfully applied in biomarker screens of several other diseases [10–21]. Additionally, in the present study, we have executed the largest ELISA validation study in BC, interrogating 30 protein biomarkers in an independent cohort consisting of 68 subjects (31 urology clinic controls (UC) and 37 BC (10 Ta, 10 Tis, 10 T1, and 7 T2–T4)). The study has uncovered novel urine protein biomarkers that have not been reported in BC patients before and that out-perform current biomarkers used in clinical practice. The reported urine biomarkers may be useful for the initial diagnosis of BC and possibly for the surveillance of the disease.

Methods

Patient cohorts

Inclusion and exclusion criteria: In all cohorts, the included bladder cancer patients were patients in whom the diagnosis was established by cystoscopy and pathology. Subjects with other malignancies were excluded. Urine samples for the initial aptamer-based screen were obtained from the University of Texas Southwestern Medical Center and Bioreclamation (Bioreclamation, RRID:SCR_004728), Westbury, NY. The samples included 15 urology clinic controls ("UC") and 27 bladder cancer subjects including 5 Ta (non-invasive papillary carcinoma), 4 Tis (flat carcinoma in situ), 9 T1 (tumor spread to connective tissue), 4 T2 (muscle-invasive bladder cancer), 3 T3, and 2 T4 BC. Of these, 18 subjects (stage Ta, Tis, and T1) were classified as non-muscle invasive bladder cancer (NMIBC) while 9 subjects (stage T2–T4) were classified as muscle invasive bladder cancer (MIBC). It should be stressed out that UC does not refer to urothelial carcinoma but for urology clinic controls. Additionally, unless stated otherwise, all bladder cancer subjects included in this study had urothelial cancer. For replication of the findings from the initial proteomic screen, the independent validation cohort for ELISA consisted of samples obtained from the University of Texas Southwestern Medical Center ("UTSW cohort"). These included 31 UC samples and 37 BC samples (10 Ta, 10 Tis, 10 T1, and 7 T2–T4). Of these, 30 subjects (stage Ta, Tis, and T1) were classified as NMIBC while 7 subjects (stage T2–T4) were classified as MIBC. UC samples

included patients investigated for hematuria but found not to have any urological cancers. Subject demographics, including age gender and ethnicity, and clinical information pertaining to these samples are detailed in Table 1. Sex as a biological variable: both genders were included in the study. The secondary validation cohort consisted of samples of Chinese ethnicity. These samples were from the Third Xiangya Hospital of the Central South University in Changsha, China, and comprised 91 BC patients and 77 UC patients (Additional file 1: Table S1). Samples in all cohorts were obtained with informed consent. The study was approved by the institutional review boards at the University of Houston, Houston, TX; UTSW, Dallas, TX; and the Third Xiangya Hospital of the Central South University in Changsha, China. In all cases, urine samples were centrifuged, aliquoted, and stored at −80 °C, and used for the assays without repeated freeze-thaws. Laboratory researchers performing the assays were blinded to the subject groupings.

Aptamer-based targeted proteomic screen of BC urine
The samples for the aptamer-based screen consisted of 15 urology controls and 27 bladder cancer subjects (BC Ta = 5, BC Tis = 4, BC T1 = 9, BC T2 = 4, BC T3 = 3, BC T4 = 2). These urine samples were screened using an aptamer-based screening platform ("SOMAScan") manufactured by Somalogic, as detailed previously [20, 22]. In short, the samples were added to the aptamer-coated beads. SOMAmer-protein binding then occurs. Following this, the unbound proteins are washed off. The remaining bound proteins are biotinylated. SOMAmer-protein complexes are next photocleaved from the beads with UV light. Incubation in a buffer with a polyanionic competitor disrupts non-specific interactions. The SOMAmer-proteins are then recaptured on a second streptavidin-coated bead. Next, the SOMAmer reagents are released from the beads in a denaturing buffer. The released SOMAmers are then hybridized onto a DNA microarray and quantified by the relative fluorescence unit for each protein.

Cross-sectional ELISA validation of urine protein biomarkers
Altogether, 34 proteins were initially selected for ELISA validation based on the aptamer-based screening. Commercially available ELISA kits were purchased, and preliminary testing was conducted. The protein, ELISA manufacturer, optimal urine sample dilution, reason for selection, and outcome of ELISA testing are listed in Additional file 1: Table S2. After preliminary testing,

Table 1 Demographic information pertaining to subjects used for the aptamer-based screen

Variable	Category	UC (N = 15)	Ta (N = 5)	Tis (N = 4)	T1 (N = 9)	T2–T4 (N = 9)
Age[a]		66.2 ± 4.1	66.6 ± 12.1	68.5 ± 4.7	69.3 ± 10.1	71.9 ± 12.4
Gender, *n* (%)	Male	15 (100.0%)	5 (100.0%)	4 (100.0%)	9 (100%)	9 (100%)
Ethnicity, *n* (%)	Caucasian	6 (40.0%)	4 (80.0%)	4 (100.0%)	5 (55.6%)	5 (55.6%)
	Others	1 (6.6%)	N/A	N/A	N/A	N/A
	Unknown	8 (53.3%)	1 (20.0%)	N/A	4 (44.4%)	4 (44.4%)
Demographic information pertaining to the ELISA validation cohort (UTSW cohort)						
Variable	Category	UC (N = 31)	Ta (N = 10)	Tis (N = 10)	T1 (N = 10)	T2–T4 (N = 7)
Age[a]		66.2 ± 10.1	71.3 ± 11.9	71.5 ± 6.3	67.4 ± 9.9	74.3 ± 7.9
Gender, *n* (%)	Male	27 (87.1%)	8 (80%)	10 (100%)	8 (80%)	7 (100%)
	Female	4 (12.9%)	2 (20%)	N/A	2 (20%)	N/A
Ethnicity, *n* (%)	Caucasian	23 (74.2%)	8 (80%)	9 (90%)	8 (80%)	6 (85.7%)
	Black	4 (12.9%)	N/A	N/A	1 (10%)	1 (10%)
	Asian	2 (6.5%)	1 (10%)	N/A	1 (10%)	N/A
	Hispanic	N/A	1 (10%)	N/A	N/A	N/A
	Others	1 (3.2%)	N/A	N/A	N/A	N/A
	Unknown	1 (3.2%)	N/A	1 (10%)	N/A	N/A

Top: Demographic information pertaining to the 42 subjects included in the aptamer-based screen. NMIBC comprised Ta, Tis, and T1 subjects. MIBC comprised T2–T4 BC. Subjects in the UC category comprised of BPH (*n* = 1), elevated PSA (*n* = 1), erectile dysfunction (*n* = 2), HLD (*n* = 2), allergic rhinitis (*n* = 2), hypothyroidism (*n* = 2), dementia (*n* = 1), insomnia (*n* = 1), impotence (*n* = 1), osteopenia (*n* = 1), thrombocytopenia (*n* = 1), hemangioma (*n* = 1), glaucoma (*n* = 1), and gout (*n* = 1)

Bottom: Demographic information pertaining to the 68 subjects used for the independent ELISA validation cohort UTSW Cohort. NMIBC comprised Ta, Tis, and T1 subjects. MIBC comprised T2–T4 BC. UC subjects comprised patients with various urological conditions including bladder diverticulum (*n* = 1), bladder stones (*n* = 1), BPH (*n* = 7), cystocele (*n* = 1), elevated PSA (*n* = 2), erectile dysfunction (*n* = 3), hematuria (*n* = 3), UTI (*n* = 1), hydrocele (*n* = 1), incomplete bladder emptying (*n* = 1), impotence (*n* = 1), kidney stones (*n* = 3), nocturia (*n* = 2), prostate cancer (*n* = 1), and stress incontinence (*n* = 1)

[a] Age displayed as mean ± standard deviation

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 4 of 17

30 protein biomarkers were assayed in an independent (UTSW) cohort which consisted of 31 UC samples and 37 BC samples (10 Ta, 10 Tis, 10 T1, and 7 T2–T4). The absolute levels of urine proteins were creatinine normalized. Secondary validation was completed for 3 protein biomarkers using an independent cohort comprising 91 BC patients and 77 patients of Chinese ethnicity. The demographic information and clinical information pertaining to the 168 subjects whose urine samples were used for the second ELISA validation are displayed in Additional file 1: Table S1. In this cohort, most patients classified as BC were diagnosed to have bladder cancer except for 7 patients with urothelial cancers (including ureteric cancer). After validation, the absolute levels of urine proteins were normalized by creatinine. The ELISA assay protocols are detailed as has been shown before [20].

D-dimer assay used in the Chinese cohort

As opposed to the ELISA used to assay D-dimer in the other cohorts, an agglutination assay was used to assay D-dimer in the Chinese cohort. A D-dimer detection kit (latex immunity ratio turbidity method) from SEKISUI was used to quantitatively detect the concentration of D-dimer in the Chinese cohort. Briefly, the D-dimer in the sample reacts with the monoclonal antibody to the mouse anti-human D-dimer. This causes agglutination and increased turbidity. Following this, the concentration of D-dimer was determined by measuring the variation of turbidity with a spectrophotometer.

Data analysis of the aptamer-based screening and ELISA results

Aptamer-based screening data was subjected to hybridization, median normalization, and creatinine normalization as detailed previously [20]. Further data analysis was completed in R version 1.4.1103 with packages readxl, readr, qvalue, and stats. A non-parametric Mann–Whitney U-test was used to identify the proteins that were significantly up- or downregulated among the subject groups. Statistical p-values were computed for each biomarker. To address multiple testing correction, q-values were calculated to adjust for the false discovery rate. Fold change values were also computed to determine the ratio of protein expression from diseased to control subjects (BC/UC) and MIBC to NMIBC subjects (T2–T4 vs Ta, Tis, T1). ROC analyses including AUC, cutoff, sensitivity, and specificity values were computed with easyROC version 1.3.1 [23]. The biomarker ELISA data was plotted and analyzed using GraphPad Prism 5 (GraphPad Prism, RRID:SCR_002798). Group comparisons were analyzed using either a non-parametric Mann–Whitney U-test or

a Kruskal–Wallis test with Dunn's multiple comparison. Statistical p-values from analyses were computed.

Gene Ontology and KEGG functional enrichment analysis

Gene Ontology (GO) and KEGG (KEGG, RRID:SCR_012773) functional enrichment analysis was completed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 8 (DAVID, RRID: SCR_001881). The top 330 proteins with a Mann–Whitney p-value < 0.05 in the aptamer-based screen (BC versus UC) were used for analysis. The top 10 biological processes, molecular functions, and KEGG pathways were plotted using R. The packages used include readxl and ggplot2 (ggplot2, RRID:SCR_014601). The size of the dots represents the count/hit number of genes belonging to the annotation term, and the color of the dots represent $-\log_{10}$FDR value.

Protein–protein interaction networks and regulatory networks

Protein–protein interaction networks for the top 330 proteins in the aptamer-based screen (BC vs UC, Mann–Whitney p-value < 0.05) were created using Cytoscape version 3.9.0 (Cytoscape, RRID:SCR_003032) using the stringApp. MCODE clustering was preformed to discern highly interconnected nodes in the network. The top 3 clusters are plotted with their associated Reactome pathways. The top transcription factor and signaling molecule regulator were identified for the top 330 proteins in the aptamer-based screening (BC vs UC, Mann–Whitney p-value < 0.05) using the iRegulon plugin available through Cytoscape. The color of each node corresponds to the fold change. Nodes with a fold change of less than 1 range in color from blue to purple while those with a fold change greater than 1 range from pink to red, when comparing BC to UC.

Volcano plot, principal component analysis (PCA), and correlation plot

The volcano, PCA, and correlation plots were created in R using the readr, readxl, gplots, ggplot2, ggplot. multistats, scatterplot3d, Hmisc, data.table, and corrplot packages. All 1317 proteins are represented in the volcano plot. The data was log-transformed, and a Mann–Whitney U-test (BC vs UC) was used to generate statistical p-values. A 2D PCA plot was generated for the top 119 proteins (BC vs UC, Mann–Whitney q-value < 0.05) where the two first principal components are plotted. Subject groups are differentiated by color and/or shape. A correlation plot for the top 50 proteins was also generated (Mann–Whitney p-value < 0.05, ordered by fold change, comparing BC to UC). Correlation coefficients for all possible protein pairs were computed using Pearson's correlation coefficients.

Heatmap analysis

Heatmaps were generated from the aptamer-based screening assay in order to cluster proteins with similar expression profiles together. Proteins significantly elevated in BC when compared to urology controls (p-value < 0.05 and a fold change > 2) were analyzed. Hierarchical clustering was performed in R. Each row corresponds to the creatinine-normalized protein level measured, and each column represents a sample (UC = 15, BC Ta = 5, BC Tis = 4, BC T1 = 9, BC T2 = 4, BC T3 = 3, BC T4 = 2). Proteins that are above the mean value for each biomarker are shaded yellow. Proteins comparable to the mean are shaded black. Those below the mean are shaded blue.

Random forest analysis

Random forest analysis, a supervised machine learning algorithm, was conducted for the purpose of identifying the relative importance of biomarker candidates in disease discrimination. The randomForest R package and the top 93 proteins identified from the aptamer-based screen (BC vs UC, Mann–Whitney $p < 0.05$ and FC > 2) were used for analysis. The importance of each biomarker was measured using the GINI coefficient. The resulting top 10 potential urine biomarkers identified by random forest classification are plotted.

Bayesian network analysis

Bayesian network analysis was executed with the BayesiaLab software. This method uses probability distributions to represent the inter-dependencies between all variables in a model and how they relate to one another. The dataset comprised 68 subjects including 31 UC and 37 BC (10 Ta, 10 Tis, 10 T1, 7 T2–T4) subjects. The urine levels of 21 protein biomarkers, features (age, gender, ethnicity), and disease status (BC vs UC) were examined. The network was constructed in an unsupervised manner with the EQ algorithm. The size of each node was determined using node force and is proportional to its impact on the other nodes in the network. The arcs that interconnect the nodes were determined with Pearson's correlation coefficient. The interconnections between nodes represent the dependencies among the variables including the correlation coefficient between nodes. The thickness of the arcs is proportional to Pearson's correlation coefficient.

STEM analysis

Short Time-series Expression Miner (STEM) version 1.3.13 was utilized for clustering, comparing, and visualizing protein expression data from the aptamer-based screen across bladder cancer (BC Ta-Tis = 9, BC T1 = 9, BC T2–T4 = 9) and urology controls (UC = 15). Protein expression among the top 330 proteins (BC vs UC, Mann–Whitney p-value < 0.05) were averaged across the subject groups analyzed. The number in the upper left-hand corner of each box is representative of the number of genes in each cluster. The number in the lower left-hand corner is the p-value significance of the number of proteins assigned to the cluster versus what was expected based on permutation testing. Protein expression profiles through UC and progressive BC stages, Ta-Tis, T1, and T2–T4 are plotted. Statistically significant profiles that are similar form a cluster and are shaded the same color. A total of 50 cluster patterns were generated by STEM analysis, of which only the statistically significant clusters that exhibited a progressive increase or decrease of urine biomarkers across BC stages are plotted. The associated Gene Ontology (GO) enrichment analysis for profiles of proteins with the same expression patterns was determined. The Reactome pathways associated with each significant cluster were identified through Cytoscape and the functional enrichment tool.

Multi-biomarker panels after adjusting for age, gender, and ethnicity

The protein level of each protein assayed by ELISA and the age of patients were standardized to have a mean of zero and a standard deviation of one, after applying $\log_2$ transformation. The proteins that best discriminate classes of subjects (BC vs UC and MIBC vs NMIBC) were identified using the predictive projection feature selection technique executed with the projpred package in R (version 4.0.3.) [24–26]. Model selection was performed using a model with the best predictive power (a reference model) in order to find a model with a smaller number of proteins. The smaller model should maintain comparable prediction performance when compared to the reference model (predictive projection). The selection process consisted of two main components. During the first step, a Bayesian regularized logistic regression model with horseshoe prior [27, 28] was fitted as a reference model. During the second step, a projected submodel with at most 5 proteins that minimized the Kullback–Leibler divergence from the posterior distribution of the reference model to that of the projected model was searched for. The selected submodel was found to exhibit similar predictive performance determined by the mean log predictive density and the mean square error. Both of the performance metrics in addition to the area under the curve and prediction accuracy were assessed through leave-one-out cross-validation in an effort to bypass potential problems resulting from overfitting [29]. The selected proteins of one model and its performance

metrics were compared to those of the counterpart model with adjustments for age, ethnicity, and gender to account for potential confounding factors from these variables.

Results

Aptamer-based targeted proteomic screen of BC urine

An overview of the study flow is depicted in Additional file 1: Fig. S1. A comprehensive aptamer-based screen of urine samples from 42 human subjects was executed to interrogate the levels of 1317 proteins, as detailed in the "Methods" section. This study group included 27 bladder cancer subjects and 15 UC (Table 1). A non-parametric Mann–Whitney *U*-test was used to identify proteins that were significantly up- or downregulated among BC vs

UC, resulting in 330 proteins. After multiple testing correction, 119 of these 330 proteins had a *q*-value < 0.05. The 330 proteins found to be statistically significantly different (both up- and downregulated proteins) in BC were subjected to functional pathway enrichment.

Functional pathway enrichment was performed using Gene Ontology analysis. The top 10 Gene Ontology biological processes identified from the differentially expressed urine proteins in BC can be found in Fig. 1A. Extracellular matrix organization, cytokine signaling, inflammatory response, and angiogenesis were identified as some of the most significant biological processes associated with the dysregulated proteins in BC. The top Gene Ontology molecular functions that were enriched included binding to collagen, integrin, heparin, and transmembrane tyrosine
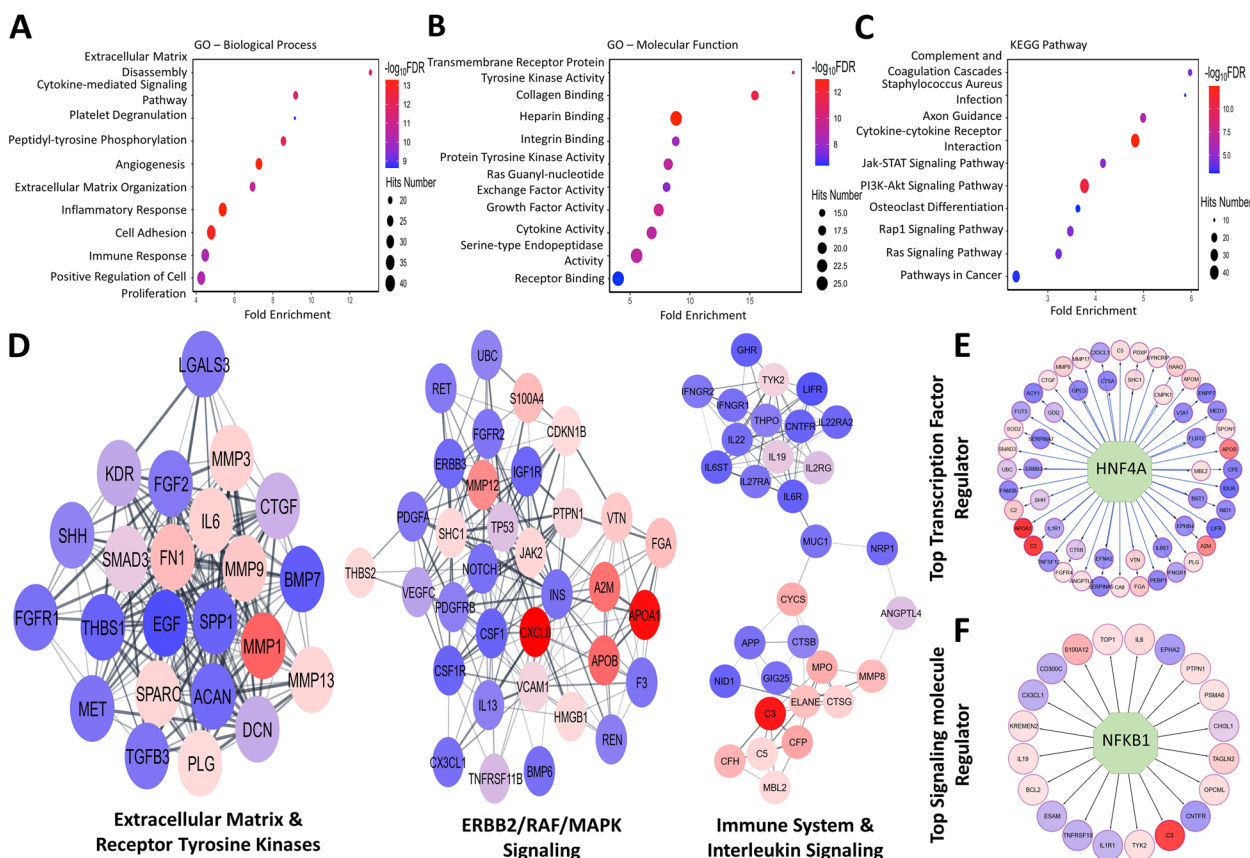

Fig. 1 Functional pathway enrichment analysis of proteins dysregulated in bladder cancer urine using Gene Ontology, KEGG pathway, and protein–protein interaction networks. All 330 proteins with a Mann–Whitney *p*-value < 0.05 (BC versus UC) in the aptamer-based screen were used for functional pathway enrichment. **A**–**C** The top 10 Gene Ontology biological process, molecular functions, and KEGG pathways obtained through GO are plotted respectively based on *p*-value significance in the order of fold enrichment. The size of the dots represents the count/hit number of genes belonging to the annotation term, and the color of the dots is representative of − log$_{10}$FDR value. **D** Protein–protein interaction networks for the top 330 proteins (BC vs UC, Mann–Whitney *p*-value < 0.05) were created using the Cytoscape stringApp. MCODE clustering was preformed to find highly interconnected nodes in the network. The top 3 clusters are plotted with their associated Reactome pathways. The color of each node corresponds to the fold change. Nodes with a fold change of less than 1 (reduced in BC) range in color from blue to purple while those with a fold change greater than 1 (increased in BC) range from pink to red. **E** The top transcription factor regulator regulating the differentially expressed proteins was identified using the iRegulon plugin available for Cytoscape. Nodes with a fold change of less than 1 range in color from blue to purple while those with a fold change greater than 1 range from pink to red

kinase receptors (Fig. 1B). KEGG pathway analysis was next executed for these 330 proteins to identify implicated pathways (Fig. 1C). Cytokine-cytokine receptor interaction and PI3K-AKT signaling pathways encompassed the largest number of proteins under these annotation terms.

Following this, protein–protein interaction networks were created with Cytoscape using the 330 differentially expressed proteins in BC to determine their interactions with one another. MCODE clustering was performed to identify highly interconnected clusters (Fig. 1D). Node color is representative of fold change, with downregulated proteins in BC urine shaded blue and upregulated proteins in BC urine shaded red. Reactome pathways enriched among the first cluster include extracellular matrix and receptor tyrosine kinases. The second cluster

encompasses ERBB2/RAF/MAPK signaling pathways, while the third cluster is enriched with the immune system and interleukin signaling pathways. HNF4A was identified as the top-most transcription factor controlling the differentially expressed proteins in BC (Fig. 1E), while NFKB1 was singled out as the topmost signaling molecule regulating these proteins (Fig. 1F), as determined using the iRegulon plugin for Cytoscape.

Of the 1317 proteins assayed in the aptamer-based screen, 93 urine proteins were found to be elevated in BC vs UC (Mann–Whitney p-value < 0.05) at a fold change of > 2, as depicted by the volcano plot in Fig. 2A. Of these 93 proteins, 7 were found to be significantly elevated with a Mann–Whitney p-value < 0.001 and a fold change of > 5 (represented as red dots in Fig. 2A). The top 119 proteins (BC vs UC, q-value < 0.05) were used as input
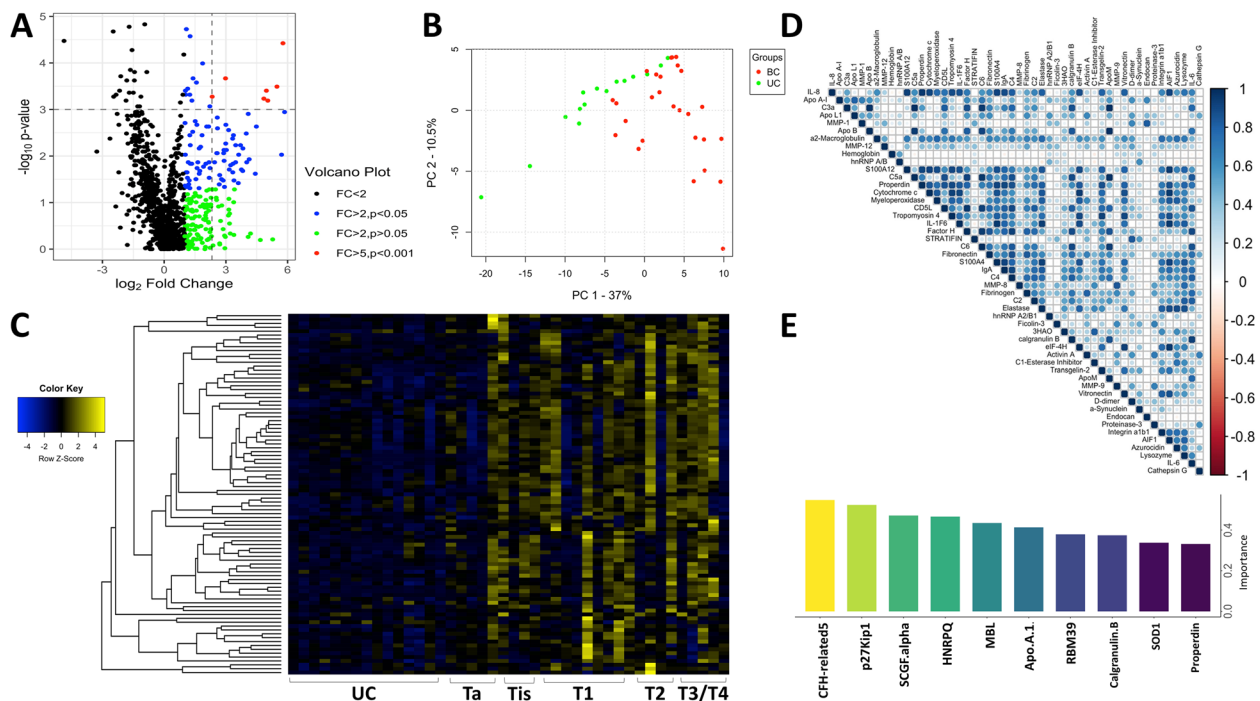
Fig. 2 Aptamer-based proteomic screening of bladder cancer urine uncovers several clusters of up- and downregulated proteins. **A** A volcano plot representation of the results of the aptamer-based screening of 1317 proteins analyzed in 42 urine samples (15 UC, BC (Ta-Tis) = 9, BC (T1) = 9, BC (T2–T4) = 9). Data was log-transformed and analyzed as detailed in the "Methods" section. Of the 330 proteins that were differentially expressed between the groups, 93 proteins were elevated at fold change > 2 in BC when compared to UC. Each dot represents one of the 1317 proteins. The x-axis plots the $\log_2$ transform of the fold change. The y-axis displays the $-\log_{10}$ transform of the p-value. **B** A 2D PCA plot of all subjects, using the 119 proteins that were differentially expressed after multiple testing corrections (BC vs UC, Mann–Whitney q-value < 0.05). Bladder cancer is represented by a red circle while urology control is represented by a green circle. The first three principal components are displayed on each axis of the plot. **C** A heatmap representation of the results of the aptamer-based screen displaying the top 93 proteins (BC vs UC, Mann–Whitney p-value < 0.05, fold change > 2) elevated in BC urine. Hierarchical clustering was performed. Each row corresponds to the creatinine-normalized protein level measured, and each column represents a patient sample (UC = 15, BC Ta = 5, BC Tis = 4, BC T1 = 9, BC T2 = 4, BC T3 = 3, BC T4 = 2). Proteins that are above the mean value for each biomarker and shaded yellow. Those below the mean are shaded blue. Proteins comparable to the mean are shaded black. **D** Correlation plot displaying the expression profiles of the upregulated proteins in BC across the entire cohort. Pearson's and Spearman's correlation coefficients were determined for each pair. The proteins were ordered based on hierarchical clustering. Each circle represents the correlation for a protein pair. Blue corresponds to a positive correlation while red corresponds to a negative correlation. **E** Random forest analysis using the top 93 proteins (BC vs UC, Mann–Whitney p-value < 0.05, fold change > 2) identified the 10 most discriminatory urine proteins with the greatest impact on distinguishing BC subjects from urology controls. These 10 proteins are ordered by their GINI coefficient (importance in discrimination)

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 8 of 17

for principal component analysis (PCA), an unsupervised machine learning algorithm, which successfully discriminated BC patients from the urology clinic controls (Fig. 2B). The first two principal components are displayed, with BC and UC samples being represented by a red and green circle, respectively. Additional file 1: Fig. S2A. and S2B display additional PCA plots for all expressed proteins and all 330 differentially expressed proteins, respectively.

A heatmap clustering of the 93 proteins significantly elevated in BC is shown in Fig. 2C. Proteins with upregulated expression are colored yellow while those downregulated are colored blue. To help select proteins for further ELISA validation, the expression profiles of the top 50 upregulated proteins (based on fold change) were next clustered using a correlation matrix (Fig. 2D). A correlation plot of the 93 significantly elevated proteins in BC was also generated (Additional file 1: Fig. S3). Several of these urine proteins were highly correlated with each other, as marked by distinct subclusters of urine proteins that became evident. As an independent approach to identify the most discriminatory proteins, a machine learning approach was used. Specifically, random forest analysis (RFA) of the top 93 proteins (BC vs UC, Mann–Whitney p-value < 0.05, fold change > 2) identified the 10 most discriminatory urine proteins with the greatest impact on distinguishing BC from UC (Fig. 2E), ordered by their importance as represented by their mean decrease in Gini coefficient. Both the correlation clusters (Fig. 2D) and the RFA (Fig. 2E) were used to select proteins for subsequent ELISA validation.

Urine biomarkers for distinguishing bladder cancer stages, based on the aptamer-based screen

Next, we evaluated if the urine proteins identified in the aptamer screen were able to distinguish earlier BC stages from later stages. Urine protein levels in NMIBC were compared to the corresponding levels in MIBC, as summarized by the volcano plot in Fig. 3A, which plots statistical significance (y-axis) versus biological significance (x-axis). With this comparison, 8 urine proteins were found to be elevated in MIBC stages compared to NMIBC (MIBC vs NMIBC, Mann–Whitney p-value < 0.05, fold change > 2), as represented by the blue dots. Principal component analysis (PCA) demonstrated that the differentially expressed proteins in BC were successful in discriminating BC patients with more advanced disease (MIBC), from the urology clinic controls (Fig. 3B). Additional PCA plots display all three subject groups (UC, NMIBC, and MIBC) for all expressed proteins and all 330 differentially expressed proteins (Additional file 1: Fig. S2C and S2D).

A Venn diagram representation of the significantly elevated proteins identified through the aptamer-based screen is shown in Fig. 3C. More proteins were found to be significantly elevated in BC compared to UC (BC vs UC, Mann–Whitney p-value < 0.05) with 93 upregulated at a fold change of 2, and 147 upregulated at a fold change of 1. In contrast, 8 proteins were identified as being significantly elevated with a fold change of 2 or greater, and 12 proteins obtained a fold change greater than 1 in MIBC compared to NMIBC (Mann–Whitney p-value < 0.05). Four urine proteins overlapped between these two comparisons with a fold change greater than 1. These four overlapping proteins included urinary Elastase, S100A12, p53, and Kallikrein 6.

Next, to identify urine proteins and functional pathways that progressively increased or decreased with the worsening BC stage, Short Time-series Expression Miner (STEM) analysis was executed. This analysis identified 2 distinct clusters of proteins that progressively increased with BC stage (clusters 1 and 2; left 2 boxes in Fig. 3D) and 4 clusters of proteins that decreased with worsening BC stage (rightmost 4 boxes in Fig. 3D). The mean expression profiles of the proteins assigned to each plot are represented by the black line within each box. Functional pathway enrichment using Reactome indicated that the proteins in cluster 1 (27 proteins) and those in cluster 2 (29 proteins) were significantly enriched with pathways related to the immune system, complement cascade, and interleukin signaling. Reactome or KEGG functional pathways identified by the proteins in clusters 3–6 included MAPK signaling, cytokine-cytokine receptor interaction, Rap 1 signaling, interleukin signaling, EPH-ephrin mediated repulsion of cells, and signaling by receptor tyrosine kinases.

Validation of urine protein biomarkers in BC urine using an additional assay platform, ELISA, using an independent patient cohort

Based on the correlation clustering (Fig. 2D) and random forest analysis (Fig. 2E) of the urine proteins identified using the aptamer-based screen, 34 proteins were selected for ELISA validation, which represents a different platform than the one used for the initial screen. The selected proteins, ELISA manufacturer, urine dilution, reason for protein selection, and outcome of the ELISA are listed in Additional file 1: Table S2. Of the 34 proteins selected, 30 proteins were advanced forward for validation in the first independent cohort based on preliminary ELISA results.

These 30 proteins were assayed by ELISA in a cohort of 68 urine samples, drawn from 31 UC, and 37 BC patients, comprising 10 Ta, 10 Tis, 10 T1, and 7 T2–T4 patients, referred to in this manuscript as the UTSW cohort (Table 1). The BC vs UC comparison of the ELISA results is detailed in Fig. 4A which includes all of the data for the 30 proteins. This data was creatinine normalized.
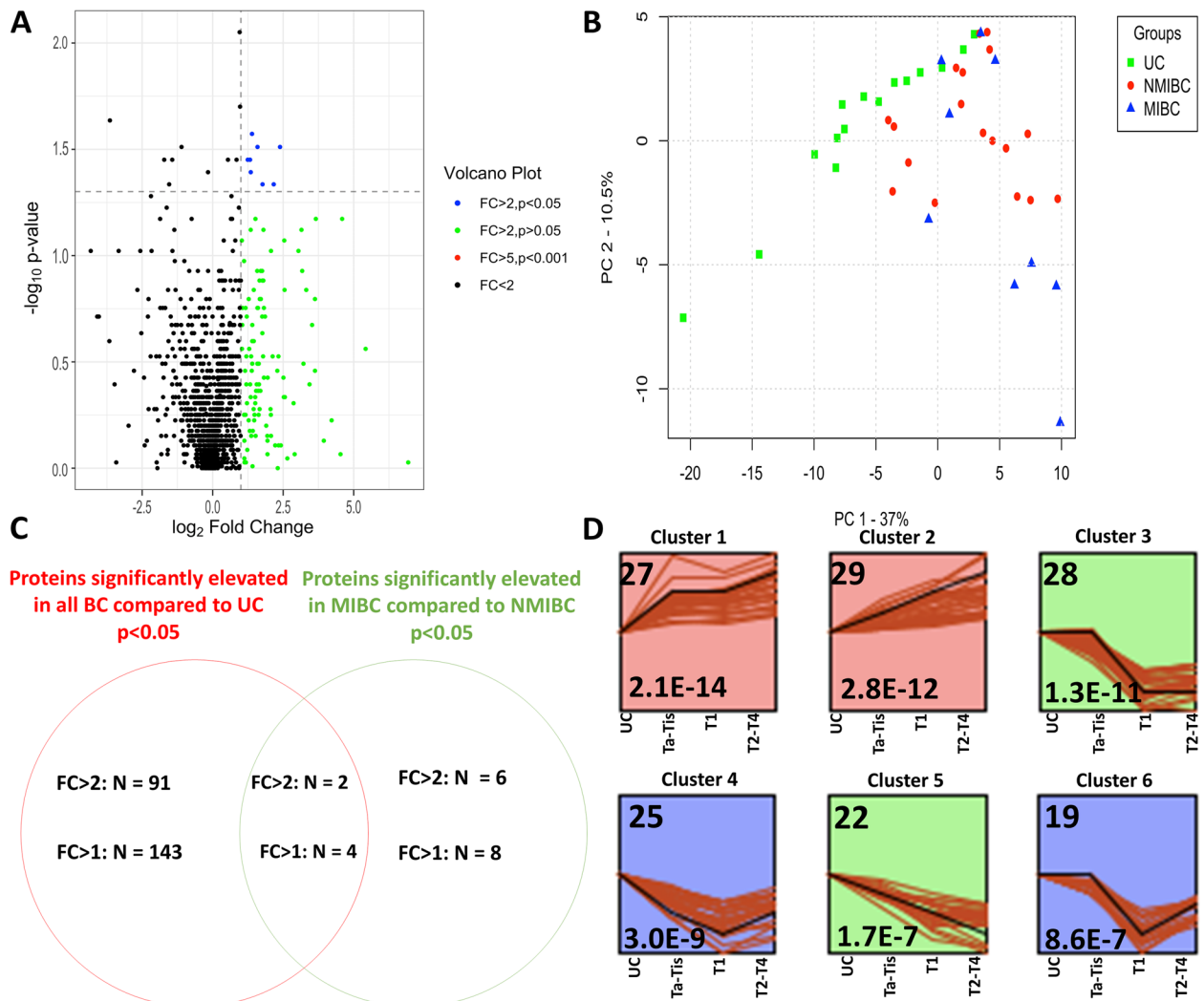
Fig. 3 Urine proteins that discriminate BC stages, based on the aptamer-based proteomic screen. **A** A volcano plot representation of the results of the aptamer-based screening of 1317 proteins analyzed in 27 urine samples, comparing disease stages (NMIBC = 18, MIBC = 9). Data was log-transformed and analyzed as detailed in the "Methods" section. Eight proteins were found to be elevated (Mann–Whitney *p*-value < 0.05, fold change > 2) in MIBC when compared to NMIBC. Each dot represents one of the 1317 proteins. The *x*-axis shows the $\log_2$ transform of the fold change. The *y*-axis displays the $-\log_{10}$ transform of the *p*-value. **B** A 2D PCA analysis of all subjects using the top 119 urine proteins (BC vs UC, Mann–Whitney *q*-value < 0.05). NMIBC are represented by a red circle while MIBC is represented by a blue triangle. Urology control is represented by a green square. The first three principal components are displayed on each axis of the plot. **C** A Venn diagram comparison of the number of proteins significantly elevated in bladder cancer versus urology control compared to proteins significantly elevated in MIBC compared to NMIBC urine. The data for two different fold change cutoffs are included. **D** Short Time-series Expression Miner (STEM) analysis was completed for the top 330 proteins (BC vs UC, Mann–Whitney *p*-value < 0.05). The number in the upper left-hand corner of each box is the number of proteins in each cluster. The number in the lower left-hand corner is the *p*-value significance of the number of proteins assigned to the cluster versus what was expected based on permutation testing. The protein expression through UC, and progressive BC stages, Ta-Tis, T1, and T2–T4, is plotted. Statistically significant profiles that are similar form a cluster of profiles and are shaded the same color. A total of 50 profiles or clusters (each representing a different pattern) were generated by STEM analysis, of which only the statistically significant clusters that exhibited a progressive increase (clusters 1 and 2) or decrease (clusters 3–6) of urine biomarkers across BC stages are plotted

A box plot view depicting the expression profile of each protein is displayed in Additional file 1: Fig. S4. The eight urine proteins with the highest area under the curve (AUC) for discriminating BC from UC include Apo A1, complement C2, Calgranulin B, D-dimer, IgA, MMP-1, MMP-9, and Properdin. Their AUC values ranged from 0.84 to 0.96. Figure 4B displays the dot plot views for these eight urine proteins. Of all ELISA-tested proteins, urine D-dimer was best at discriminating BC from UC (AUC = 0.96, sensitivity = 95%; specificity = 90%), with

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 10 of 17

A. Urinary Proteins validated by ELISA comparing BC to UC

| Protein	UC (mean/median)	BC (mean/median)	Fold Change	Cut Off	AUC	AUC CI	Specificity	Sensitivity	Sensitivity[0.8]
14-3-3 sigma[p]	45.5(12.2)	32.1(12.9)	0.7	16.3	0.50	0.36-0.64	0.74 (0.55-0.88)	0.43 (0.27-0.61)	0.22
α2 Macroglobulin[n]	10666.3(1032.9)	27085.6(9119.2)	2.5****	2241.9	0.78****	0.67-0.90	0.81 (0.63-0.93)	0.78 (0.62-0.90)	0.78
α-synuclein[n]	1.8(0.7)	4.0(0.8)	2.2	0.9	0.55	0.41-0.69	0.74 (0.55-0.88)	0.49 (0.32-0.66)	0.32
Apolipoprotein A1[n]	3.1(0)	247.1(118.3)	79.1****	10.1	0.91****	0.83-0.98	0.97 (0.83-1.00)	0.84 (0.68-0.94)	0.86
Apolipoprotein L1[n]	94.9(13.7)	615.8(184.9)	6.5**	24.5	0.72***	0.60-0.84	0.65 (0.45-0.81)	0.73 (0.56-0.86)	0.54
C2[n]	12.7(1)	523.1(42.9)	41.2****	5.7	0.84****	0.74-0.93	0.81 (0.63-0.93)	0.81 (0.65-0.92)	0.81
Calgranulin B[n]	138439.6(2650.2)	110696.2(44584.8)	0.8****	14139.2	0.85****	0.75-0.95	0.81 (0.63-0.93)	0.84 (0.68-0.94)	0.84
CLU[p]	183900.5(55358.3)	334394.9(190605.7)	1.8***	94372.3	0.74****	0.62-0.87	0.74 (0.55-0.88)	0.76 (0.59-0.88)	0.62
D-dimer[p]	100.6(1)	13561.8(4301.2)	134.8****	278.1	0.96****	0.92-1.00	0.90 (0.74-0.98)	0.95 (0.82, 0.99)	0.97
Elastase[n]	8.7(0.1)	39.1(5.7)	4.5****	3.8	0.79****	0.68-0.90	0.934 (0.79-0.99)	0.65 (0.47, 0.80)	0.65
Endocan[p]	7.5(1)	1134.7(18.2)	151.8****	10.9	0.74****	0.62-0.86	0.90 (0.74-0.98)	0.60 (0.42, 0.75)	0.62
Ficolin-3[n]	346.4(124.5)	376.7(168.3)	1.1	191.2	0.56	0.42-0.70	0.81 (0.63-0.93)	0.50 (0.29-0.63)	0.46
Fibronectin[n]	807.9(49.2)	1033.7(386.9)	1.3***	99.2	0.77****	0.65-0.88	0.64 (0.45-0.81)	0.84 (0.70-0.94)	0.54
HCC-1[p]	78.3(0)	3377.9(9.5)	43.2***	1.8	0.71***	0.60-0.82	0.84 (0.66-0.95)	0.62 (0.45-0.78)	0.62
IgA[n]	532.3(117.6)	2727.7(1521.8)	5.1****	312.7	0.85****	0.76-0.95	0.87 (0.70-0.96)	0.81 (0.65-0.92)	0.84
IL-8[p]	499.3(350.8)	365.1(1098.5)	7.4****	627.4	0.80****	0.70-0.91	0.84 (0.66-0.95)	0.73 (0.56-0.86)	0.73
Importin[n]	4.5(0.2)	11.6(3.6)	2.6**	0.7	0.71**	0.58-0.84	0.68 (0.49-0.83)	0.81 (0.65-0.92)	0.30
ITGB1[n]	32.4(11.7)	15.3(9.7)	0.5	13.5	0.46	0.32-0.60	0.71 (0.52-0.86)	0.43 (0.27-0.61)	0.30
Lysozyme[n]	145.7(18.4)	303.6(90.8)	2.1****	27.1	0.81****	0.70-0.92	0.71 (0.52-0.86)	0.89 (0.75-0.97)	0.73
M2-PK[n]	16.7(5.7)	13.6(8.5)	0.8	6.8	0.61	0.48-0.75	0.61 (0.42-0.78)	0.62 (0.45-0.78)	0.27
MMP-1[p]	1(1)	16861.9(1754.2)	16861.9****	63.1	0.89****	0.70-0.91	1.00 (0.89-1.00)	0.78 (0.62-0.90)	0.83
MMP-12[n]	0.2(0.1)	7.1(0.8)	38.5****	0.1	0.80****	0.70-0.91	0.84 (0.66-0.95)	0.76 (0.59-0.88)	0.77
MMP-9[p]	31065.8(2429.0)	580512.7(101655.7)	18.7****	16931.4	0.85****	0.76 -0.95	0.87 (0.70-0.96)	0.76 (0.59-0.88)	0.81
Properdin[p]	147.8(1)	32692.8(7098.4)	221.1****	1211.6	0.89****	0.82-0.96	0.97 (0.83-1.00)	0.78 (0.62-0.90)	0.83
Proteinase 3[p]	663.9(49.3)	798.5(271.5)	1.2****	85.9	0.81****	0.70-0.92	0.77 (0.59-0.90)	0.81 (0.65-0.92)	0.70
S100A12[n]	170.5(1)	288.7(1)	1.7	132.8	0.51	0.41-0.61	0.87 (0.70-0.96)	0.19 (0.08-0.35)	0.22
S100A4[n]	2.5(0.9)	11.8(1.1)	4.7	1.2	0.55	0.41-0.69	0.77 (0.59-0.90)	0.49 (0.32-0.66)	0.41
SOD1[p]	87977.1(22695.1)	138999.9(42063.7)	1.6*	31671.7	0.64*	0.51-0.78	0.65 (0.45-0.81)	0.65 (0.47-0.80)	0.30
SPARC[p]	6.9(2.7)	16.6(4.4)	2.4	8.7	0.64*	0.50-0.77	0.934 (0.79-0.99)	0.38 (0.22-0.55)	0.46
TXD12[n]	367.4(149.2)	648.5(173.1)	1.8	195.5	0.57	0.43-0.71	0.81 (0.63-0.93)	0.49 (0.32-0.66)	0.49
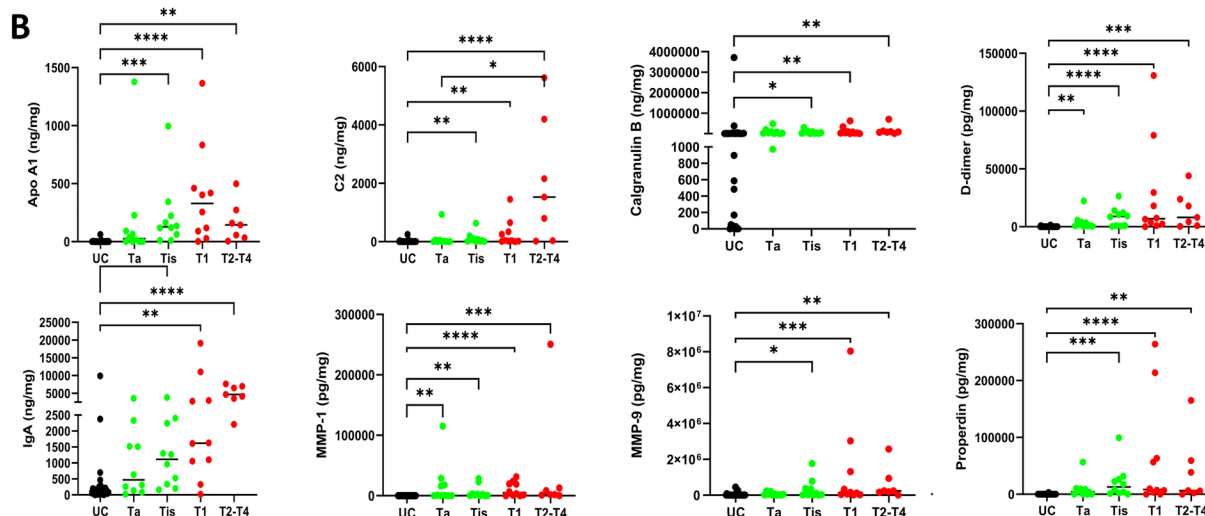
Fig. 4 The ability of 30 ELISA-validated urine proteins to discriminate bladder cancer patients from urology clinic controls. **A** Thirty urinary proteins were assayed in bladder cancer and UC by ELISA, using the UTSW cohort, comprising 31 UC, 10 Ta, 10 Tis, 10 T1, and 7 T2-T4 BC patients. The ELISA results for BC vs UC are displayed. The urology clinic controls (UC) comprise urology clinic control patients without urological cancers, as detailed in the "Methods" section. Biomarker protein units (normalized to creatinine) are as follows: n = ng/mg, p = pg/mg. CI: confidence interval. **A** Delong CI for AUC and a Clopper-Pearson CI for sensitivity and specificity is displayed. Sensitivity.[0.8] depicts the sensitivity at a fixed specificity value of 0.8. Indicated are the statistical significance p-values *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, comparing BC to UC. **B** The dot plots depict the 8 proteins with the highest ROC AUC accuracy values for discriminating BC vs UC. Creatinine normalized urine protein levels are plotted in contrasting colors representing UC, BC Ta, Tis, and T1–T4 stages. Statistical analyses of plots were carried out using a Kruskal-Walls test with Dunn's multiple comparison post hoc test. The asterisks indicate the level of significance between the groups: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$

urine properdin and MMP-1 being close behind. As a sensitivity analysis, a bootstrap logistic regression model was used to derive optimism-corrected performance metrics and to ascertain the robustness of each urine protein for distinguishing BC from urology control. This method, which is more accurate for small sample sizes, yielded similar results, as listed in Additional file 1: Table S3.

## A. ELISA validation of Urine Proteins that distinguish MIBC vs. NMIBC

| Protein	NMIBC (mean/median)	MIBC (mean/median)	Fold Change	Cut Off	AUC	AUC CI	Sensitivity (CI)	Specificity (CI)	Specificity[0.8]
14-3-3 sigma[p]	36.3(12.5)	14.3(12.9)	0.4	10.4	0.44	0.22-0.66	0.71(0.29-0.96)	0.47(0.28-0.66)	0.20
α2 Macroglobulin[n]	27502.0(4954.6)	25301.5(14389.6)	0.9	6418.3	0.72**	0.55-0.89	1(0.59-1)	0.53(0.34-0.72)	0.60
α-synuclein[n]	4.6(0.9)	1.5(0.6)	0.3	0.4	0.55	0.33-0.77	0.57(0.18-0.90)	0.37(0.20-0.56)	0.37
Apolipoprotein A1[n]	265.9(118.0)	166.9(144.1)	0.6	144.1	0.51	0.29-0.74	0.57(0.18-0.90)	0.6(0.41-0.77)	0.30
Apolipoprotein L1[n]	359.2(116.9)	1715.7(550.6)	4.8**	272.7	0.81***	0.64-0.98	0.86(0.42-1)	0.67(0.47-0.83)	0.67
C2[n]	166.8(26.1)	2050.0(1529.0)	12.3**	796.3	0.85***	0.66-1	0.71(0.29-0.96)	0.93(0.78-0.99)	0.50
Calgranulin B[n]	98065.4(38551.5)	164828.2(81775.8)	1.7	62392.2	0.67	0.45-0.88	0.86(0.42-1)	0.6(0.41-0.77)	0.60
CLU[p]	333363.0(173919.1)	338817.3(235125.1)	1	157994.9	0.61	0.40-0.82	0.86(0.42-1)	0.47(0.28-0.66)	0.47
D-dimer[p]	13414.3(3881.8)	14193.7(8111.01)	1.1	4301.2	0.59	0.34-0.84	0.71(0.29-0.96)	0.53(0.34-0.72)	0.27
Elastase[n]	36.2(5.3)	51.6(28.6)	1.4	28.6	0.66	0.42-0.90	0.57(0.18-0.90)	0.8(0.61-0.92)	0.47
Endocan[p]	470.6(11.0)	3980.9(990.5)	8.5*	51.9	0.81***	0.64-0.99	0.71(0.29-0.96)	0.8(0.61-0.92)	0.60
Ficolin-3[n]	349.5(169.2)	493.0(127.8)	1.4	626.6	0.46	0.16-0.75	0.29(0.04-0.71)	0.9(0.74-0.98)	0.00
Fibronectin[n]	713.4(260.0)	2406.3(2226.4)	3.4**	417	0.87****	0.75-0.99	1(0.59-1)	0.67(0.47-0.83)	0.80
HCC-1[p]	1924.1(7.0)	9608.5(41.0)	5	12685.9	0.6	0.34-0.86	0.29(0.04-0.71)	0.93(0.78-0.99)	0.28
IgA[n]	2175.0(1182.7)	5096.1(4671.6)	2.3***	3546.7	0.89****	0.78-1	0.86(0.42-0.1)	0.87(0.69-0.96)	0.80
IL-8[p]	1789.2(824.5)	11863.1(9786.4)	6.6****	6824.8	0.99****	0.96-1	1(0.59-1)	0.93(0.78(0.99)	0.28
Importin[n]	10.7(3.0)	15.9(7.3)	1.5	6.8	0.72	0.50-0.94	0.71(0.29-0.96)	0.8(0.61-0.92)	0.63
ITGB1[n]	16.5(9.8)	10.0(8.4)	0.6	4.8	0.47	0.27-0.66	1(0.59-1)	0.23(0.10(0.42)	0.27
Lysozyme[n]	311.9(77.6)	268.6(230.9)	0.9	207.4	0.73*	0.51-0.95	0.86(0.42-1)	0.73(0.54-0.88)	0.73
M2-PK[n]	13.2(9.0)	15.3(6.8)	1.2	15.2	0.54	0.27-0.80	0.43(0.10-0.82)	0.8(0.61-0.92)	0.20
MMP-1[p]	11582.3(1945.2)	39489.2(1754.1)	3.4	1239.4	0.56	033-0.79	0.86(0.42-1)	(0.47(0.28-0.66)	0.47
MMP-12[n]	2.4(0.6)	27.3(1.8)	11.5*	1.1	0.75**	0.56-0.93	0.86(0.42-1)	0.6(0.41-0.77)	0.60
MMP-9[p]	571808.7(58493.6)	617815.5(233576.7)	1.1	160857.6	0.67	0.43-0.90	0.86(0.42-1)	0.67(0.47-0.83)	0.67
Properdin[p]	31064.9(7425.5)	39669.7(6359.0)	1.3	38589.3	0.55	0.30-0.81	0.43(0.10-0.82)	0.8(0.61-0.92)	0.13
Proteinase 3[p]	552.2(232.6)	1854.2(1402.6)	3.4**	773.2	0.84****	0.68-1	0.86(0.42-1)	0.77(0.58-0.90)	0.77
S100A12[n]	314.2(1.0)	179.4(1.0)	0.6	1250	0.46	0.29-0.63	0.14(0-0.58)	0.93(0.78-0.99)	0.18
S100A4[n]	14.1(1.0)	1.6(1.3)	0.1	0.7	0.62	0.39-0.85	0.86(0.42-1)	0.43(0.26-0.63)	0.43
SOD1[p]	143150.4(43048.9)	121211.7(39984.9)	0.8	8245.7	0.49	0.24-0.73	1(0.59-1)	0.1(0.02-0.27)	0.20
SPARC[p]	12.5(4.0)	34.4(11.3)	2.8	8.7	0.69	0.43-0.94	0.71(0.29-0.96)	0.7(0.51-0.85)	0.30
TXD12[n]	576.5(163.4)	956.9(280.3)	1.7	200	0.62	0.36-0.88	0.71(0.29-0.96)	0.6(0.41-0.77)	0.33
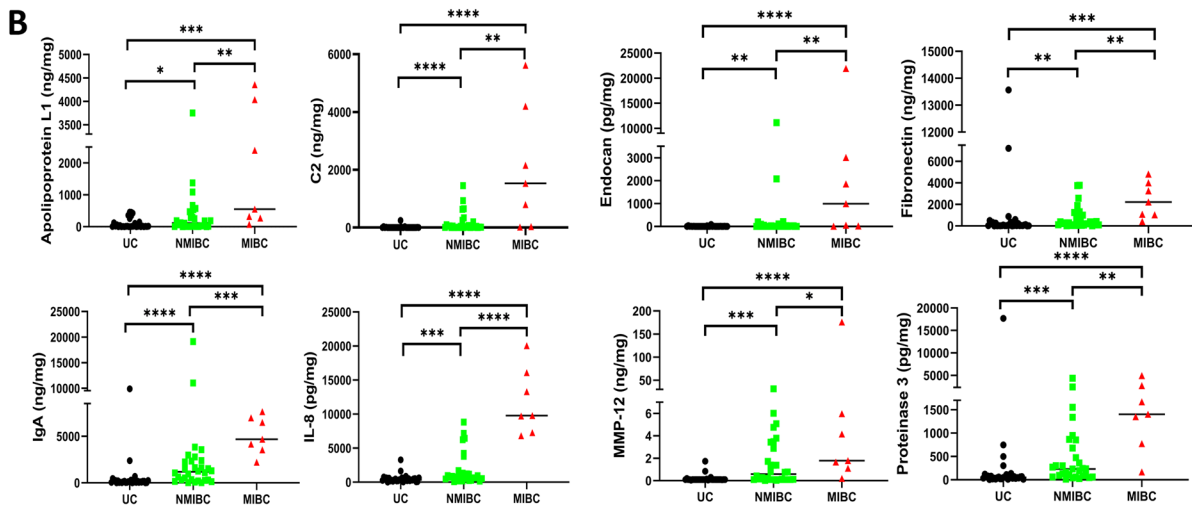
Fig. 5 The ability of 30 ELISA-validated urine proteins to discriminate bladder cancer patients by their disease stage. **A** Thirty urinary proteins were assayed by ELISA, using the UTSW cohort, comprised of 7 MIBC and 30 NMIBC subjects. The ELISA results for MIBC vs NMIBC are displayed. Biomarker protein units (normalized to creatinine) are as follows: n = ng/mg, p = pg/mg. Specificity.[0.8] depicts the specificity at a fixed sensitivity of 0.8. Indicated are the statistical significance *p*-values as determined by a Mann-Whitney *U*-test *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. **B** The dot plots depict the 8 proteins with the highest AUC value for MIBC vs NMIBC comparison. Tested samples include 30 NMIBC (10 Ta, 10 Tis, 10 T1) and 7 MIBC (T2–T4). Creatine-normalized urine protein levels are shown in different colors representing UC, NMIBC, and MIBC. The asterisks indicate the level of significance between the groups: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$

Besides comparing the different BC groups to UC, MIBC (T2–T4) was also compared to NMIBC (Ta, Tis, and T1). The results of the 30 ELISAs for the MIBC vs NMIBC stage comparisons are detailed in Fig. 5A and Additional file 1: Fig. S5. This data was creatinine-normalized. Figure 5B displays the dot plot view for the eight urine proteins with the highest AUC values for discriminating MIBC from NMIBC. These included urine Apolipoprotein L1, complement C2, Endocan, Fibronectin, IgA, IL-8, MMP-12, and Proteinase 3. Their AUC values ranged from 0.75 to 0.99. Urine IL-8 was best at discriminating these two groups (AUC=0.99, sensitivity=100%; specificity=93%). Urine IgA also outperformed other markers, with an AUC of 89%, a sensitivity
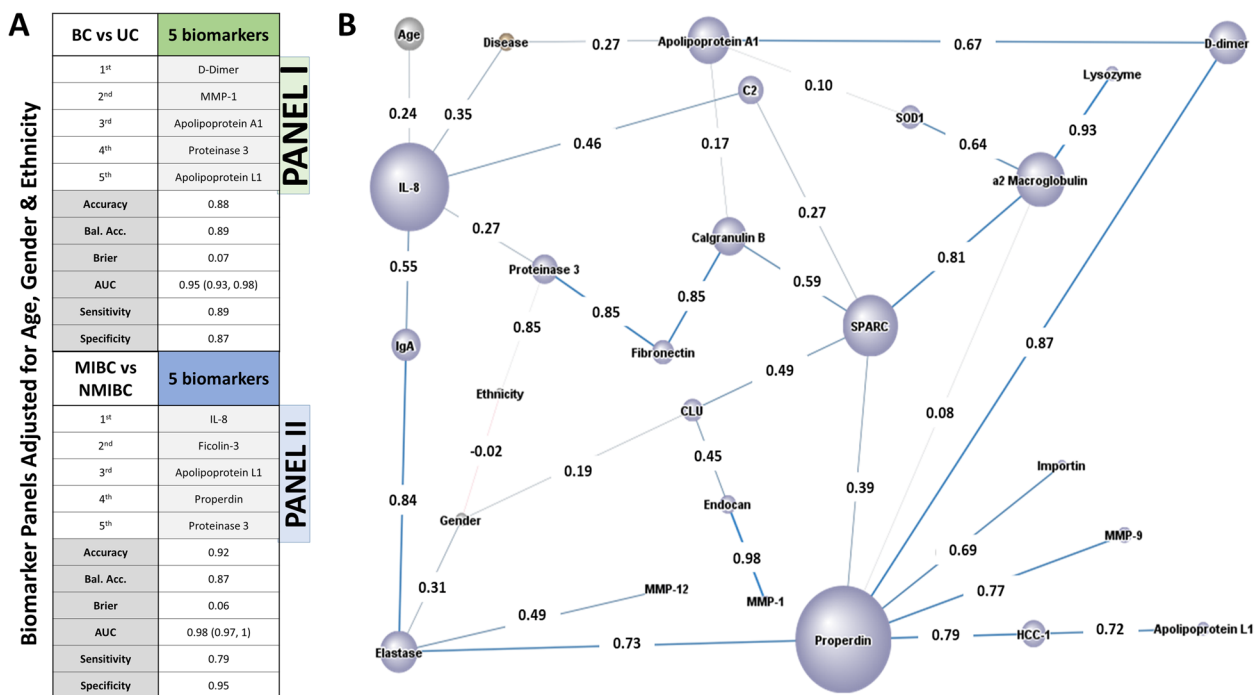
Fig. 6 Analysis of BC biomarkers using multi-marker panels and Bayesian network analysis. **A** Panel 1 displays the top 5-biomarker panel that distinguishes BC from UC after adjusting for age, gender, and ethnicity. A positive prediction approach was implemented as detailed in the "Methods" section for panels 1 and 2. The combined statistics for the panel of markers are displayed. Panel 2 displays the top 5-biomarker panel that distinguishes MIBC from NMIBC after adjusting for age, gender, and ethnicity. **B** The 21 proteins that showed significant AUC values for BC vs UC and fold change > 1 were subjected to Bayesian network analysis using BayesiaLab. The network was constructed as detailed in the "Methods" section. The circular nodes represent the urine biomarkers (colored purple), features (colored gray), and disease (BC vs UC; colored brown). The size of each node was determined using "node force" and is proportional to its impact on the other nodes in the network, based on conditional probabilities. The arcs that interconnect the nodes were determined with Pearson's correlation coefficient. The interconnections between nodes represent dependencies among the variables including the correlation coefficient between nodes. The thickness of the arcs is proportional to the correlation coefficient
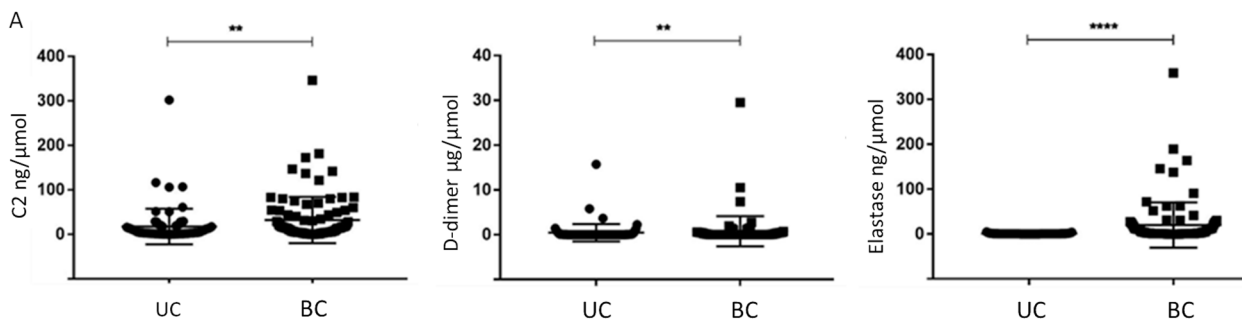
Fig. 7 Independent ELISA validation of elevated urinary proteins in a second validation cohort of Chinese BC patients. Dot plots depict the three urine proteins that were ELISA tested in a second validation cohort of Chinese ethnicity, comprised of BC patients ($N=91$) and UC ($N=77$), including 19 with kidney cancer, 4 with kidney cyst, 1 with kidney harmatoma, 50 with kidney stones, 1 with fibrous epithelial polyp, and 2 with kidney angiomyolipoma, as detailed in the supplementary figures/methods. Creatinine-normalized protein values are shown for each group (black dot = UC, black square = BC). The asterisks indicate the level of significance between the groups: ** $p < 0.01$, **** $p < 0.0001$

of 86%, and a specificity of 87%. Of particular note, urine IgA exhibited the highest specificity of 80% for MIBC, at 80% sensitivity, outperforming IL-8. As a sensitivity analysis, a bootstrap logistic regression model was used to

derive optimism-corrected performance metrics and to ascertain the robustness of each urine protein for distinguishing MIBC from NMIBC. This method yielded similar results, as listed in Additional file 1: Table S4.

Multi-marker panels and Bayesian network analysis of urine protein biomarkers for BC

Two multi-marker biomarker panels were constructed, after adjusting for age, gender, and ethnicity, using the positive prediction approach (Fig. 6A). Panel 1 was constructed for the BC vs UC comparison. The most discriminatory 5-marker panel consisted of urine D-dimer, MMP-1, Apolipoprotein A1, Proteinase 3, and Apolipoprotein L1, with an AUC of 0.95, a sensitivity value of 0.89, and a specificity value of 0.87. Not surprisingly, the top 3 proteins in panel 1 (D-dimer, MMP-1, Apolipoprotein A1) also ranked as the best single-marker performers, based on their individual AUC values (Fig. 4). Panel 2 was constructed for the MIBC vs NMIBC comparison. The most discriminatory 5-marker panel consisted of urine IL-8, Ficolin-3, Apolipoprotein L1, Properdin, and Proteinase 3. The panel produced an AUC of 0.98, with a sensitivity of 0.79 and a specificity of 0.95. Not surprisingly, 3 proteins in this panel (IL-8, Proteinase 3, Apolipoprotein L1) also ranked among the best single-marker performers, based on their individual AUC values (Fig. 5).

We next subjected all 21 proteins that showed significant ROC AUC values in discriminating BC from UC and clinical diagnosis to an unsupervised Bayesian network analysis. This analysis uses probability distributions to represent the inter-dependencies between all variables in a model and how they relate to one another. Figure 6B displays the derived Bayesian network, where the size of each node is representative of its impact on the other nodes in the network. Pearson's correlation coefficient is displayed between the nodes and the interconnects represent dependencies among the variables. This network illustrates significant interactions between biomarkers, demographics, and disease status. As indicated in the Bayesian network, urine Apolipoprotein A1, D-dimer, and IL-8 had the largest direct impact on BC versus UC discrimination, consistent with the findings from the additional analytic approaches described above. In addition, SPARC (ON) and α2-macroglobulin also exercised a large impact on other nodes in this network. Thus, this independent machine learning algorithm re-affirms the diagnostic potential of Apolipoprotein A1, D-dimer, and IL-8 in BC.

Secondary validation of urine protein biomarkers in BC urine by ELISA using a Chinese cohort

A secondary validation of upregulated urine proteins in BC vs UC was conducted in subjects of Chinese ethnicity. Three proteins were further assayed using ELISA in a cohort consisting of 77 UC and 91 BC patients (Additional file 1: Table S1). These urine proteins included complement C2, D-dimer, and Elastase which were among the best-performing markers in the first independent validation. Their associated creatinine normalized protein values are displayed in Fig. 7. The selected urine proteins were once again able to distinguish between BC and UC subjects.

Discussion

Research over the past several years has uncovered potentially important urine biomarkers and tests for BC, including BTA and NMP22. BTA-stat, an FDA-approved urine biomarker, is used clinically to detect bladder tumor-associated antigen (human complement factor H-related protein) in the urine. A meta-analysis of BTA stat reported a specificity of 67% and a sensitivity of 75% after reviewing 13 studies [8]. The sensitivity levels of BTA-stat have been shown to positively correlate with increasing grade of BC [8]. However, BTA-stat has several limitations. These include lower specificity values and issues relating to false-positive results in benign conditions [6]. Hence, urine BTA-stat may have limitations in the diagnosis and monitoring of disease progression. Similarly, NNP22 is an FDA-approved urine biomarker designed to detect the NMP22 protein levels which are high due to cell turnover from tumor apoptosis. A meta-analysis of 19 studies has identified this marker to have a pooled specificity of 88% and a sensitivity of 56% [7].

As opposed to studies looking at an isolated protein in the urine, a few screens have been reported where multiple proteins were examined simultaneously. Summarized below are a couple of studies documenting biomarkers with both sensitivity and specificity values greater than or equal to 85%. Goodison et al. performed a validation study for the urinary concentrations of 14 proteins (A1AT, APOE, ANG, CA9, CCL18, CD44, IL-8, MMP-9, MMP-10, OPN, PAI-1, PTX3, SDC1, and VEGF) using an ELISA [30]. An 8-biomarker panel (ANG, APOE, CA-9, IL-8, MMP-9, MMP-10, PAI-1, and VEGF) achieved the most accurate BC diagnosis with a sensitivity of 92% and a specificity of 97%. However, a panel of 3 biomarkers (APOE, IL-8, and VEGF) also performed well with a sensitivity of 90% and a specificity of 97% for the detection of BC [30]. Kumar et al. identified a 5-biomarker panel consisting of Apolipoprotein A4, Coronin-1A, DJ-1/PARK7, Gamma synuclein, and Semenogelin-2. ELISA and western blot data obtained an AUC of 0.92 and 0.98, respectively, in diagnosing Ta/T1 BC (sensitivity 79.2% and 93.9% for ELISA; specificity 100% and 96.7% for western blot) [31]. For the diagnosis of T2/T3 BC, the panel of markers achieved an AUC of 0.94 and 1, respectively, using the same methods (sensitivity 86.4% and 100%; specificity 100%) [31].

Low-grade BC has a high recurrence rate; therefore, identifying biomarkers for the surveillance of BC is

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 14 of 17

essential for the potential clinical management of the disease. Rosser et al. identified 10 biomarkers (ANG, APOE, CA9, IL-8, MMP-9, MMP-10, SDC1, SERPINA1, SERPINE1, and VEFGA) using ELISA for monitoring urine for recurrent BC. The complete panel achieved an AUC of 0.90, a sensitivity of 79%, and a specificity of 88% [32]. De Paoli et al. identified a panel of 6 biomarkers (cadherin-1, EN2, ErbB2, IL-6, IL-8, and VEGF-A) and three clinical parameters including BCG therapies, stage at the time of diagnosis, and past recurrences. The panel achieved an AUC of 0.91 and was identified through microarray and ELISA analysis [33].

There are several reasons to discriminate patients with bladder cancer from benign conditions. In patients with hematuria, it would be helpful to identify who needs cystoscopic evaluation which is invasive. Given that urine proteins are easily measurable and are compatible with point-of-care monitoring, a quick urine test could dramatically impact triage and workflow in urology outpatient clinics. Likewise, in bladder cancer surveillance, a reliable urine biomarker can help determine if cancer (like CIS) was missed or to avoid cystoscopy in marker-negative patients. Similarly, urine biomarkers that can reliably distinguish MIBC from NMIBC can inform us as to who has the more aggressive disease. When used as a routine point-of-care test (either at home or at outpatient visits), these urinary biomarkers may facilitate earlier identification of aggressive disease and design of tailored therapy.

The present work represents the first attempt to screen > 1000 urine proteins for urine biomarker candidates in BC, using a relatively novel aptamer-based screen. Systems biology analysis implicated molecular functions related to the extracellular matrix, collagen, integrin, heparin, and transmembrane tyrosine kinase signaling in BC susceptibility, with HNF4A and NFKB1 being key regulators. STEM analysis of the dysregulated pathways implicated a functional role for the immune system, complement, and interleukins in BC disease progression (Fig. 3D). This study has also uncovered urine proteins that outperform current FDA-approved markers in many respects. Several urine proteins (D-dimer, Apolipoprotein A1, MMP-1, Properdin, Calgranulin B) significantly discriminate BC from UC with AUC values from 0.85 to 0.96 (p-value < 0.0001). As a single biomarker, urine D-dimer was able to discriminate BC from UC with 96% accuracy (sensitivity = 95%; specificity = 90%). Likewise, several urine proteins (IL-8, IgA, Fibronectin, C2, Proteinase 3) significantly discriminate MIBC from NMIBC with AUC 0.84–0.99 (p-value < 0.001). Interestingly, several of the proteins described above have been documented to be elevated in bladder cancer tissue (at the RNA or protein level) and/or implicated in tumor biology at some level, as summarized in Additional file 1:

Table S5. Considering their biomarker potential and functional properties based on the literature (Additional file 1: Table S5), these urine proteins warrant further investigation, including, D-dimer [34], Apolipoprotein A1 [35, 36], Apolipoprotein L1, Calgranulin B [37, 38], complement C2 [39], Fibronectin [40–43], Ficolin-3, IL-8 [44–49], IgA [50], MMP-1 [51, 52], Properdin, and Proteinase 3 [53]. A summary of previous research on these proteins can be found in Additional file 1: Table S6. Additional markers increased in tissues are described in Additional file 1: Table S7. D-dimer is a specific cleavage product of fibrin and a symbol of hyperfibrinolysis [34]. It is the primary diagnostic tool in various diseases, such as deep venous thrombosis, systemic illness, and cancers [54]. Previous studies have reported that molecules in the coagulation/fibrinolysis system, especially plasma fibrinogen and D-dimer, are abnormal in cancer patients [34]. In the present study, urine D-dimer levels show a significant ability to differentiate BC from UC (AUC = 0.96) ($p < 0.0001$). After correcting for patient demographics, urine D-dimer is still eligible for inclusion within the 5-biomarker panel for best distinguishing BC from UC. Perhaps most impressive is the observation that urine D-dimer demonstrates a high sensitivity for the detection of BC (95%), and at a fixed specificity of 0.8, it can achieve a sensitivity of 0.97. Hence, as a single biomarker, urine D-dimer outperforms current FDA-approved biomarkers and competing biomarkers in the research literature as a sensitive biomarker for BC detection.

Apolipoproteins (Apolipoprotein A1 and Apolipoprotein L1) are proteins known to interact with the lipids of the lipoprotein core and also the aqueous environment of the plasma. Apolipoprotein A1 is the primary protein component of high-density lipoprotein while Apolipoprotein L1 is a minor component. Previous studies have validated Apolipoprotein A1 as a novel urinary biomarker for BC [35, 36]. In the current research, Apolipoprotein A1 was the second-best performing protein in terms of the AUC value (0.91) in distinguishing BC from UC.

After adjusting for demographics, this protein ranked within a 5-marker panel for distinguishing BC from UC. Similarly, Apolipoprotein L1 also ranked within the 5-marker panel for distinguishing BC from UC and MIBC from NMIBC.

Calgranulin B (S100A9) is a zinc- and calcium-binding protein that plays a prominent role in regulating inflammatory and immune responses. Several S100 proteins, including S100A9, have received attention regarding their possible role in tumor development and progression and studies report an increased expression in a variety of tumors, including ovarian, colon, gastric, and prostate cancer [37]. Increased expression of S100A9 protein in the serum has been previously associated with tumor

grade [37]. Current validation results of Calgranulin B are promising as it was among the top markers in discriminating BC from UC with an AUC of 0.85.

Complement proteins may promote tumor growth in the context of chronic inflammation [39]. Complement C2's relation to BC at this time is unknown. However, the present study identified this protein as the fourth best single protein for differentiating MIBC from the NMIBC stage. Properdin is also a member of the complement system, controlling the alternate pathway of complement activation. Research on properdin in BC is limited. However, in the current study, properdin demonstrated the third highest AUC value (AUC = 0.89, $p < 0.0001$) in discriminating BC from UC. These biomarker findings are consistent with the observation that changes in complement activation constitute one of the major pathways that predict BC disease progression, based on STEM analysis (Fig. 3D).

Fibronectin is a glycoprotein component of the extracellular matrix. Tumor cells can attach to fibronectin via integrins or other cell surface receptors [55]. Its effectiveness as a urine biomarker for BC has been explored in a variety of studies [40–43]. Here, fibronectin showed the third best discriminatory ability in identifying MIBC compared to NMIBC. The marker exhibited an AUC value of 0.87 ($p < 0.0001$).

IL-8 is a proinflammatory CXC chemokine. It has previously been associated with the promotion of neutrophil chemotaxis and degranulation [56]. Increased expression of IL-8 has been associated with endothelial cells, infiltrating neutrophils, tumor-associated macrophages, and cancer cells [56]. Therefore, IL-8 may be a significant regulatory factor within the tumor microenvironment. Previous studies have identified urinary IL-8 as a potential marker for BC [44–49]. In the present study, urine IL-8 was the best-performing protein in the MIBC vs NMIBC comparison, with an AUC of 0.99 ($p < 0.0001$), although its specificity was modest at a fixed sensitivity of 80%. Taken together with a wealth of supporting literature, this marker has the potential to be a monitoring tool for BC disease progression and warrants further analysis in this context.

IgA is an immunoglobulin and is often the first line of defense in the resistance against infections, particularly in mucosal tissues. A correlation of intra-tumor IgA1 and poor overall survival in BC patients has been identified in a previous study [50]. However, research regarding IgA in BC urine is limited. The data presented in this study indicated that urinary IgA may differentiate MIBC from NMIBC (AUC = 0.89, $p < 0.0001$). Overall, IgA performed 2nd best out of a total of 30 urinary markers validated for this comparison. Of particular note, urine IgA exhibited the highest specificity of 80% for MIBC, at 80% sensitivity, out-performing IL-8.

Matrix metalloproteinases (MMP) are a group of zinc-dependent proteolytic enzymes. Their role involves remodeling of the extracellular matrix. Many studies have evaluated the levels of MMPs in cancer patients and have reported the vital roles of some MMPs as potential diagnostic and prognostic biomarkers in tumorigenesis [57]. The current study has uncovered MMP-1 as the fourth best-performing molecule for distinguishing BC from UC (AUC = 0.89, $p < 0.0001$). This protein was also included in a 5-marker panel for distinguishing BC from UC. At the mechanistic level, one can envision tissue matrix remodeling as an important pre-requisite for cancer progression.

Conclusions

There is a need for these findings to be validated in additional cohorts. However, the urine proteins reported in this research exhibit great potential for use in a clinical setting. D-dimer, Apolipoprotein A1, MMP-1, Properdin, and Calgranulin B were identified as the most discriminatory urine markers in distinguishing BC from UC. Given that urine D-dimer has 97% sensitivity (at 80% specificity) for detecting BC, it may have a role in the initial diagnosis of BC, or for the detection of BC recurrence during surveillance follow-up. Urine Apolipoprotein A1, Properdin, and MMP-1 are the next best-performing biomarkers in this respect. On the other hand, urine IL-8 and IgA may have the potential in identifying disease progression during BC patient follow-up.

Several aspects of this study could be improved. Our study is limited by the relatively small sample size and the imbalance between the number of patients with NMIBC and MIBC. Because of these limitations, the generalizability of the study's findings warrants caution. This calls for future investigation in larger patient populations. Additionally, a larger sample size could uncover markers that are less discriminatory. Given that the current study only pursued the validation of 34 urine protein biomarkers, a larger number of additional proteins found to be significant could be assessed for validation in future studies. Urine D-dimer, MMP-1, Apolipoprotein A1, Proteinase 3, and Apolipoprotein L1 need to be validated independently and in multimarker panels in additional cross-sectional and longitudinal cohorts to confirm if they are superior to current FDA-approved markers for BC detection. Urine IL-8, Ficolin-3, Apolipoprotein L1, Properdin, and Proteinase 3 also need to be validated both independently and as a multi-marker panel in future cross-sectional and longitudinal cohorts to confirm if they are good indicators for bladder cancer disease progression. These novel urine biomarkers for BC also warrant systemic testing to assess their utility in BC surveillance and in predicting or monitoring response to treatment.

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 16 of 17

Abbreviations

AUC	Area under the curve
BC	Bladder cancer
DAVID	Database for Annotation, Visualization, and Integrated Discovery
ELISA	Enzyme-linked immunosorbent assay
FDA	United States Food and Drug Administration
GO	Gene Ontology
MIBC	Muscle invasive bladder cancer
NMIBC	Non-muscle invasive bladder cancer
PCA	Principal component analysis
RFA	Random forest analysis
STEM	Short Time-series Expression Miner

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-023-02813-x.

Additional file 1: Fig. S1. Consort diagram outlining the flow of the study. **Fig. S2.** Principal component analyses of the SOMAscan results. **Fig. S3.** Correlation analysis of the top 93 proteins elevated in BC urine. Fig. S4. Box plot expression profiles of the 30 ELISA validated proteins (BC vs UC). **Fig. S5.** Box plot expression profiles of the 30 ELISA validated proteins (MIBC vs NMIBC). **Table S1.** Demographic and clinical information of the secondary validation cohort of Chinese ethnicity. **Table S2.** ELISA validation kits selected for 34 proteins. **Table S3.** Single marker AUC analysis for the comparison of BC vs UC. **Table S4.** Single marker AUC analysis for the comparison of MIBC vs NMIBC. **Table S5.** Literature and public database profiles of shortlisted urine proteins. **Table S6.** Literature profiles of the outstanding proteins in urine and serum. **Table S7.** Literature profiles of the outstanding proteins in tissues.

Availability of data and materials

Supplementary information is available for this paper. All primary data will be made available by the corresponding author, if not in the supplementary data. The proteomic data will be made available upon request from the corresponding author.

Declarations

Ethics approval and consent to participate

The study was approved by the institutional review boards at the University of Houston, Houston, TX, and UTSW, Dallas, TX. The protocol number is 16441-EX.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department Biomedical Engineering, University of Houston, 3517 Cullen Blvd., Room 2027, Houston, TX 77204-5060, USA. [2]Department of Urology, Third Xiangya Hospital of Central South University, Changsha, China. [3]Center for Clinical Research and Evidence-Based Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. [4]Department of Urology, UT Southwestern Medical Center, Dallas, TX, USA.

References

1. Cancer facts and statistics: American Cancer Society; [Available from: https://www.cancer.org/research/cancer-facts-statistics/.
2. Richters A, Aben KKH, Kiemeney LALM. The global burden of urinary bladder cancer: an update. World J Urol. 2020;38(8):1895–904.
3. Reid MD, Osunkoya AO, Siddiqui MT, Looney SW. Accuracy of grading of urothelial carcinoma on urine cytology: an analysis of interobserver and intraobserver agreement. Int J Clin Exp Pathol. 2012;5(9):882–91.
4. Sugeeta SS, Sharma A, Ng K, Nayak A, Vasdev N. Biomarkers in bladder cancer surveillance. Front Surg. 2021;8:735868.
5. Wang Z, Que H, Suo C, Han Z, Tao J, Huang Z, et al. Evaluation of the NMP22 BladderChek test for detecting bladder cancer: a systematic review and meta-analysis. Oncotarget. 2017;8(59):100648–56.
6. Guo A, Wang X, Gao L, Shi J, Sun C, Wan Z. Bladder tumour antigen (BTA stat) test compared to the urine cytology in the diagnosis of bladder cancer: a meta-analysis. Can Urol Assoc J. 2014;8(5–6):E347–52.
7. Barocas DA, Boorjian SA, Alvarez RD, Downs TM, Gross CP, Hamilton BD, et al. Microhematuria: AUA/SUFU Guideline. J Urol. 2020;204(4):778–86.
8. Chang SS, Boorjian SA, Chou R, Clark PE, Daneshmand S, Konety BR, et al. Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO Guideline. J Urol. 2016;196(4):1021–9.
9. Lei R, Huo R, Mohan C. Current and emerging trends in point-of-care urinalysis tests. Expert Rev Mol Diagn. 2020;20(1):69–84.
10. Albaba D, Soomro S, Mohan C. Aptamer-based screens of human body fluids for biomarkers. Microarrays (Basel). 2015;4(3):424–31.
11. Sattlecker M, Kiddle SJ, Newhouse S, Proitsi P, Nelson S, Williams S, et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. Alzheimers Dement. 2014;10(6):724–34.
12. Kiddle SJ, Sattlecker M, Proitsi P, Simmons A, Westman E, Bazenet C, et al. Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study. J Alzheimers Dis. 2014;38(3):515–31.
13. De Groote MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, et al. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. PLoS ONE. 2013;8(4):e61002.
14. Nahid P, Bliven-Sizemore E, Jarlsberg LG, De Groote MA, Johnson JL, Muzanyi G, et al. Aptamer-based proteomic signature of intensive phase treatment response in pulmonary tuberculosis. Tuberculosis (Edinb). 2014;94(3):187–96.
15. Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P, et al. Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. Proc Natl Acad Sci U S A. 2015;112(23):7153–8.
16. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, et al. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PLoS ONE. 2010;5(12):e15003.
17. Ostroff RM, Mehan MR, Stewart A, Ayers D, Brody EN, Williams SA, et al. Early detection of malignant pleural mesothelioma in asbestos-exposed individuals with a noninvasive proteomics-based surveillance tool. PLoS ONE. 2012;7(10):e46091.
18. Mehan MR, Williams SA, Siegfried JM, Bigbee WL, Weissfeld JL, Wilson DO, et al. Validation of a blood protein signature for non-small cell lung cancer. Clin Proteomics. 2014;11(1):32.
19. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. JAMA. 2016;315(23):2532–41.
20. Stanley S, Vanarsa K, Soliman S, Habazi D, Pedroza C, Gidley G, et al. Comprehensive aptamer-based screening identifies a spectrum of urinary biomarkers of lupus nephritis across ethnicities. Nat Commun. 2020;11(1):2197.

Vanarsa *et al. BMC Medicine*   (2023) 21:133

Page 17 of 17

21. Soomro S, Venkateswaran S, Vanarsa K, Kharboutli M, Nidhi M, Susarla R, et al. Predicting disease course in ulcerative colitis using stool proteins identified through an aptamer-based screen. Nat Commun. 2021;12(1):3989.

22. SomaLogic. SOMAscan proteomic assay technical white paper. 2015. p. 1–14.

23. Goksuluk D, Korkmaz S, Zararsiz G, Karaagaoglu AE. easyROC: an interactive web-tool for ROC curve analysis using R language environment. R J. 2016;8(2):213.

24. Catalina A, Bürkner P-C, Vehtari A. Projection predictive inference for generalized linear and additive multilevel models. 2020. arXiv preprint arXiv:201006994.

25. Piironen J, Paasiniemi M, Vehtari A. Projective inference in high-dimensional problems: prediction and feature selection. Electron J Statist. 2020;14(1):2155–97. https://doi.org/10.1214/20-EJS1711.

26. Piironen J, Paasiniemi M, Catalina A, Vehtari A. Projpred: projection predictive feature selection. R package version 2.0.22020.

27. Carvalho CM, Polson NG, Scott JG. Handling sparsity via the horseshoe. PMLR. 2009;5:73–80.

28. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Statist. 2017;11(2):5018–51. https://doi.org/10.1214/17-EJS1337SI.

29. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2015;27:1413–32.

30. Goodison S, Chang M, Dai Y, Urquidi V, Rosser CJ. A multi-analyte assay for the non-invasive detection of bladder cancer. PLoS ONE. 2012;7(10):e47469.

31. Kumar P, Nandi S, Tan TZ, Ler SG, Chia KS, Lim WY, et al. Highly sensitive and specific novel biomarkers for the diagnosis of transitional bladder carcinoma. Oncotarget. 2015;6(15):13539–49.

32. Rosser CJ, Chang M, Dai Y, Ross S, Mengual L, Alcaraz A, et al. Urinary protein biomarker panel for the detection of recurrent bladder cancer. Cancer Epidemiol Biomarkers Prev. 2014;23(7):1340–5.

33. De Paoli M, Gogalic S, Sauer U, Preininger C, Pandha H, Simpson G, et al. Multiplatform biomarker discovery for bladder cancer recurrence diagnosis. Dis Markers. 2016;2016:4591910.

34. Li X, Shu K, Zhou J, Yu Q, Cui S, Liu J, et al. Preoperative plasma fibrinogen and D-dimer as prognostic biomarkers for non-muscle-invasive bladder cancer. Clin Genitourin Cancer. 2020;18(1):11-9.e1.

35. Li H, Li C, Wu H, Zhang T, Wang J, Wang S, et al. Identification of Apo-A1 as a biomarker for early diagnosis of bladder transitional cell carcinoma. Proteome Sci. 2011;9(1):21.

36. Li C, Li H, Zhang T, Li J, Liu L, Chang J. Discovery of Apo-A1 as a potential bladder cancer biomarker by urine proteomics and analysis. Biochem Biophys Res Commun. 2014;446(4):1047–52.

37. Minami S, Sato Y, Matsumoto T, Kageyama T, Kawashima Y, Yoshio K, et al. Proteomic study of sera from patients with bladder cancer: usefulness of S100A8 and S100A9 proteins. Cancer Genomics Proteomics. 2010;7(4):181–9.

38. Bansal N, Gupta A, Sankhwar SN, Mahdi AA. Low- and high-grade bladder cancer appraisal via serum-based proteomics approach. Clin Chim Acta. 2014;436:97–103.

39. Pio R, Corrales L, Lambris JD. The role of complement in tumor growth. Adv Exp Med Biol. 2014;772:229–62.

40. Shen Z, Wei K, Yang S, Shi S, Chen Z, Li Y, et al. Measurement of urine fibronectin in the diagnosis of invasive bladder transitional carcinoma. Chin J Urol. 1993;14:27–9.

41. Menéndez V, Fernández-Suárez A, Galán JA, Pérez M, García-López F. Diagnosis of bladder cancer by analysis of urinary fibronectin. Urology. 2005;65(2):284–9.

42. Sánchez-Carbayo M, Urrutia M, GonzálezdeBuitrago JM, Navajo JA. Evaluation of two new urinary tumor markers: bladder tumor fibronectin and cytokeratin 18 for the diagnosis of bladder cancer. Clin Cancer Res. 2000;6(9):3585–94.

43. Eissa S, Zohny SF, Zekri AR, El-Zayat TM, Maher AM. Diagnostic value of fibronectin and mutant p53 in the urine of patients with bladder cancer: impact on clinicopathological features and disease recurrence. Med Oncol. 2010;27(4):1286–94.

44. Urquidi V, Chang M, Dai Y, Kim J, Wolfson ED, Goodison S, et al. IL-8 as a urinary biomarker for the detection of bladder cancer. BMC Urol. 2012;12:12.

45. Kamat AM, Briggman J, Urbauer DL, Svatek R, Nogueras González GM, Anderson R, et al. Cytokine Panel for Response to Intravesical Therapy (CyPRIT): nomogram of changes in urinary cytokine levels predicts patient response to Bacillus Calmette-Guérin. Eur Urol. 2016;69(2):197–200.

46. Kumari N, Agrawal U, Mishra AK, Kumar A, Vasudeva P, Mohanty NK, et al. Predictive role of serum and urinary cytokines in invasion and recurrence of bladder cancer. Tumour Biol. 2017;39(4):1010428317697552.

47. Reis ST, Leite KR, Piovesan LF, Pontes-Junior J, Viana NI, Abe DK, et al. Increased expression of MMP-9 and IL-8 are correlated with poor prognosis of bladder cancer. BMC Urol. 2012;12:18.

48. Koçak H, Oner-Iyidoğan Y, Koçak T, Oner P. Determination of diagnostic and prognostic values of urinary interleukin-8, tumor necrosis factor-alpha, and leukocyte arylsulfatase-A activity in patients with bladder cancer. Clin Biochem. 2004;37(8):673–8.

49. Sheryka E, Wheeler MA, Hausladen DA, Weiss RM. Urinary interleukin-8 levels are elevated in subjects with transitional cell carcinoma. Urology. 2003;62(1):162–6.

50. Welinder C, Jirström K, Lehn S, Nodin B, Marko-Varga G, Blixt O, et al. Intra-tumour IgA1 is common in cancer and is correlated with poor prognosis in bladder cancer. Heliyon. 2016;2(8):e00143.

51. Nutt JE, Mellon JK, Qureshi K, Lunec J. Matrix metalloproteinase-1 is induced by epidermal growth factor in human bladder tumour cell lines and is detectable in urine of patients with bladder tumours. Br J Cancer. 1998;78(2):215–20.

52. Durkan GC, Nutt JE, Rajjayabun PH, Neal DE, Lunec J, Mellon JK. Prognostic significance of matrix metalloproteinase-1 and tissue inhibitor of metalloproteinase-1 in voided urine samples from patients with transitional cell carcinoma of the bladder. Clin Cancer Res. 2001;7(11):3450–6.

53. Zoidakis J, Makridakis M, Zerefos PG, Bitsika V, Esteban S, Frantzi M, et al. Profilin 1 is a potential biomarker for bladder cancer aggressiveness. Mol Cell Proteomics. 2012;11(4):M111.009449.

54. Vikey A. D-dimer as an alarming biomarker in various cancers: a review of literature. Glob Med Therap. 2018;1:3–4.

55. Kang HW, Kim W-J, Yun S-J. The role of the tumor microenvironment in bladder cancer development and progression – Kang. Transl Cancer Res. 2022;6(Supplement 4).

56. Waugh DJ, Wilson C. The interleukin-8 pathway in cancer. Clin Cancer Res. 2008;14(21):6735–41.

57. Miao C, Liang C, Zhu J, Xu A, Zhao K, Hua Y, et al. Prognostic role of matrix metalloproteinases in bladder carcinoma: a systematic review and meta-analysis. Oncotarget. 2017;8(19):32309–21.

Publisher's Note