

## RESEARCH ARTICLE

# Aptamer based proteomic pilot study reveals a urine signature indicative of pediatric urinary tract infections

Liang Dong<sup>1</sup>, Joshua Watson<sup>2</sup>, Sha Cao<sup>3</sup>, Samuel Arregui<sup>1</sup>, Vijay Saxena<sup>1</sup>, John Ketz<sup>4</sup>, Abduselam K. Awol<sup>5</sup>, Daniel M. Cohen<sup>6</sup>, Jeffrey M. Caterino<sup>7</sup>, David S. Hains<sup>1</sup>, Andrew L. Schwaderer<sup>1\*</sup>

**1** Department of Pediatrics, Indiana University, Indianapolis, Indiana, United States of America, **2** Division of Infectious Disease, Nationwide Children's Hospital, Columbus, Ohio, United States of America, **3** Department of Biostatistics, Indiana University, Indianapolis, Indiana, United States of America, **4** The Research Institute at Nationwide Children's Hospital, Columbus, Ohio, United States of America, **5** Earlham College, Richmond, Indiana, United States of America, **6** Division of Emergency Medicine, Nationwide Children's Hospital, Columbus, Ohio, United States of America, **7** Division of Emergency Medicine, The Ohio State University, Columbus, Ohio, United States of America

\* [schwadea@iu.edu](mailto:schwadea@iu.edu)



## OPEN ACCESS

**Citation:** Dong L, Watson J, Cao S, Arregui S, Saxena V, Ketz J, et al. (2020) Aptamer based proteomic pilot study reveals a urine signature indicative of pediatric urinary tract infections. *PLoS ONE* 15(7): e0235328. <https://doi.org/10.1371/journal.pone.0235328>

**Editor:** Mehreen Arshad, Northwestern University Feinberg School of Medicine, UNITED STATES

**Received:** October 23, 2019

**Accepted:** June 14, 2020

**Published:** July 6, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0235328>

**Copyright:** © 2020 Dong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and Supporting Information files.

**Funding:** Eli Lilly foundation award (<https://www.lilly.com/who-we-are/lilly-foundation>) ALS

## Abstract

### Objective

Current urinary tract infection (UTI) diagnostic strategies that rely on leukocyte esterase have limited accuracy. We performed an aptamer-based proteomics pilot study to identify urine protein levels that could differentiate a culture proven UTI from culture negative samples, regardless of pyuria status.

### Methods

We analyzed urine from 16 children with UTIs, 8 children with culture negative pyuria and 8 children with negative urine culture and no pyuria. The urine levels of 1,310 proteins were quantified using the Somascan™ platform and normalized to urine creatinine. Machine learning with support vector machine (SVM)-based feature selection was performed to determine the combination of urine biomarkers that optimized diagnostic accuracy.

### Results

Eight candidate urine protein biomarkers met filtering criteria. B-cell lymphoma protein, C-X-C motif chemokine 6, C-X-C motif chemokine 13, cathepsin S, heat shock 70kDA protein 1A, mitogen activated protein kinase, protein E7 HPV18 and transgelin. AUCs ranged from 0.91 to 0.95. The best prediction was achieved by the SVMs with radial basis function kernel.

### Conclusions

Biomarkers panel can be identified by the emerging technologies of aptamer-based proteomics and machine learning that offer the potential to increase UTI diagnostic accuracy, thereby limiting unneeded antibiotics.

and DSH The Research Institute at Nationwide Hospital intramural funds (<https://www.nationwidechildrens.org/research>). ALS National Institute on Aging (<https://www.nia.nih.gov>) National Institute of Diabetes and Digestive and Kidney Diseases (<https://www.niddk.nih.gov>) Grant numbers: R01AF050801 (salary support for JC, ALS and DSH), R01DK106286 (salary support for ALS and DSH) and R01DK117934 (salary support for ALS and DSH). The sponsors played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I (Andrew Schwaderer) have consulted for Allena Pharmaceuticals on a topic unrelated to this manuscript; This does not alter our adherence to PLOS ONE policies on sharing data and materials. Otherwise there are no conflicts to report. The work in this manuscript is original, not previously published, and not submitted for publication or consideration elsewhere.

## Background

Urinary tract infections (UTIs) are frequently encountered. UTIs account for 7 million office visits and 400,000 hospitalizations annually in the United States [1, 2]. The aforementioned hospitalizations for UTIs increased 52% between 1998 and 2011 and resulted in an estimated 2.8 billion dollars of cost. In children, UTIs account for 7% of emergency department (ED) antibiotic prescriptions [3]. Unfortunately, antibiotic resistance in uropathogens is increasing [4]. The diagnosis of UTIs is typically made at the point of care by symptoms and the identification of nitrites and /or leukocyte esterase (LE) on urinalysis (UA) and/or urine dipstick [5]. Ultimately urine culture results with 50,000 colony forming units (cfu)/ml of a uropathogen is used to confirm a clinical UTI [5]. However accurate urine culture results can be dependent on proper collection methodology and take 24–48 hours to complete [6]. Further, the use of LE on urine dipsticks has limitations including limited sensitivity and specificity [5]. Specifically non UTI, infectious and/or inflammatory conditions such chlamydia, appendicitis and interstitial nephritis can result in positive urine leukocyte esterase and negative urine cultures [7].

A timely and accurate UTI diagnosis is important in clinical care. Initiating antibiotics in a patient with suspected UTI that is actually culture negative pyuria exposes the patient to unneeded antibiotics and potentially increases the risk of antibiotic resistant bacteria [8]. Conversely, waiting for culture results before treating a patient who has a true UTI risks complication from disease progression from cystitis to pyelonephritis or even urosepsis [9]. Methodologies that increase accuracy of UTI diagnosis are needed.

The discoveries of new biomarkers are fundamental to improving clinical care. Several challenges have limited protein-based biomarker discovery with traditional antibody based ELISAs. Specifically, ELISAs are time consuming to perform and the required antibodies have inherent costs, instability, batch to batch variation, storage requirements limited dynamic ranges and are difficult to multiplex [10–12]. SOMAscan™ (Somologics, Boulder CO) uses SOMAmer (slow-off-rate modified aptamer) protein binding reagents. Aptamers are modified DNA with high affinity ( $10^9$ – $10^{12}$  M) and high specificity for their cognate analytes comparable to sandwich ELISA performance that have been used for biomarker discovery [13]. For example, in 2018 the Somascan aptamer-based platform was utilized to discover unique protein profiles in autoimmune cholangitis [14]. Aptamers are being explored as affordable, sensitive, specific, user-friendly point of care tests on a variety of platforms [15]. Aptamers have the ability to perform in formats where antibodies often perform poorly such as homogenous multiplex assays, do not degrade when stored at room temperature as a dry lyophilized reagent at room temperature and have minimal to no batch to batch variation [12]. Thus there is speculation that aptamers may replace antibodies in future diagnostics [16]. The objectives of this pilot study are to determine if SOMAscan<sup>M</sup> aptamer-based comparison of (a) children with neither pyuria nor growth on culture, (b) children with pyuria but no growth on urine culture and (c) children with pyuria and 50,000 cfu/ml of *E.coli* on urine culture will reveal a protein profile unique to children with UTI along with performing an initial experiment comparing aptamer based detection with antibody based detection using an adult emergency cohort for the latter.

## Methods

### Study approval and patients

The study was approved and granted a waiver of informed consent by the Nationwide Children's Hospital Institutional review board (IRB13-00090). Samples were prospectively obtained in the ED and main campus Urgent Care at Nationwide Children's Hospital, Columbus, Ohio as previously described [17]. Inclusion criteria consisted of dipstick urinalysis and urine culture

performed for any clinical indication and availability of excess urine sample. Children who had received antibiotic treatment within 7 days before ED presentation were excluded. Samples from a second cohort were collected from the Ohio State University Wexner Medical Center Emergency Department (ED) patients. Institutional review board approval was obtained and informed consent was obtained from each patient or appropriate proxy. Inclusion criteria consisted of ED  $\geq 65$  years of age that had a urinalysis for suspected UTI. Exclusion criteria consisted of chronic intermittent catheterization, UTI or positive urine culture or genitourinary procedure in the prior 30 days, antibiotic use in the past 14 days, immunosuppression, hemodialysis, homelessness, previous enrollment, incarceration, non-English speaking, trauma team activation, and lack of patient or proxy ability to give consent or respond to the survey.

### Sample collection and processing

We prospectively collected urine samples as previously described [17]. Samples were collected when a research assistant was available to collect samples resulting in a convenience collection. After ensuring sufficient urine volume was available for clinical diagnostic tests, excess urine was immediately collected in AssayAssure urine collection tubes (Thermo Scientific, Waltham, MA) containing a bacteriostatic preservative that suppresses nuclease and protease activity and preserves urine specimens at room temperature for up to 26 days per the manufacturer. We previously independently confirmed protein stability for 14 days [17]. Samples were processed within 7 days of collection and centrifuged at 3,000 rpm for 5 minutes with the supernatant saved in 300 to 500  $\mu$ l aliquots then stored at  $-80^{\circ}\text{C}$  until use.

### Sample selection and groups

Samples were selected from our previously reported cohort of pediatric ED patients that had sufficient urine volume for Somalogic analysis, were collected by clean catch and fit criteria for the following groups: (a) *UTI* defined by  $\geq 1+$  LE on urine dipstick and  $\geq 50,000$  cfu/ml of *E. coli* on urine culture; (b) *Culture negative (CN) pyuria* defined by  $\geq 1+$  LE on urine dipstick and no growth on urine culture and (d) *CN no pyuria* defined by negative LE on urine dipstick and no growth on urine culture [5, 17, 18]. The UTI group was further divided between those with and without fevers  $\geq 100.4^{\circ}$  Fahrenheit/ $38^{\circ}$  Celsius (either measured in the ED or by report at home). Urine culture functioned as the “reference standard”. The adult ED cohort was divided into a culture negative and culture positive group.

### Urine aptamer proteomic evaluation

One aliquot of the selected samples were sent on dry ice to Somalogic Inc. (Boulder, CO) to measure concentrations of 1,310 proteins via SOMAscan™ analysis. The SOMAscan™ results were presented as relative fluorescent units (RFU) per ml. Urine protein levels were normalized to urine creatinine which were measured using the Oxford colorimetric assay (Oxford Biomedical Research, Oxford, MI).

### Urine antibody based protein detection

A V-PLEX Human GM-CSF Kit (Mesoscale Discovery, Rockville, MD, catalog # K151RID-1) was used for validation, in an adult ED cohort, of urine granulocyte-macrophage colony stimulating factor (GM-CSF) levels according to the manufacturer’s directions. We chose the mesoscale array because it has a large dynamic range of detection (0.14–770 pg/ml), is labeled for use in urine and we have experience with the assay [19]. Results were normalized to urine creatinine as described previously for the urine aptamer levels.

## Statistical analysis

Demographics and presenting symptoms were compared with Graphpad Prism (La Jolla, CA) using the chi square test if percentages or proportions were evaluated and the t-test if continuous data was evaluated. Groups were compared by the Wilcoxon test with SPSS software (IBM corporation, Armonk NY). Proteins were filtered by meeting all of the following criteria: (a) significantly different levels between the UTI group (febrile and afebrile combined) versus the CN-pyuria group; (b) significantly different levels between the UTI group (febrile and afebrile combined) vs the CN-no pyuria group and (c) but not significantly different levels between the CN-pyuria group vs CN-no pyuria group. Significance was assigned for a p value of  $< 0.05$  and the results were further filtered for a p value  $< 0.01$ . Next, proteins with that had a p-value  $< 0.01$  and under curve (AUC) of  $> 0.9$  were selected as candidate biomarkers. A general guide for interpreting the utility of a biomarker based on AUC is: “fail” = 0.5–0.6, “poor” = 0.6–0.7, “fair” = 0.7–0.8, “good” = 0.8–0.9 and “excellent” = 0.9–1.0 [20]. AUCs and the concentration with likelihood ratios (LRs) were calculated using Graphpad Prism.

## Support vector machine (SVM) predictive model optimization

Feature selection plays a crucial role in biomedical data mining. [21] Three different feature selection approaches were considered to reduce the data dimensionality before the model was trained on training subset in each fold of inner leave-one-out cross-validation: (a) feature selection based on the Wilcoxon rank sum test to screen proteins with expression strongly associated with UTI. (b) Feature ranking on the basis of random forest feature importance scores computed from the Gini impurity reduction. (c) ReliefF feature selection techniques [22]. Given our sample size, we performed hyperparameter tuning and model optimization using leave-one-out cross-validation in an inner loop. We conducted a grid search to explore the optimal hyperparameter space including a range of values for gamma and/or C for support vector classifiers with either linear or RBF kernel. The accuracy was calculated at each cross-validation split on the validation set. The mean accuracy was used as a metric for model selection. To assess the predictive performance, we further computed the performance estimates of our models on unseen data (test set) using 5-fold cross-validation in the outer loop. The overall unbiased generalization performance of the optimal model was evaluated by the mean AUC values of the receiver operating characteristic (ROC) curve, obtained in each iteration of the cross-validation split. The class probability estimate of each sample was calculated based on decision values of SVM using the parameters learned in Platt scaling [23]. A number of Python libraries and R packages were used in data analysis and machine learning processes including Pandas, Scikit-Learn, skrebate, ggplot2, dplyr, ROCR, and pROC. A schematic of the methodology for feature selection and nested cross validation is presented as [S2 Material](#).

## Figure generation

Figures were generated using Graphpad Prism, Microsoft Powerpoint (Microsoft corporation, Redmond, WA) or by web-based Lucidchart tool (<https://www.lucidchart.com>).

## Results

### Patients

We included urine samples from 32 patients (4 males and 28 females) with a median age of 7.1 years (interquartile range, 4.7–14.0). Sixteen patients met criteria for UTI group, 8 patients had CN-pyuria and 8 patients had CN-no pyuria. The UTI group was evenly divided between those with and without fevers. No patients were immunosuppressed. Two patients in the UTI

**Table 1. Epidemiology and presenting symptoms<sup>^</sup> of groups.**

	UTI (n = 16)	CN pyuria (n = 8)	CN no pyuria (n = 8)	P value
Mean age (years) ± std dev	8.2 ± 4.7	11.4 ± 5.9	7.2 ± 4.2	0.217
Female:male (% female)	15:1 (94%)	7:1 (88%)	5:3 (63%)	0.133
Fever	8 (50%)	2 (25%)	0 (0%)	*0.041
Dysuria	5 (31%)	2 (25%)	3 (38%)	0.793
Frequency	4 (25%)	1 (13%)	0 (0%)	0.272
Urgency/enuresis	4 (25%)	2 (25)	1 (13%)	0.760
Suprapubic pain	4 (25%)	1 (13%)	0 (0%)	0.272
Abdominal pain	8 (50%)	4 (50%)	3 (38%)	0.828
Back/flank pain	2 (13%)	3 (38%)	0 (0%)	0.105

<sup>^</sup>other presenting symptoms included syncope (1 in UTI group and one in normal U urine group, “bump” on testicle (1 in CN no pyuria group), headache (one in normal urine group), foul smelling urine (1 in normal urine group) and memory loss (1 in CN pyuria group)

\*statistically significant,  $p < 0.05$  with the significant difference between UTI and CN no pyuria group.

<https://doi.org/10.1371/journal.pone.0235328.t001>

group had a history of kidney stones. One 5-year old patient in the UTI group had a history of congenital hydronephrosis. There were no statistically significant differences in age, sex or presenting symptoms among groups, with the exception of a higher percentage of patients with fever in the UTI compared to the CN no pyuria group (Table 1).

### Identification of proteins elevated during UTI

We identified 133 proteins that were significantly elevated ( $p$  value  $< 0.05$ ) in UTI vs. the culture negative pyuria comparison and) UTI vs. the CN no pyuria group but were not statistically different when the CN pyuria group was compared to the CN no pyuria urine group (S3 Material). To focus on the most differentiating proteins between groups, we filtered for a  $p$  value  $< 0.01$  and identified 32 proteins that were elevated in the UTI group, but not the CN-pyuria or CN no pyuria groups (Table 2). The candidates that meet the  $p$  value  $< 0.01$  criteria were filtered for AUC curves  $> 0.9$  to determine candidate proteins as “excellent” biomarkers to differentiate culture positive (febrile + afebrile UTI) samples from the combined culture negative samples (CN-pyuria and CN-no pyuria) with the results presented in Fig 1. Scatterplots of the urine biomarker to creatinine ratio in each group along with the urine biomarker to creatinine ratio threshold (“cut off” levels, sensitivity, specificity and likelihood ratios to differentiate UTI (febrile + afebrile) samples from the control samples (CN no pyuria + CN pyuria) are presented in Fig 2.

### Antibody based protein detection

Antibody based protein detection was performed on GM-CSF in the adult patients enrolled from the Ohio State University Wexner Medical Center ED, with the results divided into culture negative ( $n = 10$ ) and culture positive ( $n = 6$ ). GM-CSF was selected from Table 1 because it could be tested in a commercially available V-PLEX assay and has previously reported relevance in UTI pathophysiology. The Mesoscale V-PLEX antibody-based protein detection had an AUC of 0.9333, comparable to the AUC of 0.8906 with the Nationwide Children’s ED cohort and Somascan™ aptamer-based measurement (Fig 3).

### Support vector machine (SVM) predictive model optimization

The Random forest, ReliefF, and Wilcoxon rank-sum test were applied in feature selection to determine the best combination of urine protein biomarkers that achieved the best prediction

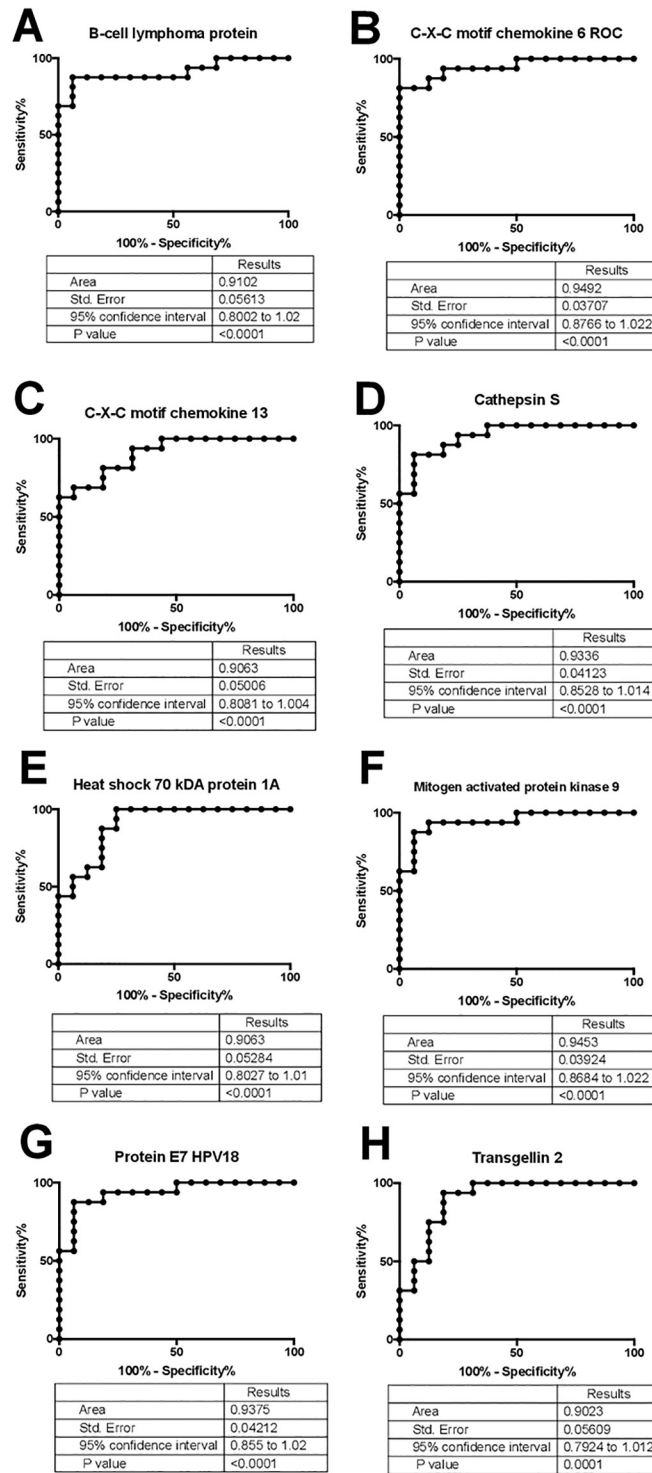
Table 2. Urine biomarker levels (urine biomarker (RFU)/ urine creatinine (mg)).

Protein	Median CN no pyuria	Median CN pyuria	Median UTI	p value		
				UTI vs CN-no pyuria	UTI vs CN-pyuria	CN-no pyuria vs CN-pyuria
Alpha-2-macroglobulin	812 ± 1055	1170 ± 2104	15462 ± 153990	0.001	0.006	0.442
B-cell lymphoma 6 protein	678 ± 1132	542 ± 27213	46537 ± 656650	<0.001	0.006	0.721
BH3-interacting domain death agonist	341 ± 1900	630 ± 263	33323 ± 6533	0.006	<0.001	0.234
C-X-C motif chemokine 11	55.12 ± 439	3604 ± 548	1331 ± 95541	0.005	0.001	0.328
C-X-C motif chemokine 13	44 ± 972	80 ± 163	474 ± 8735	<0.001	0.003	0.195
C-X-C motif chemokine 6	185 ± 942	145 ± 131	3342 ± 35842	0.001	<0.001	0.645
Calcium/calmodulin-dependent protein kinase type 1	1452 ± 3393	2126 ± 1951	8080 ± 18809	0.004	0.009	0.328
Cathepsin S	114 ± 645	314 ± 255	2513 ± 33882	<0.001	<0.001	0.195
Endothelial monocyte-activating polypeptide 2	277 ± 698	394 ± 295	1788 ± 5360	0.002	0.006	0.382
Granulocyte-macrophage colony-stimulating factor	44 ± 111	46 ± 40	296 ± 930	0.004	<0.001	0.959
Gro-beta/gamma	162 ± 843	377 ± 259	6564 ± 93973	0.002	0.001	0.161
Growth-regulated alpha protein	287 ± 1952	1490 ± 1675	27124 ± 146868	<0.001	0.002	0.195
Heat shock 70 kDa protein 1A	244 ± 1044	1190 ± 1379	12440 ± 78706	<0.001	0.003	0.083
Heat shock cognate 71 kDa protein	5838 ± 30168	18610 ± 13963	69018 ± 160844	0.002	0.003	0.382
Histone H2A type 3	5148 ± 7319	6517 ± 42825	56331 ± 130985	0.001	0.005	0.574
Immunoglobulin A	39421 ± 73514	30451 ± 67150	222417 ± 221679	0.009	0.007	0.878
Interstitial collagenase	55.8 ± 1.30	94 ± 5270	890 ± 33547	<0.001	0.009	0.130
Macrophage-capping protein	1917 ± 6597	1899 ± 1555	33978 ± 107421	0.004	0.001	0.878
Mitogen-activated protein kinase 9	371 ± 478	346 ± 21489	55799 ± 136700	<0.001	0.001	0.574
Mothers against decapentaplegic homolog 3	164 ± 535	204 ± 189	1162 ± 1576	0.009	0.004	0.645
Nucleoside diphosphate kinase A	1850 ± 10050	3016 ± 4232	45384 ± 429.0	0.003	0.007	0.645
Proteasome activator complex subunit 1	1800 ± 7904	1590 ± 1585	27296 ± 74459	0.005	0.001	0.798
Proteasome activator complex subunit 3	54 ± 57	73 ± 244	773 ± 2986	0.001	0.007	0.382
Protein E7_HPVI8	96 ± 131	247 ± 6253	17510 ± 51559	<0.001	0.001	0.195
Pulmonary surfactant-associated protein D	9529 ± 45588	15407 ± 33968	185604 ± 371714	0.005	0.007	1.000
Ras GTPase-activating protein 1	55 ± 225	74 ± 49	486 ± 1565	0.004	<0.001	0.721
Small nuclear ribonucleoprotein F	141 ± 295	336 ± 183	1155 ± 9453	0.002	0.007	0.279
Stress-induced-phosphoprotein 1	1331 ± 15285	1889 ± 3156	31714 ± 142272	0.005	0.001	0.721
Tissue-type plasminogen activator	1828 ± 1290	360 ± 724	2326 ± 15039	0.004	0.004	0.645
Transgelin-2	3383 ± 10436	6614 ± 5343	50796 ± 135169	0.001	0.001	0.645
Tumor necrosis factor receptor superfamily member 13C	180 ± 593	324 ± 242	1649 ± 3945	0.002	0.001	0.382
Ubiquitin carboxyl-terminal hydrolase isozyme L1	791 ± 1828	991 ± 676	76223 ± 38825	0.004	0.001	0.959

<https://doi.org/10.1371/journal.pone.0235328.t002>

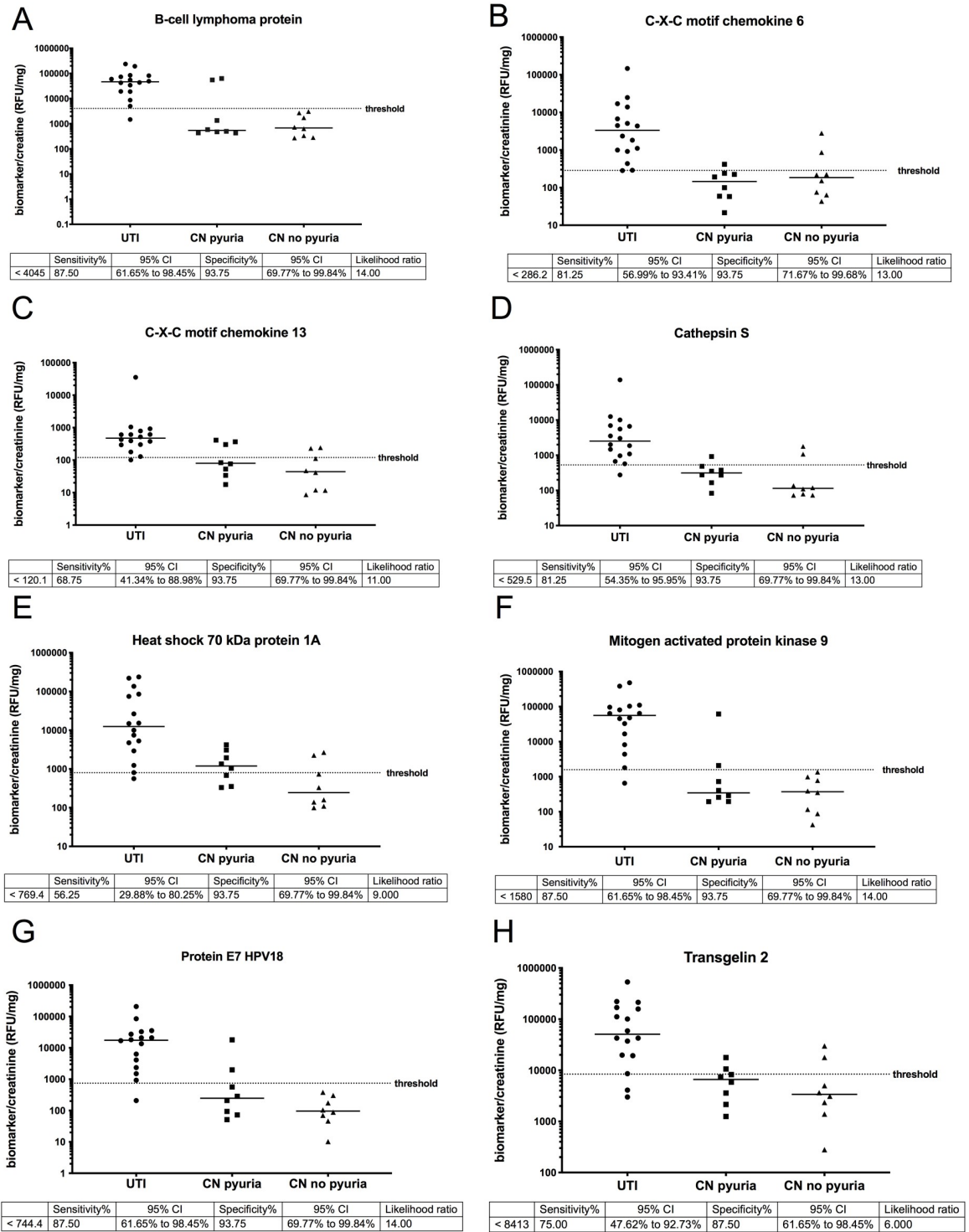
performance. A total of 45 most frequently occurring urine proteins were selected during the feature selection process, with 29% of them overlapping with each other. As shown by the Venn diagram, the overlapped urine protein biomarkers selected by at least two methods include *MAPK9*, *CXCL1*, *CXCL6*, *CXCL13*, *HSPA1A*, *E7*, *TYK2*, *PAME3*, *BCL6*, *LTF*, *HIST3H2A* and *SUMO3* (Fig 3).

The best AUC score was achieved with the SVM classifier with a radial basis function kernel (AUC score of 0.91). SVM worked best with the dataset consisting of Random forest algorithm selected urine proteins. The thirteen most frequently occurring proteins identified in feature



**Fig 1. Candidate biomarker ROCs: Area under the curve (AUCs) demonstrating the 8 candidate biomarkers that meet p value filtering criteria and had AUCs > 0.9.** B-cell lymphoma protein (A), C-X-C motif chemokine 6 (B) C-X-C motif chemokine 13 (C), Cathepsin S (D), Heat shock 79kDA protein 1A (E), Mitogen activated protein kinase (F), Protein E7 HPV18 (G) and Transgelin 2 (H) are presented.

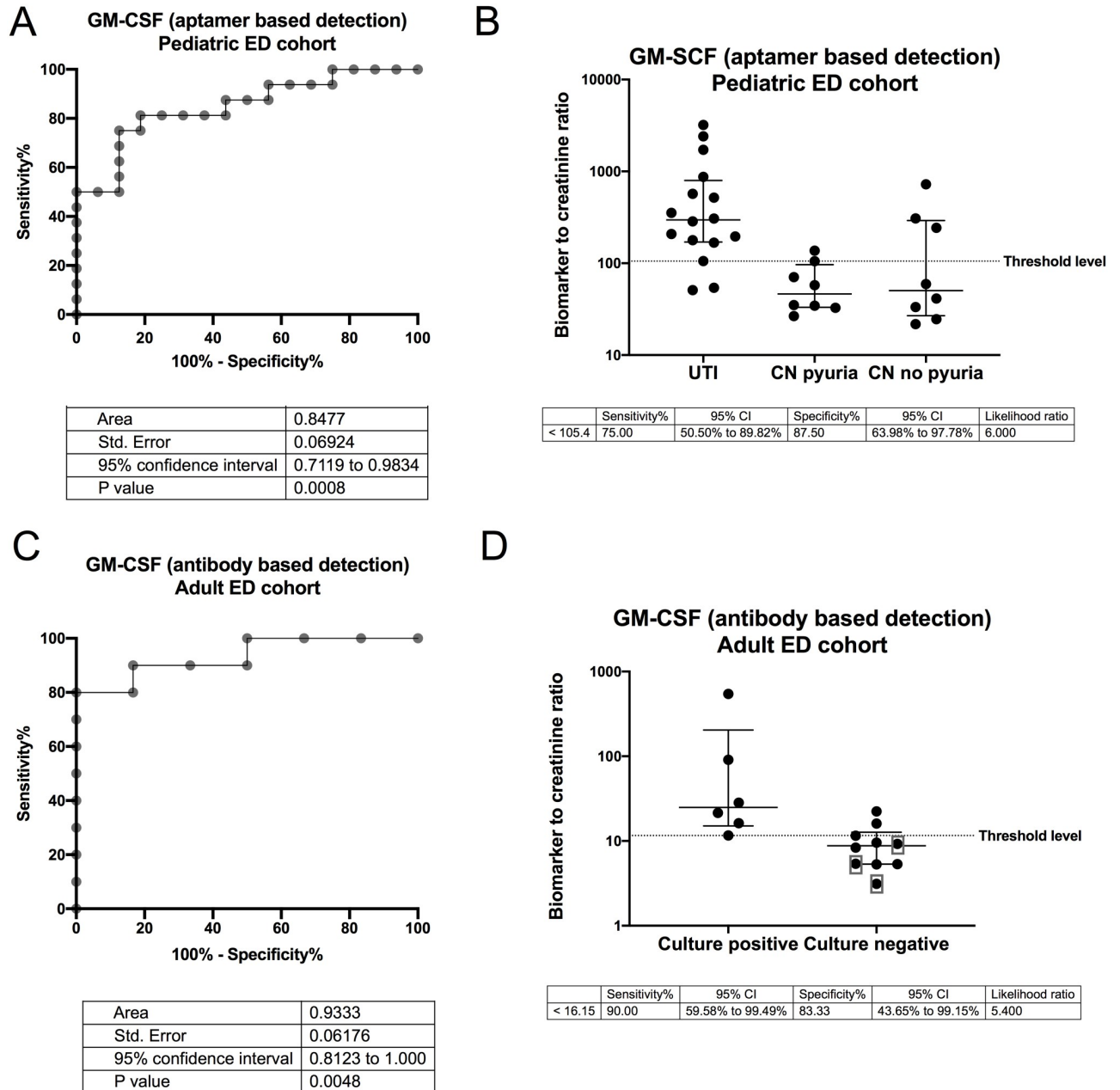
<https://doi.org/10.1371/journal.pone.0235328.g001>



**Fig 2. Candidate biomarkers scatterplots:** Scatter plots of urine biomarkers that met p value and AUC criteria are presented to show threshold values that differentiate between UTI and no UTI (CN pyuria and CN no pyuria urine). The CN pyuria and CN no pyuria samples were separated for graphical, but not for determination of the likelihood ratio (LR). Threshold levels and LRs are presented for B-cell lymphoma protein (A), C-X-C motif chemokine 6 (B) C-X-C motif chemokine 13 (C), Cathepsin S (D), Heat shock 79kDa protein 1A (E), Mitogen activated protein kinase (F), Protein E7 HPV18 (G) and Transgelin 2 (H) Y axis units are presented as relative fluorescent units (RFU) of biomarker/ mg creatinine.

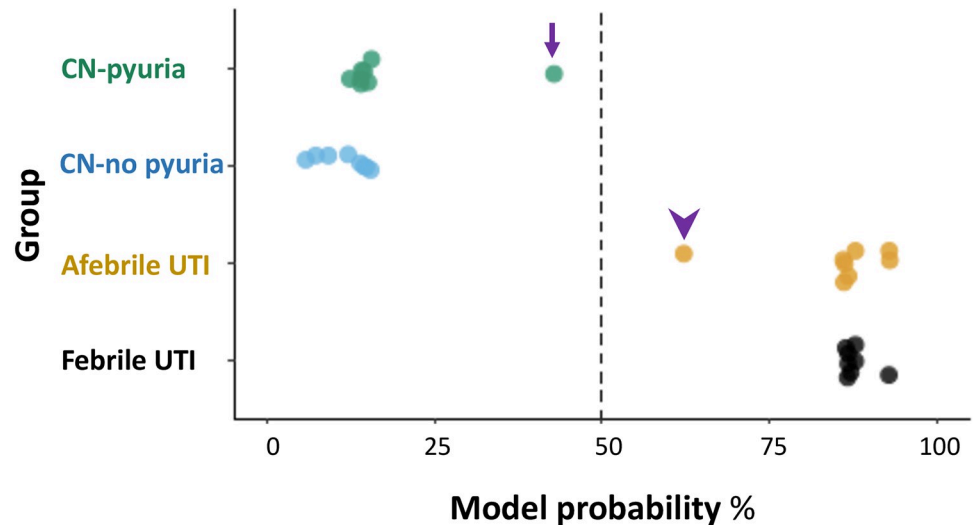
<https://doi.org/10.1371/journal.pone.0235328.g002>





**Fig 3.** Scatter plots for GM-CSF measured adult ED urine samples using an antibody based Mesoscale platform (A) and measured in pediatric ED urine samples using an aptamer based Somalogic platform (B) are presented. AUCs for the antibody based (C) and the aptamer-based platforms were similar. Units for aptamer-based detection are RFU of urine GM-CSF per mg of urine creatinine and units for antibody based detection are pg of urine GM-CSF per mg of urine creatinine. The patients were selected from an adult population from the Ohio State University Wexner Medical Center Emergency Department and included 6 culture positive and 10 culture negative samples. The mean (range) ages in years were 74 (65–89) for the culture positive and 62 (47–74) for the culture negative group. The majority, 4/6 (67%) in the culture positive and 8/10 (80%) in the culture negative group were female. Urine cultures grew *E. coli* in four patients, *E. Faecalis* in one patient and both *E. coli* and *E. Faecalis* in one patient. Seven of the culture negative group had no growth and three represented by gray boxes grew mixed flora. Of note these results may not differentiate UTI from asymptomatic bacteriuria which will need investigated in future studies.

<https://doi.org/10.1371/journal.pone.0235328.g003>



**Fig 4. The UTI class probability estimate for each sample by the optimal SVM classifier.** The dashed black line shows where the 50% probability lies. Generally, model probability of predicting UTI samples was  $> 80\%$ . There are 2 outliers, one 18-year old female with CN pyuria (purple arrow) who presented with left flank pain, fever and dysuria along with 1+ LE on UA and had 43.4% UTI probability. The other outlier (purple arrowhead) was a 3-year-old female who presented with fever and abdominal pain, along with 1+ LE and had 62.7% UTI probability.

<https://doi.org/10.1371/journal.pone.0235328.g004>

selection during the 5-fold cross-validation process are shown in Fig 4A and S4 Material. The UTI class probability estimate calculated based on the SVM decision values for each sample are presented as Fig 4B.

### Source material

The source data is presented as S5 Material.

### Discussion

Analysis of the human urine is the 1<sup>st</sup> known type of laboratory medicine and dates back to 4,000 BCE [24]. Hippocrates (460–355 BCE) associated increasing urine sediment with increasing fevers [24]. If the aforementioned urine sediment was due to WBCs this would be the earliest known description of a UTI biomarker [24]. Urine test strips, sometimes referred to urine dipsticks were developed in the 1950s and 1960s and have been used since then in the point-of-care diagnosis of UTIs [25]. However, UAs have limitations regarding sensitivity and specificity and its key UTI diagnostic components detects LE in the urine, a finding not necessarily specific to UTIs. Despite the need for more judicious use of antibiotics secondary to increasing rates of antibiotic resistant bacteria, point of care diagnosis of UTIs has remained largely unchanged since the introduction of urine dipsticks. One strategy is to detect bacterial products such as bacterial nuclease activity, however identifying the bacterial load indicative of a UTI may be problematic because urine contains a microbiota [26]. We and others have identified increased urine levels of innate immune proteins in the urine compared to normal controls [27, 28]. However, these innate immune proteins might not associate with UTIs if compared to the urine of ill patients without UTI. Here we use patients for which a urine dipstick and culture was obtained for clinical indications highlighting advantages of an emergence department for biomarker studies. Specifically, we identified a protein profile that differentiates UTI from children who had similar symptoms but did not have a UTI (e.g. CN-no pyuria and CN-pyuria).

We identified 8 proteins that were significantly ( $p < 0.01$ ) elevated in UTI samples compared to CN-pyuria and CN-no pyuria samples with “excellent” biomarker potential based on an AUC of 0.9–1 (Fig 1). Some of the candidate proteins (Fig 1) have been associated with bacterial interactions with mucosal surfaces other than the urinary tract. Cathepsin S (CTSS) expression is upregulated during periodontal infections [29]. Transgelin 2 mimics bacterial SipA, a protein that promotes bacterial entry into cells, and promotes phagocytosis in lipopolysaccharide activated macrophages [30]. C-X-C motif chemokine 13 is required for recruitment of specialized B cells, antibody production and the bacterial defense of the peritoneal and pleural cavities [31]. The involvement of cathepsin S, transgelin 2 and C-X-C motif chemokine 13 with other infections provides a foundation for the evaluation of the potential role of these proteins in *E.coli* UTI pathophysiology while other proteins have been linked to viral infections. B cell lymphoma protein 6 was initially described for its regulation of lymphocyte growth and development, but has been demonstrated to function as a checkpoint regarding the initiation of the innate immune response to cytosolic RNA viruses [32]. Mitogen protein kinase 9 and Protein E7 HPV18 are also involved in innate response to viral infections [33, 34]. Past studies of the human virome have had variable results regarding increased Protein E7 HPV18 expression during UTI [35, 36]. HPV18 is included in the vaccine for this virus, however 24/32 (75%) of included patients were  $< 11$  years of age, younger than the recommend age for the HPV vaccine [37]. It is possible that Protein E7 HPV18 represents a virus with homologous regions such as adenovirus E1a [38]. Other proteins not included in our top 8 candidate biomarkers were significantly elevated in the UTI group, but had AUCs  $< 0.9$  (Table 1). In some cases, these proteins have more established roles in UTI pathophysiology. Pulmonary surfactant-associated protein D (SP-D) inhibits the growth of uropathogenic *E.coli* and regulates renal inflammation via the p38 MAPK related pathway during UTI [39]. Granulocyte macrophage colony-stimulating factor (GM-CSF) has been shown to be expressed by murine urothelial cells in response to lipopolysaccharides [40]. Our findings provide further biological relevance for SP-D and GM-CSF in human UTIs. We speculate that a panel containing some of the aforementioned biomarkers might lead to improved UTI diagnosis compared to urine dipstick results. Further we obtained similar results using aptamer and antibody-based GM

Current urine dipstick markers of host immune response (e.g. LE) are limited to proteins produced by WBCs, and thus may be produced nonspecifically when WBCs are present in the urine. In our previously reported larger cohort of 199 patients that we selected our samples from for this study, had a sensitivity of 83% and specificity of 85%. Other pediatric meta-analyses/studies reported sensitivities of 72–83% and specificities of 78–87% for LE. In comparison B-cell lymphoma protein and mitogen activated protein kinase 9 each had a sensitivity of 88% and specificity of 94%. Immunohistochemistry images for from the Human Protein Atlas version 18.1 (<https://www.proteinatlas.org>) are consistent with mitogen activated protein 5, along with C-X-C chemokine ligand 13 and heat shock protein are expression in the spleen, bladder lumen and collecting duct of the kidney [41–43]. We have previously demonstrated that the renal collecting duct, the initial kidney tubular section encountered by ascending uropathogens has innate immune functions [44]. Since many of the biomarkers that we evaluated are expressed in the bladder and kidney, in addition to the spleen (e.g. myeloid cells), they may represent a more specific innate immune response to infection compared to white blood cell limited proteins such as the LE.

To the best of our knowledge, this is the first study that uses Somacan™ proteomics data to construct a machine learning predictive model for urinary tract infection. In this study, we explored the application of SVM classifier in solving the classification problem on proteomics data. To obtain an unbiased performance estimation, we have adopted a nested cross-validation approach that performing hyperparameter tuning and model optimization in the inner

cross-validation loop and evaluated the optimal models independently in the outer cross-validation loop [45]. This design avoids the optimistic bias introduced into the performance estimate due to the use of the same cross-validation procedure for both hyperparameter optimization and performance evaluation [46]. Our SVM model had a slightly lower AUC than for some of the individual proteins. This is likely because for the SVM model that divided our results between a test and validation cohort. We anticipate that the SVM model will outperform single biomarkers in future studies with many more samples. It is also possible that our SVM model may be more accurate than urine culture. The patient that was assigned to the CN pyuria with a UTI probability of score of 43.4% (Fig 4) presented with left flank pain, fever, dysuria and UTI history; we speculate that this patient might have had an actual UTI with an organism that could not be isolated in standard urine. In the future, enhanced urine culture and sequencing could be applied to similar samples to help determine if these actually represent culture negative UTIs [7].

There are limitations with this study. First, we did not attempt to identify a biomarker to distinguish between pyelonephritis and cystitis in this study because we did not have radiologic evidence to distinguish between these conditions. Second, because of the expense of analyzing >1300 proteins rather than a targeted study comparing 1–10 proteins, a relatively small number of urine samples, 32, were analyzed and stringent filtering criteria such as using a p value of 0.1 rather than 0.05 and an AUC of 0.9 rather than 0.8. With this strict criterion we may have excluded some relevant biomarkers. Third, the Somascan™ platform does not include all proteins. For example, human alpha defensin 5 and human neutrophil proteins 1–3 which we reported on in 2016 in a hypothesis based UTI biomarker study are not included in the Somascan™ platform [17]. Fourth, the Somascan™ platform reports results a relative fluorescent units (RFU/ml); use of additional methodologies such as ELISA will be needed for further quantification of the candidate proteins. Fifth, future studies will be needed to see if different populations such as children less than 2 years of age, patients with urine collected by catheter or patients with positive urine cultures and no LE have different urine biomarker profiles. Last, we limited our pediatric UTI samples to cultures from which *E.coli*, the bacteria species that accounts for 75%–90% of UTIs, was isolated, it is possible that other bacteria species may result in distinct proteomic patterns [47].

LE is a nonspecific marker for pyuria while nitrites are a bacterial product. LE may not differentiate culture negative pyuria from UTI and nitrites might not differentiate asymptomatic bacteriuria from UTI. Addition of urine levels of renal urothelium and collecting duct expressed innate immune proteins adds a distinct biomarker mechanism to current UTI diagnosis methodology. Our 8 leading proteins had AUCs between 0.91 and 0.95 indicating that they represent “excellent” biomarker candidates [20]. We propose that including the urine levels of these candidate biomarkers, either alone or in combination with traditional UA biomarkers will help clinicians identify true UTI from culture negative pyuria at the point of care. Future directions will be testing a larger number of patients with ELISAs to further refine our SVM model. Ultimately any biomarker panel will need to be converted to an aptamer-based point of care test and validated with an antibody-based assay, similar to what was done with GM-CSF in this study.

## Supporting information

**S1 Material. Primers used for RT-PCR.**  
(PDF)

**S2 Material. Illustration of machine learning process using a nested cross-validation strategy.** The support vector machine (SVM) models are trained and selected in an inner leave-

one-out cross-validation loop, which involves model-based feature selection and SVM hyperparameter optimization. An outer 5x cross-validation loop evaluates the generalized performance of the optimal model selected from the inner loop on the test set.

(PDF)

**S3 Material. Table of proteins that were statistically different in UTI vs CN no pyuria urine and UTI versus CN pyuria with a p value of < 0.05 but were not statistically different between CN pyuria and CN no pyuria samples.**

(XLSX)

**S4 Material. Expression patterns of the 17 proteins selected by random forest algorithm presented as heatmap with UTI versus non-UTI (CN pyuria and CN no pyuria).**

(PDF)

**S5 Material. Source data.**

(XLS)

## Acknowledgments

We would like to acknowledge Jennifer Kline (Nationwide Children's) and Michael Hill (The Ohio State University) for coordinating sample collection.

## Author Contributions

**Conceptualization:** Joshua Watson, David S. Hains, Andrew L. Schwaderer.

**Data curation:** Joshua Watson, Samuel Arregui, John Ketz.

**Formal analysis:** Liang Dong, Sha Cao, Jeffrey M. Caterino, David S. Hains, Andrew L. Schwaderer.

**Funding acquisition:** Andrew L. Schwaderer.

**Investigation:** Samuel Arregui, Vijay Saxena, Abduselam K. Awol, David S. Hains, Andrew L. Schwaderer.

**Methodology:** Liang Dong, Vijay Saxena, Andrew L. Schwaderer.

**Project administration:** Andrew L. Schwaderer.

**Resources:** Andrew L. Schwaderer.

**Supervision:** Joshua Watson, Daniel M. Cohen.

**Writing – original draft:** Liang Dong, Andrew L. Schwaderer.

**Writing – review & editing:** Joshua Watson, Sha Cao, Samuel Arregui, Vijay Saxena, John Ketz, Abduselam K. Awol, Daniel M. Cohen, Jeffrey M. Caterino, David S. Hains, Andrew L. Schwaderer.

## References

1. Simmering JE, Tang F, Cavanaugh JE, Polgreen LA, Polgreen PM. The Increase in Hospitalizations for Urinary Tract Infections and the Associated Costs in the United States, 1998–2011. *Open Forum Infect Dis.* 2017; 4(1):ofw281. <https://doi.org/10.1093/ofid/ofw281> PMID: 28480273.
2. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am J Med.* 2002; 113 Suppl 1A:5S–13S. [https://doi.org/10.1016/s0002-9343\(02\)01054-9](https://doi.org/10.1016/s0002-9343(02)01054-9) PMID: 12113866.

3. Poole NM, Shapiro DJ, Fleming-Dutra KE, Hicks LA, Hersh AL, Kronman MP. Antibiotic Prescribing for Children in United States Emergency Departments: 2009–2014. *Pediatrics*. 2019. <https://doi.org/10.1542/peds.2018-1056> PMID: 30622156.
4. Sanchez GV, Babiker A, Master RN, Luu T, Mathur A, Bordon J. Antibiotic Resistance among Urinary Isolates from Female Outpatients in the United States in 2003 and 2012. *Antimicrob Agents Chemother*. 2016; 60(5):2680–3. <https://doi.org/10.1128/AAC.02897-15> PMID: 26883714.
5. Subcommittee on Urinary Tract Infection SCoQI, Management, Roberts KB. Urinary tract infection: clinical practice guideline for the diagnosis and management of the initial UTI in febrile infants and children 2 to 24 months. *Pediatrics*. 2011; 128(3):595–610. Epub 2011/08/30. <https://doi.org/10.1542/peds.2011-1330> PMID: 21873693.
6. Visser VE, Hall RT. Urine culture in the evaluation of suspected neonatal sepsis. *J Pediatr*. 1979; 94(4):635–8. Epub 1979/04/01. [https://doi.org/10.1016/s0022-3476\(79\)80040-2](https://doi.org/10.1016/s0022-3476(79)80040-2) PMID: 430312.
7. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, et al. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol*. 2014; 52(3):871–6. <https://doi.org/10.1128/JCM.02876-13> PMID: 24371246.
8. Luciano R, Piga S, Federico L, Argentieri M, Fina F, Cuttini M, et al. Development of a score based on urinalysis to improve the management of urinary tract infection in children. *Clin Chim Acta*. 2012; 413(3–4):478–82. Epub 2011/11/29. <https://doi.org/10.1016/j.cca.2011.11.005> PMID: 22120731.
9. Wise GJ, Schlegel PN. Sterile pyuria. *N Engl J Med*. 2015; 372(11):1048–54. <https://doi.org/10.1056/NEJMra1410052> PMID: 25760357.
10. Solier C, Langen H. Antibody-based proteomics and biomarker research—current status and limitations. *Proteomics*. 2014; 14(6):774–83. Epub 2014/02/13. <https://doi.org/10.1002/pmic.201300334> PMID: 24520068.
11. Sakamoto S, Putalun W, Vimolmangkang S, Phoolcharoen W, Shoyama Y, Tanaka H, et al. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. *J Nat Med*. 2018; 72(1):32–42. Epub 2017/11/23. <https://doi.org/10.1007/s11418-017-1144-z> PMID: 29164507.
12. C W.C. S TK B JG. The Point behind Translation of Aptamers for Point of Care Diagnostics. *Aptamers and Synthetic Antibodies*. 2017; 2(1):36–42.
13. Hensley P. SOMAmers and SOMAscan—A Protein Biomarker Discovery Platform for Rapid Analysis of Sample Collections From Bench Top to the Clinic. *J Biomol Tech*. 2013; 24(Suppl)(S5).
14. Zhang W, Zhang R, Zhang J, Sun Y, Leung PS, Yang GX, et al. Proteomic analysis reveals distinctive protein profiles involved in CD8(+) T cell-mediated murine autoimmune cholangitis. *Cell Mol Immunol*. 2018. <https://doi.org/10.1038/cmi.2017.149> PMID: 29375127.
15. Kaur H, Bruno JG, Kumar A, Sharma TK. Aptamers in the Therapeutics and Diagnostics Pipelines. *Theranostics*. 2018; 8(15):4016–32. Epub 2018/08/22. <https://doi.org/10.7150/thno.25958> PMID: 30128033.
16. Toh SY, Citartan M, Gopinath SC, Tang TH. Aptamers as a replacement for antibodies in enzyme-linked immunosorbent assay. *Biosens Bioelectron*. 2015; 64:392–403. Epub 2014/10/04. <https://doi.org/10.1016/j.bios.2014.09.026> PMID: 25278480.
17. Watson JR, Hains DS, Cohen DM, Spencer JD, Kline JM, Yin H, et al. Evaluation of novel urinary tract infection biomarkers in children. *Pediatr Res*. 2016; 79(6):934–9. <https://doi.org/10.1038/pr.2016.33> PMID: 26885759.
18. Shaikh N, Shope TR, Hoberman A, Vigliotti A, Kurs-Lasky M, Martin JM. Association Between Uropathogen and Pyuria. *Pediatrics*. 2016; 138(1). <https://doi.org/10.1542/peds.2016-0087> PMID: 27328921.
19. Kusumi K, Ketz J, Saxena V, Spencer JD, Safadi F, Schwaderer A. Adolescents with urinary stones have elevated urine levels of inflammatory mediators. *Urolithiasis*. 2019; 47(5):461–6. Epub 2019/04/18. <https://doi.org/10.1007/s00240-019-01133-1> PMID: 30993354.
20. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics*. 2013; 9(2):280–99. <https://doi.org/10.1007/s11306-012-0482-9> PMID: 23543913.
21. Pok G, Liu JC, Ryu KH. Effective feature selection framework for cluster analysis of microarray data. *Bioinformatics*. 2010; 4(8):385–9. <https://doi.org/10.6026/97320630004385> PMID: 20975903.
22. Kononenko I, editor *Estimating attributes: Analysis and extensions of RELIEF1994*; Berlin, Heidelberg: Springer Berlin Heidelberg.
23. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *ADVANCES IN LARGE MARGIN CLASSIFIERS*. 1999:61–74.

24. Armstrong JA. Urinalysis in Western culture: a brief history. *Kidney Int.* 2007; 71(5):384–7. <https://doi.org/10.1038/sj.ki.5002057> PMID: 17191081.
25. Voswinckel P. A marvel of colors and ingredients. The story of urine test strip. *Kidney Int Suppl.* 1994; 47:S3–7. PMID: 7869669.
26. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat Commun.* 2018; 9(1):1557. <https://doi.org/10.1038/s41467-018-03968-5> PMID: 29674608.
27. Forster CS, Jackson E, Ma Q, Bennett M, Shah SS, Goldstein SL. Predictive ability of NGAL in identifying urinary tract infection in children with neurogenic bladders. *Pediatr Nephrol.* 2018; 33(8):1365–74. <https://doi.org/10.1007/s00467-018-3936-0> PMID: 29532235.
28. Spencer JD, Schwaderer AL, Wang H, Bartz J, Kline J, Eichler T, et al. Ribonuclease 7, an antimicrobial peptide upregulated during infection, contributes to microbial defense of the human urinary tract. *Kidney Int.* 2013; 83(4):615–25. <https://doi.org/10.1038/ki.2012.410> PMID: 23302724.
29. Memmert S, Damanaki A, Nogueira AVB, Eick S, Nokhbehsaim M, Papadopoulou AK, et al. Role of Cathepsin S in Periodontal Inflammation and Infection. *Mediators Inflamm.* 2017; 2017:4786170. <https://doi.org/10.1155/2017/4786170> PMID: 29362520.
30. Jun CD, Kim HR. Transgelin-2 mimics bacterial SipA and promotes membrane ruffling and phagocytosis in lipopolysaccharide-activated macrophages. *Journal of Immunology.* 2016; 196.
31. Ansel KM, Harris RB, Cyster JG. CXCL13 is required for B1 cell homing, natural antibody production, and body cavity immunity. *Immunity.* 2002; 16(1):67–76. [https://doi.org/10.1016/s1074-7613\(01\)00257-6](https://doi.org/10.1016/s1074-7613(01)00257-6) PMID: 11825566.
32. Xu F, Kang Y, Zhuang N, Lu Z, Zhang H, Xu D, et al. Bcl6 Sets a Threshold for Antiviral Signaling by Restraining IRF7 Transcriptional Program. *Sci Rep.* 2016; 6:18778. <https://doi.org/10.1038/srep18778> PMID: 26728228.
33. Chu WM, Ostertag D, Li ZW, Chang L, Chen Y, Hu Y, et al. JNK2 and IKKbeta are required for activating the innate response to viral infection. *Immunity.* 1999; 11(6):721–31. [https://doi.org/10.1016/s1074-7613\(00\)80146-6](https://doi.org/10.1016/s1074-7613(00)80146-6) PMID: 10626894.
34. Borysiewicz LK, Fiander A, Nimako M, Man S, Wilkinson GW, Westmoreland D, et al. A recombinant vaccinia virus encoding human papillomavirus types 16 and 18, E6 and E7 proteins as immunotherapy for cervical cancer. *Lancet.* 1996; 347(9014):1523–7. [https://doi.org/10.1016/s0140-6736\(96\)90674-1](https://doi.org/10.1016/s0140-6736(96)90674-1) PMID: 8684105.
35. Moustafa A, Li W, Singh H, Moncera KJ, Torralba MG, Yu Y, et al. Microbial metagenome of urinary tract infection. *Sci Rep.* 2018; 8(1):4333. <https://doi.org/10.1038/s41598-018-22660-8> PMID: 29531289.
36. Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT. The human urine virome in association with urinary tract infections. *Front Microbiol.* 2015; 6:14. <https://doi.org/10.3389/fmicb.2015.00014> PMID: 25667584.
37. Petrosky E, Bocchini JA Jr., Hariri S, Chesson H, Curtis CR, Saraiya M, et al. Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR Morb Mortal Wkly Rep.* 2015; 64(11):300–4. PMID: 25811679.
38. Barbosa MS, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. *EMBO J.* 1990; 9(1):153–60. PMID: 2153075.
39. Hu F, Ding G, Zhang Z, Gatto LA, Hawgood S, Poulain FR, et al. Innate immunity of surfactant proteins A and D in urinary tract infection with uropathogenic *Escherichia coli*. *Innate Immun.* 2016; 22(1):9–20. Epub 2015/10/30. <https://doi.org/10.1177/1753425915609973> PMID: 26511057.
40. Li Y, Lu M, Alvarez-Lugo L, Chen G, Chai TC. Granulocyte-macrophage colony-stimulating factor (GM-CSF) is released by female mouse bladder urothelial cells and expressed by the urothelium as an early response to lipopolysaccharides (LPS). *Neurourol Urodyn.* 2017; 36(4):1020–5. Epub 2016/06/24. <https://doi.org/10.1002/nau.23057> PMID: 27337494.
41. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347(6220):1260419. <https://doi.org/10.1126/science.1260419> PMID: 25613900.
42. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhorji G, et al. A pathology atlas of the human cancer transcriptome. *Science.* 2017; 357(6352). <https://doi.org/10.1126/science.aan2507> PMID: 28818916.
43. Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. *Science.* 2017; 356(6340). <https://doi.org/10.1126/science.aal3321> PMID: 28495876.

44. Saxena V, Fitch J, Ketz J, White P, Wetzel A, Chanley MA, et al. Whole Transcriptome Analysis of Renal Intercalated Cells Predicts Lipopolysaccharide Mediated Inhibition of Retinoid X Receptor alpha Function. *Sci Rep*. 2019; 9(1):545. <https://doi.org/10.1038/s41598-018-36921-z> PMID: 30679625.
45. Stone M. Cross-validators choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. 1974:111–47.
46. Cawley G. T N. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*. 2010; 11:2079–107.
47. Zhanel GG, Hisanaga TL, Laing NM, DeCorby MR, Nichol KA, Weshnoweski B, et al. Antibiotic resistance in *Escherichia coli* outpatient urinary isolates: final results from the North American Urinary Tract Infection Collaborative Alliance (NAUTICA). *Int J Antimicrob Agents*. 2006; 27(6):468–75. <https://doi.org/10.1016/j.ijantimicag.2006.02.009> PMID: 16713191.