OPEN

Discovery and Validation of Urinary Molecular Signature of Early Sepsis

Sabyasachi Bandyopadhyay, MS[1–4]; Nicholas Lysak, MD[3,5]; Lasith Adhikari, PhD[1,3]; Laura M. Velez, BS[1]; Larysa Sautina , MS[1,4]; Rajesh Mohandas, MD[1,6]; Maria-Cecilia Lopez, BS[4,7]; Ricardo Ungaro, BS[5]; Ying-Chih Peng, MS[3,8]; Ferdous Kadri , MD[1,3,6]; Philip Efron, MD[4,5]; Scott Brakenridge, MD[4,5]; Lyle Moldawer, PhD[4,5]; Frederick Moore, MD[4,5]; Henry V. Baker, MD[4,7]; Mark S. Segal, MD[1,4,6]; Tezcan Ozrazgat-Baslanti, PhD[1,3,4]; Parisa Rashidi, PhD[2,3]; Azra Bihorac, MD[1,3,4]

Objectives: Identify alterations in gene expression unique to systemic and kidney-specific pathophysiologic processes using whole-genome analyses of RNA isolated from the urinary cells of sepsis patients.

Design: Prospective cohort study.

Setting: Quaternary care academic hospital.

Patients: A total of 266 sepsis and 82 control patients enrolled between January 2015 and February 2018.

Interventions: Whole-genome transcriptomic analysis of messenger RNA isolated from the urinary cells of sepsis patients within 12 hours of sepsis onset and from control subjects.

Measurements and Main Results: The differentially expressed probes that map to known genes were subjected to feature selection using multiple machine learning techniques to find the best subset of probes that differentiates sepsis from control subjects. Using differential expression augmented with machine learning ensembles, we identified a set of 239 genes in urine, which show excellent effectiveness in classifying septic patients from those with chronic systemic disease in both internal and independent external validation cohorts. Functional analysis indexes disrupted biological pathways in early sepsis and reveal key molecular networks driving its pathogenesis.

Conclusions: We identified unique urinary gene expression profile in early sepsis. Future studies need to confirm whether this approach can complement blood transcriptomic approaches for sepsis diagnosis and prognostication.

Key Words: cells; gene expression; machine learning; messenger ribonucleic acid; sepsis; urine

[1]Department of Medicine, Division of Nephrology, Hypertension, and Renal Transplantation, University of Florida, Gainesville, FL.

[2]J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL.

[3]Precision and Intelligent Systems in Medicine (PrismaP), University of Florida, Gainesville, FL.

[4]Sepsis and Critical Illness Research Center, University of Florida, Gainesville, FL.

[5]Department of Surgery, University of Florida, Gainesville, FL.

[6]North Florida/South Georgia Veterans Health System, Gainesville, FL.

[7]Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL.

[8]Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

Crit Care Expl 2020; 2:e0195

DOI: 10.1097/CCE.0000000000000195

Sepsis, a dysregulated host immune response to infection leading to acute organ dysfunction, continues to be a significant challenge for the healthcare system (1). The introduction of the Surviving Sepsis Campaign (2) led to a decrease in hospital mortality rates, yet total sepsis deaths continue to rise, and its treatment carries significant resource consumption (3). Identification of altered molecular profiles and biochemical pathways in septic patients has generated enthusiasm for the discovery of novel blood biomarkers (4). Multiple studies have examined the gene expression of immune cells in the blood from sepsis patients (5), and the U. S. Food and Drug Administration recently approved the first diagnostic test for sepsis based on expression of four genes in peripheral blood (6).

Urine is a readily available biofluid that does not require invasive collection, yet its wealth of molecular information is underutilized. The kidney is one of the most commonly affected organs in sepsis with profound effect on outcomes (7) with both systemic and local inflammations playing a role in the pathophysiology of sepsis-induced renal injury (8). The kidney filters

150 L of circulating plasma daily to produce 1.5 L of urine where highly concentrated potential biomarkers, such as metabolites, proteins, and nucleic acids, reflect both renal and systemic pathologies (9). Immune and renal cells appearing in the urine in response to systemic or renal inflammations have been used as prognostic markers in lupus, systemic vasculitis, and glomerular diseases (10–15) but not in sepsis. Previous work with urine in sepsis has focused on protein biomarkers of acute kidney injury (16, 17), whereas urinary RNA has been tested mainly for the diagnosis of transplant rejection and urologic cancers (18–20). We tested the hypothesis that application of machine learning (ML) for whole-genome transcriptomic analysis of RNA isolated from the urinary cells of septic patients can be used to identify alterations in gene expression unique to systemic and kidney-specific pathophysiologic processes in sepsis. This work is not intended to provide a diagnostic tool for sepsis. The authors only want to highlight that cells leaching into the urine due to sepsis are enriched with molecular information capable of discerning sepsis from noninfected controls. **Figure 1*A*** shows how sepsis-associated kidney injury can lead to leaching of immune cells and pathogen/damage-associated molecular patterns into the urine from blood.

MATERIALS AND METHODS

Participants

Sepsis patients were prospectively recruited between January 2015 and August 2017 from a prospective longitudinal cohort of surgical patients with sepsis at the University of Florida Health (UFH) (NCT02276066) that examines the immunologic mechanisms of chronic critical illness in sepsis. For the control group, we used preoperative urine samples from patients prospectively recruited between July 2015 and February 2018 to a prospective observational study (Network Analysis of Urinary Molecular Signature Complements Clinical Data to Predict Postoperative Acute Kidney Injury [NavigateAKI]; NCT02114138), characterizing the urinary molecular response to surgical stress among patients undergoing high-risk vascular surgery at UFH (**Supplementary Fig. S1** http://links.lww.com/CCX/A265). The study protocols were finalized (21) and ethics approvals were obtained from the UF Institutional Review Board (IRB201400611 and IRB201400127) prior to the recruitment of patients. All study participants were provided written informed consent. There was no overlap of patients between the two cohorts. The informed consent form for NavigateAKI permits the usage of these data in a limited way in other research projects.

The inclusion criteria for the sepsis cohort were admission to the surgical ICU, greater than or equal to 18 years old, and a diagnosis of sepsis (clinically adjudicated by attending physician and investigators according to the American College of Chest Physicians consensus criteria [22]) with subsequent initiation of the computerized sepsis protocol (23). Excluded patients fell into three categories: 1) patients taking immunosuppressive drugs or with a history of autoimmune diseases, 2) patients with advanced liver or heart disease, and 3) patients whose primary cause of sepsis was end-stage renal disease or urinary tract infection. All control patients were adjudicated as having no evidence of infection prior to surgery by attending surgeons and investigators.

All relevant clinical data were prospectively collected. Severity of illness was defined within the first 24 hours using the Sequential Organ Failure Assessment (SOFA) score (21). Patient outcomes, including hospital and 12-month mortality, were prospectively recorded for both studies (24). The first blood and urine samples for experimental analyses were collected within 12 hours of sepsis onset for sepsis patients and within 4 hours prior to scheduled surgery for control patients.

Discovery and Validation Cohorts

The discovery cohort consisted of RNA isolated from 238 patients recruited between January 2015 and March 2016 (Supplementary Fig. S1 http://links.lww.com/CCX/A265). The validation cohort consisted of RNA isolated from 110 patients recruited between February 2017 and February 2018. Complete data were available for 146 sepsis and 32 control patients in the discovery cohort and 41 sepsis and 32 control patients in the validation cohort. This sample size enabled us to ensure that for at least 85% of probes, we have power greater than 80% to detect a two-fold change between the mean expressions for sepsis and control patients using a two-sided independent *t* test with Bonferroni adjustment at a family-wise type 1 error of 0.05.

Processing of Urine Samples and RNA Purification

Using standardized protocols to separate cell pellets from urine supernatant (Fig. 1*A*), approximately 50 mL of urine was collected in sterile manner at the bedside and processed within 2 hours of collection. We used previously described protocols to isolate total cellular RNA from the urinary cell pellet containing all cellular elements. In brief, the 50 mL of urine was spun down at 1,500 g for 30 minutes at 4°C. The pellet was collected, lysed using 1-mL rolling liquid transporter lysis buffer with 10-uL β-mercaptoethanol from the kit, and processed according the manufacturer's protocol. Total RNA was extracted using the RNeasy mini kit (Qiagen, Leusden, The Netherlands) (250) Catalog Number—74106 according to the manufacturer's protocol. To determine the quality of isolated cellular RNA, we measured the quantity (absorbance at 260 nm) and purity (ratio of absorbance at 260 and 280 nm). An RNA sample was classified as having passed quality control if the optical density 260:280 ratio was between 1.5 and 2.2 and final concentration was at least 8.7 μg/mL (25) (**Supplementary Table S1** http://links.lww.com/CCX/A275).

Microarrays

Biotin-labeled sense strand complementary DNA was prepared from 300 ng of total RNA per sample using an Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay per standard protocol (more details of which are provided in **Supplementary Methods** http://links.lww.com/CCX/A264). Hybridization to GeneChip Human Transcriptome Array 2.0 (Affymetrix, Thermo Fisher Scientific, Santa Clara, CA) was carried out at 45°C for 16 hours, and the arrays were scanned on an Affymetrix GeneChip Scanner 3000 7G using the Affymetrix GeneChip Command Console software, which produced a set of files with extensions .DAT, .CEL, .JPG, and .XML for each array. Image analysis and probe quantification were performed using the Affymetrix software that produced raw probe intensity data in the Affymetrix CEL files. Transcriptome Analysis

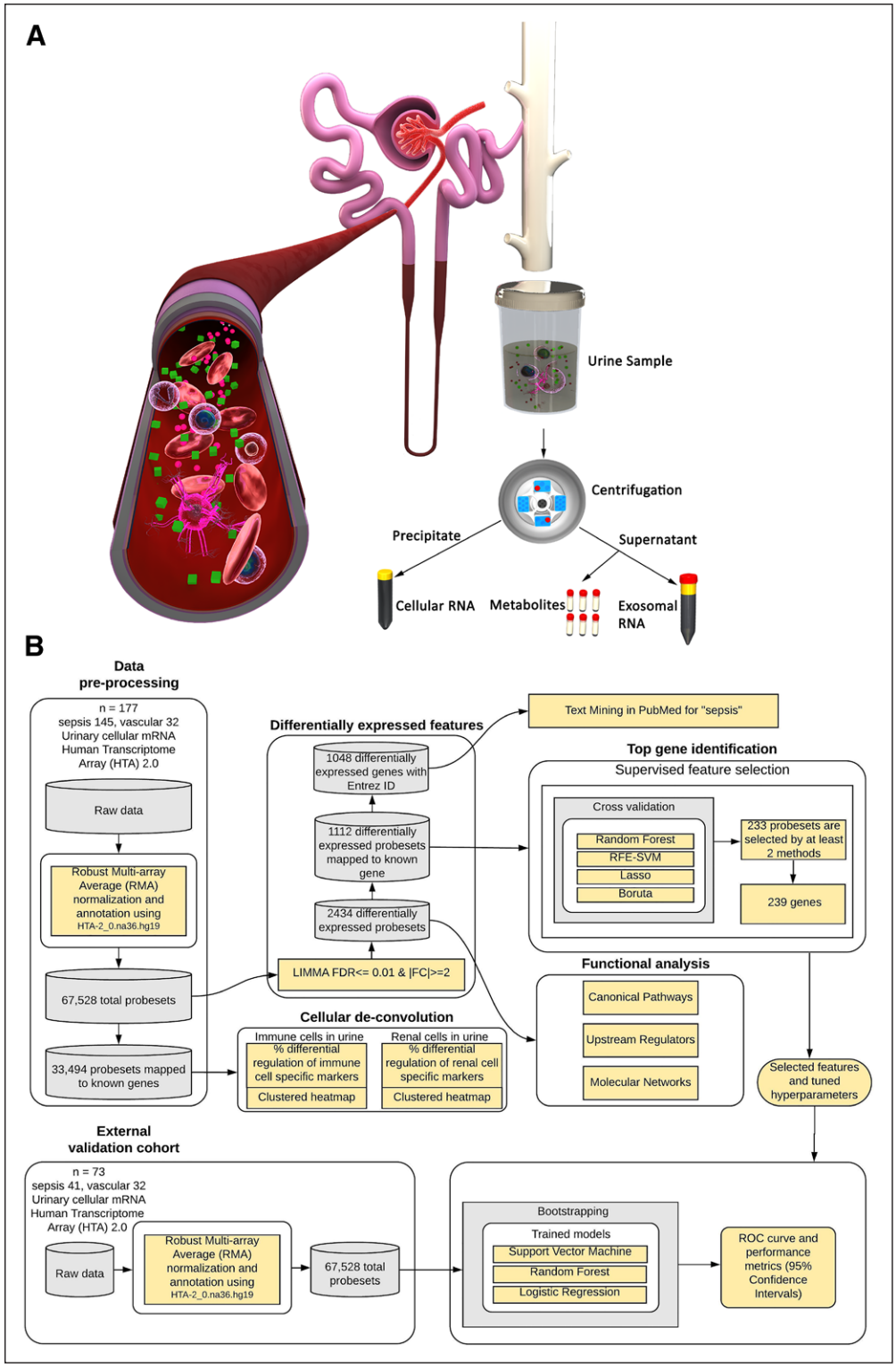

Figure 1. Workflow. **A**, Workflow for isolation of urinary markers. **B**, Conceptual workflow from data acquisition to analysis. FC = fold change, FDR = false discovery rate, ID = identity, LIMMA = linear models for microarray analysis, RFE-SVM = recursive feature elimination with support vector machine, ROC = receiver operating characteristics.

(available as GSE112098, GSE112099, and GSE112100 Gene Expression Omnibus series accessions).

Identification of Cell-Specific Transcripts

The 33,494 probes mapped to known genes were used to estimate the immune and kidney cell composition of the samples (Fig. 1*B*). The immune response in silico (IRIS) repository of 1,622 genes, classified by their specific expression in multiple immune cell lineages (27) and previously described transcript sets of 637 genes for kidney-specific cell lineages (28), was used to estimate the immune and renal cell composition in urine, respectively. Urine samples from 10 random septic patients were analyzed via flow cytometry using an Life Science Research II flow cytometer (Becton Dickinson, Franklin, NJ). Approximately 50–200 mL of urine was collected and processed within 30 minutes of sample collection. The samples were stained with CD3-AF488 (Number 557694; Becton Dickinson), CD4-AF700 (Number 566318; Becton Dickinson), CD8-BV650 (Number 565289; Becton Dickinson), CD14-PE (Number 561707; Becton Dickinson), CD19-APC (Number 561742; Becton Dickinson), and Sytox Blue (Number S34857; Invitrogen; Thermo Fisher Scientific, Waltham, MA).

Identification and Characterization of Discriminating Set of Genes in Sepsis

We applied empirical Bayes method in the linear models for microarray analysis (29) to identify differentially expressed probes between sepsis and control patients. The significance threshold was adjusted for multiple testings using the Benjamini-Hochberg false discovery rate (FDR) (30). Probes with an FDR of less than or equal to 0.01 and an absolute fold change greater than or equal to 2 were considered differentially expressed. Gene expression patterns were elucidated using Euclidean distance heatmaps with ComplexHeatmap (31). The ingenuity pathway analysis (IPA) software (http://www.ingenuity.com) was used to identify significantly enriched biologic functions, pathways, molecular networks,

Console Version 4.0.1 (Thermo Fisher Scientific, Santa Clara, CA) was used for microarray signal summarization and normalization (**Fig. 1*B***) using robust multiarray average (26). The final microarray dataset consisted of log2 transformed expression values for 67,528 probes of which 33,494 were mapped to one or more known genes

and regulatory molecules concerning the differentially expressed genes (Fig. 1*B*).

The differentially expressed probes were subjected to feature selection using an ensemble of four ML algorithms deployed in parallel (random forest [32], recursive feature elimination using support vector classifier [33], logistic regression with lasso [34], and Boruta [35]) to find the best subset of probes that differentiate sepsis from control subjects (Fig. 1*B*). For random forest, we ranked selected probes by their importance score (the set was truncated when the cumulative importance of the model upon inclusion of the next probe did not increase by 0.1%) to select 200 probes. Recursive feature elimination recursively dropped low-importance features from the support vector machine algorithm to select 200 probes. Lasso used L1-regularization (36) on the coefficients obtained from logistic regression to select 266 probes. Boruta selected 49 probes at the end of 100 iterations based on rankings provided by an internal random forest (37). Each of the methods was parameterized inside a five-fold cross validation design (**Supplementary Table S2** http://links.lww.com/CCX/A276). To obtain the final feature set from the ensemble, we used a voting strategy that retained probes that appear in at least two, three, or all four algorithms (Supplementary Methods http://links.lww.com/CCX/A264).

The final subset of probes was validated using independent validation cohort normalized separately from discovery cohort to prevent any information leakage. We employed three ML models (support vector machine, random forest, and logistic regression) that were trained and tuned on the discovery cohort. We calculated multiple performance metrics including area under curve, sensitivity, specificity, accuracy, and positive and negative predictive values. The 95% CIs for every performance metric in each model were estimated by bootstrapping the validation cohort without replacement 100 times.

We used R, version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria) and Python language, version 2.7 (Python Software Foundation, Fredericksburg, VA) as programming software and SAS, version 9.4 (SAS Institute, Cary, NC) for descriptive analyses. PubMed was searched using text mining to identify articles that match this final subset of genes to the keyword "sepsis" using the R package "rentrez" (38). The resulting articles were reviewed by authors (S. Bandyopadhyay, K.F., H.V.B., A.B.) to provide an overview of biologic functions of identified genes in sepsis. Boruta was implemented using "BorutaPy" package in Python. An automated analytic framework for the entire process in Figure 1*B* was implemented using Bioconductor (Version 3.7, Bioconductor Project, Roswell Park Comprehensive Cancer Center, NY) in R and scikit-learn (Version 0.19.2) (39) in Python and is available on Github at https://github.com/Prisma-pResearch/Urinary-signature-of-sepsis-.

RESULTS

Patient Cohorts

Compared with control patients, sepsis patients were younger but had similar comorbidity burden (**Table 1**). None of the control patients experienced sepsis within 7 days of surgery. **Supplementary Table S3** http://links.lww.com/CCX/A277 shows the different surgery types control patients were scheduled for. Proportion of patients who had preexisting kidney disease within sepsis and control cohorts was not significantly different. The groups did not differ in the SOFA score obtained on the day of urine sampling although sepsis patients had higher biomarkers of infection, as expected. For control patients, SOFA scores were obtained on the day of the surgery and included both preoperative and postoperative evaluations. The initial urine samples were collected within a median of 7 hours (range 3–11 hr) of sepsis onset. We excluded patients whose primary cause of sepsis was urinary tract infection, because these patients had a significantly higher total RNA mass that is indicative of a greater urinary cell count. The p value of a single-tailed t test assuming equal variance between the two groups was 0.008. F test showed that the two groups have equal variance with an F value of 2.41E-6, whereas the F critical is 1.714. Furthermore, we did sensitivity analysis, which showed that the differentially expressed gene sets with and without urinary tract infection (UTI) patients are nearly identical. If the UTI sepsis patients are added, eight genes, namely, *ADGRE2*, *ADGRE5*, *IFITM1*, *IFRD1*, *KLHL2*, *LYST*, *MAPK14*, and *STX11*, are added to the previous list of 1,048 differentially expressed genes, making the change insignificant.

The Acute Urinary Molecular Response to Sepsis

Within 12 hours of sepsis onset, we identified a distinct transcriptomic profile in the urinary cells retrieved from the fresh pellet with 2,434 (3.6%) of 67,528 probes being differentially expressed compared with control patients (FDR ≤ 0.01 and absolute fold change ≥ 2) (**Supplementary Fig. S2 *A–C*** http://links.lww.com/CCX/A266). Majority of probes were up-regulated (1,186 probes for 905 genes) compared with controls with a good separation in a principal component analysis (**Supplementary Fig. S2*D*** http://links.lww.com/CCX/A266).

The IPA functional analysis showed up-regulation of pathways related to innate immunity, actin cytoskeleton, cell cycle, protein synthesis, and presence of reactive oxygen species in sepsis. Nuclear factor of activated T cells in regulation of immune response, cell division cycle 42 signaling, neuroinflammation signaling pathway, fragment crystallizable gamma receptor-mediated phagocytosis in macrophages and monocytes, integrin and hypoxia signaling were the top five up-regulated canonical pathways in sepsis patients. The peroxisome proliferator-activated receptor alpha/retinoid X receptor alpha pathway was significantly down-regulated (**Fig. 2*A***). We used pathway overlap graph that connects pathways that have at least 10 molecules in common to reveal the presence of four major different clusters of genes related to innate immunity, cell cycle and metabolism, cell morphology, and motility and hypoxia (**Supplementary Fig. S3** http://links.lww.com/CCX/A267). Biofunctions concerning infection, cellular movement, and migration and leukocyte quantity, migration, invasion, and proliferation were significantly up-regulated in sepsis patients, whereas cell death, apoptosis, and necrosis were down-regulated compared with control preoperative patients (**Fig. 2*B***). Interferon-gamma, interleukin (IL)-1 beta, tumor necrosis factor, IL-6, and IL-5 were identified as key upstream regulators for the 1,048 genes (**Supplementary Table S4** http://links.lww.com/CCX/A278). The primary gene coexpression network was associated with lipid metabolism and molecular transport (**Supplementary Fig. S4** http://links.lww.com/CCX/A268). Approximately 23% of

TABLE 1. Clinical Characteristics of Patients in Discovery and Validation Cohorts

Variables	Discovery Cohort			Validation Cohort		
	Sepsis Patients (*n* = 145)	Control Patients (*n* = 32)	*p*	Sepsis Patients (*n* = 41)	Control Patients (*n* = 32)	*p*
Baseline characteristics						
Female sex, *n* (%)	67 (46)	9 (28)	0.076	17 (41)	16 (50)	0.488
Age, yr, mean (SD)	59 (15)	70 (9)	< 0.001	55 (18)	64 (11)	0.019
Race, *n* (%)			0.457			0.175
White	130 (90)	30 (94)		37 (90)	27 (84)	
African American	12 (8)	1 (3)		4 (10)	2 (6)	
Other	3 (2)	1 (3)		0 (0)	3 (9)	
Body mass index, median (25–75th)	29 (25–34)	26 (22–32)	0.064	29 (25–40)	27 (24–34)	0.114
Comorbidities, *n* (%)						
Charlson comorbidity index, median (25–75th)	1 (0–3)	1 (0–1)	0.084	1 (0–2)	1 (0–2)	0.826
Chronic kidney disease	19 (13)	7 (22)	0.267	6 (15)	7 (22)	0.5478
Hypertension	102 (70)	23 (72)	1	29 (71)	27 (84)	0.264
Diabetes	43 (30)	6 (19)	0.277	9 (22)	10 (31)	0.427
Chronic pulmonary disease	51 (35)	12 (38)	0.84	9 (22)	9 (28)	0.592
Congestive heart failure	23 (16)	5 (16)	1	6 (15)	8 (25)	0.37
Interfacility hospital transfer, *n* (%)	72 (50)	10 (31)	0.078	16 (39)	7 (22)	0.136
Time between sepsis onset and sample collection (hr), median (25–75th)	7 (3–11)	NA		7 (4–12)	NA	
Acuity at the time of sampling						
Sequential Organ Failure Assessment score, median (25–75th)	6 (3–8)	5 (3–8)	0.269	6 (3–7)	6 (5–8)	0.939
Primary sepsis source, *n* (%)						
Intra-abdominal sepsis	61 (42)	NA		18 (44)	NA	
Pneumonia	31 (21)	NA		8 (20)	NA	
Necrotizing soft-tissue infection	26 (18)	NA		7 (17)	NA	
Surgical site infection	19 (13)	NA		1 (2)	NA	
Other[a]	8 (6)	NA		7 (17)	NA	
Sepsis severity on enrollment, *n* (%)						
Sepsis 2 criteria						
Sepsis/severe sepsis	112 (77)	NA		33 (80)	NA	
Septic shock	33 (23)	NA		8 (20)	NA	
Sepsis 3 criteria						
Sepsis	108 (74)	NA		32 (78)	NA	
Septic shock	28 (19)	NA		5 (12)	NA	
Lactate (mmol/L), median (25–75th)	1.8 (1.3–2.9)	0.7 (0.6–1)	< 0.001	1.7 (1.2–2.5)	1.9 (1.1–4.6)	0.747
Serum creatinine (mg/dL), median (25–75th)	1.0 (0.7–1.5)	1.1 (0.9–1.3)	0.676	1.1 (0.9–1.7)	0.9 (0.7–1.1)	0.08
WBC count (thou/cu mm), median (25–75th)	17 (12–22)	10 (8–15)	< 0.001	19 (14–26)	10 (8–16)	< 0.001
Outcomes						
Hospital mortality, *n* (%)	11 (8)	1 (3)	0.697	6 (15)	0 (0)	0.032
Days in ICU, median (25–75th)	8 (4–18)	6 (4–10)	0.245	10 (5–15)	5 (3–11)	0.064
Days in hospital, median (25–75th)	18 (9–28)	11 (6–16)	0.016	17 (11–30)	9 (7–16)	< 0.001

NA = not available.

[a]Other primary sepsis source includes catheter-related bloods, empyema, bacteremia, and esophageal perforation.

Significance level is set to be 0.05.

1,048 genes were cross-referenced with sepsis-related literature in PubMed (**Fig. 2*C***) with greater than 150 citations associated with each of the top five cited genes.

Immune and Kidney Cell–Specific Transcripts in the Urine

IRIS deconvolution methodology was used to examine leukocyte populations in the urine cell pellet in the early sepsis. Deconvolution identified up-regulation of marker genes for neutrophils and monocytes and down-regulation for T-lymphocytes (**Fig. 3**, ***A*** and ***B*** and **Supplementary Table S5** http://links.lww.com/CCX/A279). We calculated an average expression of the signature transcripts of each immune subset and used it as a proxy for the relative amount of that cell type in the urine pellet. Cell proportions by in silico deconvolution demonstrated significant increases in neutrophils and monocytes in septic patients (**Supplementary Fig. S5*A*** http://links.lww.com/CCX/A269). Applying similar methodology on previously described signature transcripts for different nephron segments, we identified the up-regulation of marker genes for tubular epithelial cells from the collecting duct (**Fig. 3**, ***C*** and ***D*** and **Supplementary Table S6** http://links.lww.com/CCX/A280). Cell proportions demonstrated significant increases in epithelial cells from all nephron regions in sepsis patients (**Supplementary Fig. S5*B*** http://links.lww.com/CCX/A269). Flow cytometry of the urine samples from 10 randomly selected septic patients demonstrated the presence of both CD4 + (3.3% of all cells) and CD8 + T cells (0.7% of all cells), CD14 + macrophages (0.7% of all cells), and CD19 + B cells (1.6% of all cells) (**Supplementary Fig. S6** http://links.lww.com/CCX/A270).

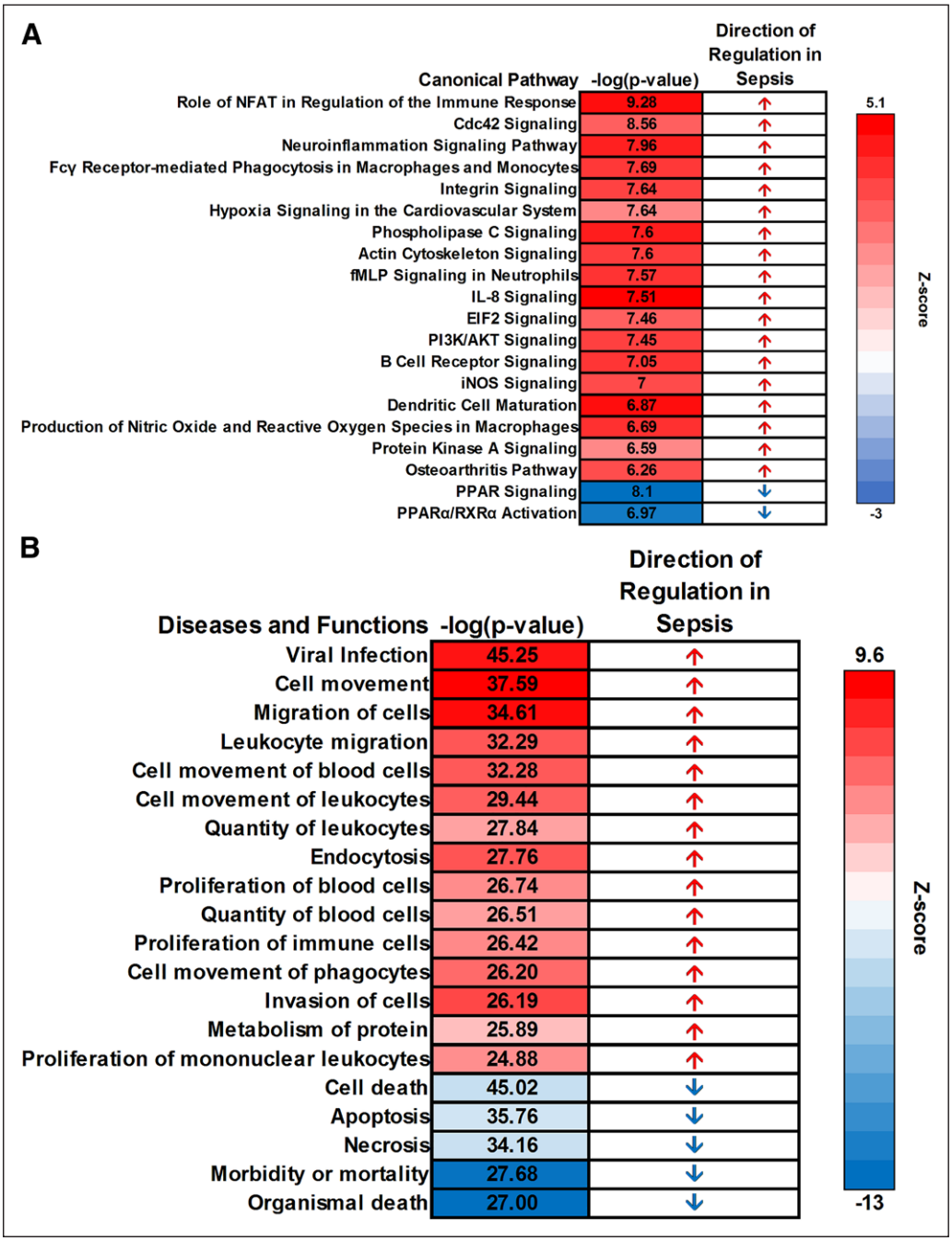

Figure 2. Pathways and biofunctions in the acute response to sepsis (within 12 hr of sepsis onset), compared with control patients. **A**, Ingenuity pathway analysis (IPA) of differentially expressed probes showed up-regulation of pathways mainly related to innate immunity, actin cytoskeleton, cell cycle and protein synthesis, and presence of reactive oxygen species in sepsis. The few pathways that were down-regulated in sepsis patients mainly corresponded to peroxisome proliferator-activated receptor pathway. *p* values are calculated by IPA software using the right-tailed Fisher exact test to measure likelihood that pathways or functions are overrepresented by molecules in dataset. **B**, Ingenuity disease and biofunction analysis of differentially expressed probes in sepsis patients.

Identification and Validation of Best Discriminating Set of Genes

To identify the subset of probes that best discriminates sepsis patients from controls, we trained an ensemble of four ML algorithms (Fig. 1*B*) using discovery cohort and performed voting to identify the probes that appeared in at least two (233 probes), three (64 probes), or all of the four (42 probes) models. All three subsets were subsequently validated using validation cohort using random forest, support vector machine, and logistic regression models whose hyperparameters were tuned on the discovery cohort. The support vector machine

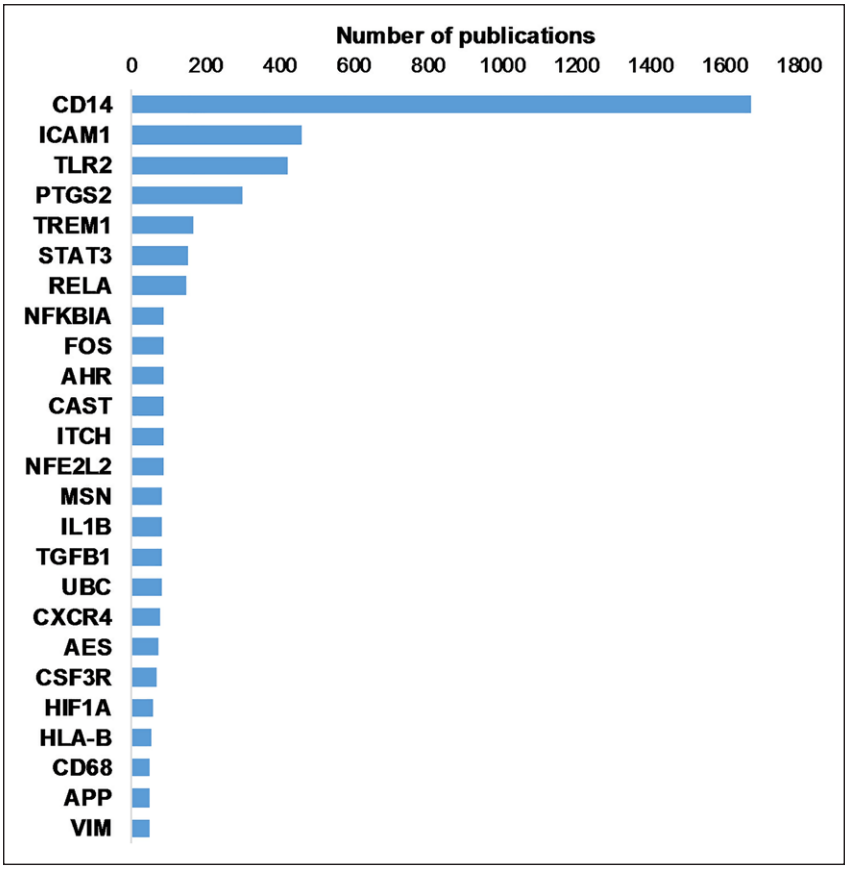

Figure 2. *(Continued)* **C**, Genes with the highest number of publications cross-referenced in PubMed with the term "sepsis." Among 1,048 genes that were described by the differentially expressed probes, 23% were cross-referenced in PubMed. Genes that were cross-referenced with at least 50 PubMed references are shown here. AES = amino enhancer of split protein, AHR = aryl hydrocarbon receptor, APP = amyloid precursor protein, CAST = calpastatin, CD14 = cluster of differentiation 14, CD68 = cluster of differentiation 68, Cdc42 = cell division control protein 42 homolog, CSF3R = colony stimulating factor 3 receptor, CXCR4 = chemokine (C-X-C motif) receptor 4, EIF2 = eukaryotic initiation factor 2, Fcγ = Fc gamma, fMLP = N-Formyl-methionyl-leucyl-phenylalanine, FOS = FBJ murine osteosarcoma viral oncogene homolog B, HIF1A = hypoxia-inducible factor 1 alpha subunit inhibitor, HLA-B = major histocompatibility complex, class I B, ICAM1 = intercellular adhesion molecule 1, IL-1B = interleukin-1 beta, IL-8 = interleukin-8, iNOS = inducible nitric oxide synthase, ITCH = E3 ubiquitin-protein ligase itchy homolog, MSN = moesin, NFAT = nuclear factor of activated T-cells, NFE2L2 = nuclear factor, erythroid 2 like 2, NFKBIA = NF-kappa-B inhibitor alpha, PI3K/AKT = phosphoinositide 3-kinase/protein kinase B, PPARα = peroxisome proliferator-activated receptor alpha, PTGS2 = prostaglandin-endoperoxide synthase 2, RELA = V-rel avian reticuloendotheliosis viral oncogene homolog A, RXRα = retinoid X receptor alpha, STAT3 = signal transducer and activator of transcription 3 (acute-phase response factor), TGFB1 = transforming growth factor beta 1, TLR2 = toll-like receptor 2, TREM1 = triggering receptor expressed on myeloid cells 1, UBC = ubiquitin C, VIM = vimentin.

performed best across all gene subsets with similar performance between 233 and 64 probe sets, whereas reduction to the 42 probes resulted in a decrease in performance (**Table 2**, **Supplementary Table S7** http://links.lww.com/CCX/A281 and **Supplementary Fig. S7** http://links.lww.com/CCX/A271). The functional analysis of these 239 genes (**Supplementary Fig. S8** http://links.lww.com/CCX/A272) revealed up-regulation in pathways related to migration and adhesion of neutrophils and phagocytic cells, IL-8 signaling, and neuroinflammation, whereas peroxisome proliferator-activated receptor alpha/retinoid X receptor alpha pathways remained significantly down-regulated ($p < 0.01$) (**Supplementary Fig. S9** http://links.lww.com/CCX/A273). We investigated the presence of biologically interconnected subsets of genes among the 239 genes (Supplementary Methods http://links.lww.com/CCX/A264) and discovered that several gene products repeatedly co-occurred in the significant pathways, including transcription factor p65 (16 of 17 pathways), IL-1B, protein kinase C delta and prostaglandin-endoperoxide synthase 2 (eight of 17 pathways), transforming growth factor beta 1, IL-8, and toll-like receptor 2 (six of 17 pathways) (**Supplementary Fig. S10** http://links.lww.com/CCX/A274).

DISCUSSION

In a single-center prospective cohort of patients with sepsis, an ensemble of four ML algorithms identified 239 gene expressions unique to systemic immune and kidney-specific processes using whole-genome transcriptomic analysis of cellular RNA isolated from urine samples within 12 hours of sepsis onset. The functional analysis of these genes displays the up-regulation of innate immune response, cellular motility and extravasation, cellular hypoxia, and production of oxidative species. This pattern resembles gene expression signatures observed in studies of circulating immune cells from the blood of septic patients with activation of the innate immune response (40–42), up-regulated cellular motility (43), cellular hypoxia (43), and production of oxidative species (44), and demonstrates the potential use of urinary immune cells as an indicator of systemic processes. The overall pathway activations exhibited by the urinary signature are comparable with functional analysis using blood transcriptomics reported in previous studies (45–48). Immune deconvolution analysis showed the amplification of gene markers for an array of immune cell types, particularly monocytes and neutrophils, underlining the role of innate immune response in early sepsis and confirming the ability to identify immune cell subsets in the urine. Additionally, the deconvolution analysis found kidney-specific transcripts in urine with up-regulation of tubular epithelial cells.

To our knowledge, this study represents the first urine-based gene expression study in sepsis. Although a readily available biofluid rich in cells and nucleic acids, urine has been underutilized for the development of biomarkers in sepsis. Previous studies have shown diagnostic and prognostic potential of urine messenger RNA and micro RNA in identifying acute and chronic graft rejection in kidney transplant recipients (18, 49, 50) and in diagnosis and risk stratification of urologic malignancies (51, 52). Unlike blood, urine can reflect the kidney response in sepsis, which shows damage associated with oxidative stress, hypoxia, and inflammation leading to cell death and epithelial-to-mesenchymal tissue remodeling (53).

Our study has several unique strengths. We applied an ensemble of ML algorithms to find an optimal subset of genes discriminative of sepsis while preserving relevant nonlinear relationships among them. This ensemble tuned to discovery cohort yielded

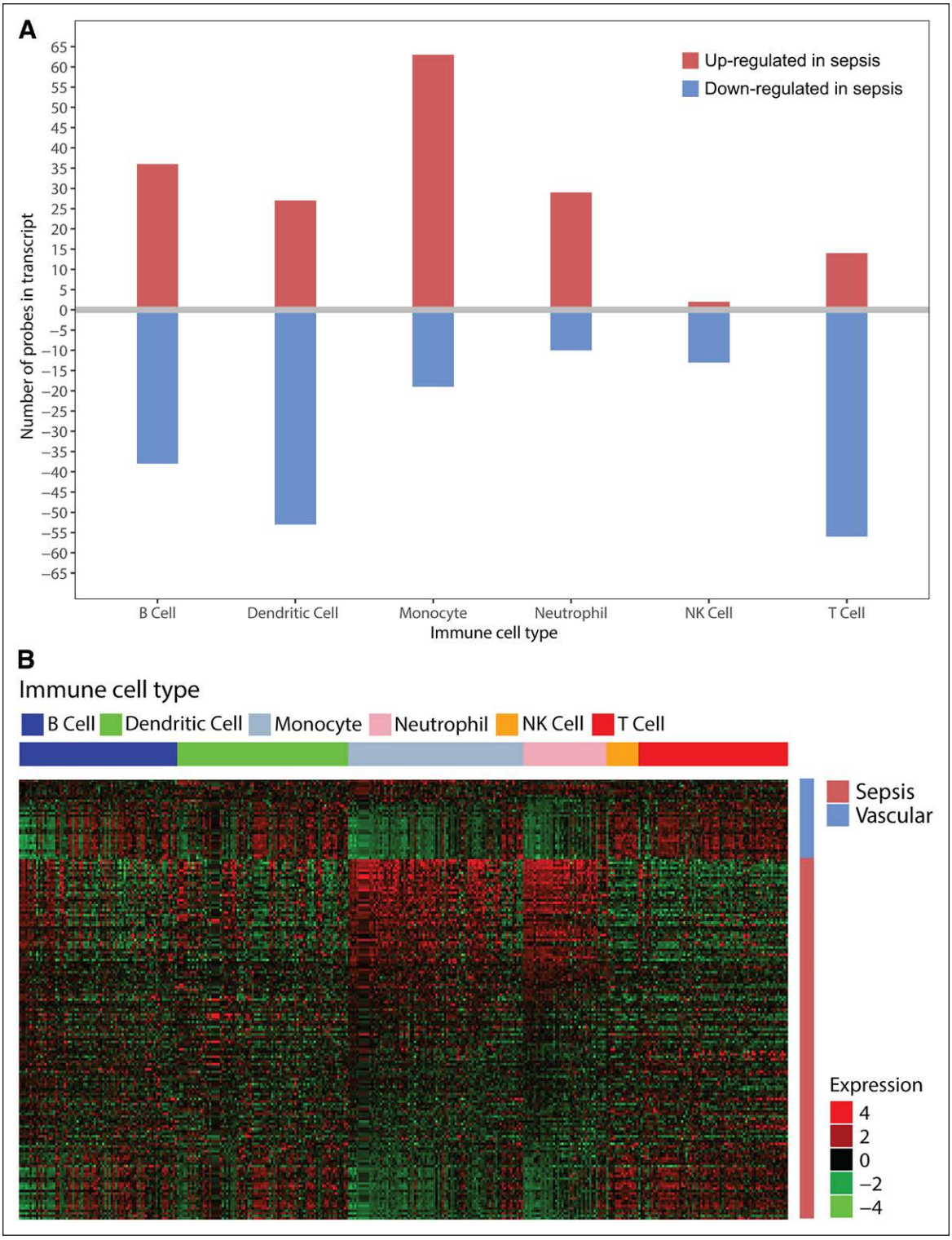

Figure 3. Immune and kidney cell–specific transcript changes in the acute response to sepsis. **A**, Immune cell deconvolution showing the overall percentage differential regulation of immune cell-specific markers (selected from the 1,622 genes from immune response in silico [IRIS] resource; see Materials and Methods) between sepsis and control patients. There was predominant up-regulation of neutrophil and monocyte markers, a mixed-response in B cells and dendritic cells, and down-regulation of NK cells and T cells. **B**, Heatmap of immune cell-specific/enriched markers (selected from the 823 genes from IRIS resource) in the sepsis and control patients. Most of the signature genes in B cells, T cells, and NK cells are underexpressed, and most of the signature genes in neutrophils and monocytes are overexpressed in sepsis compared with controls.

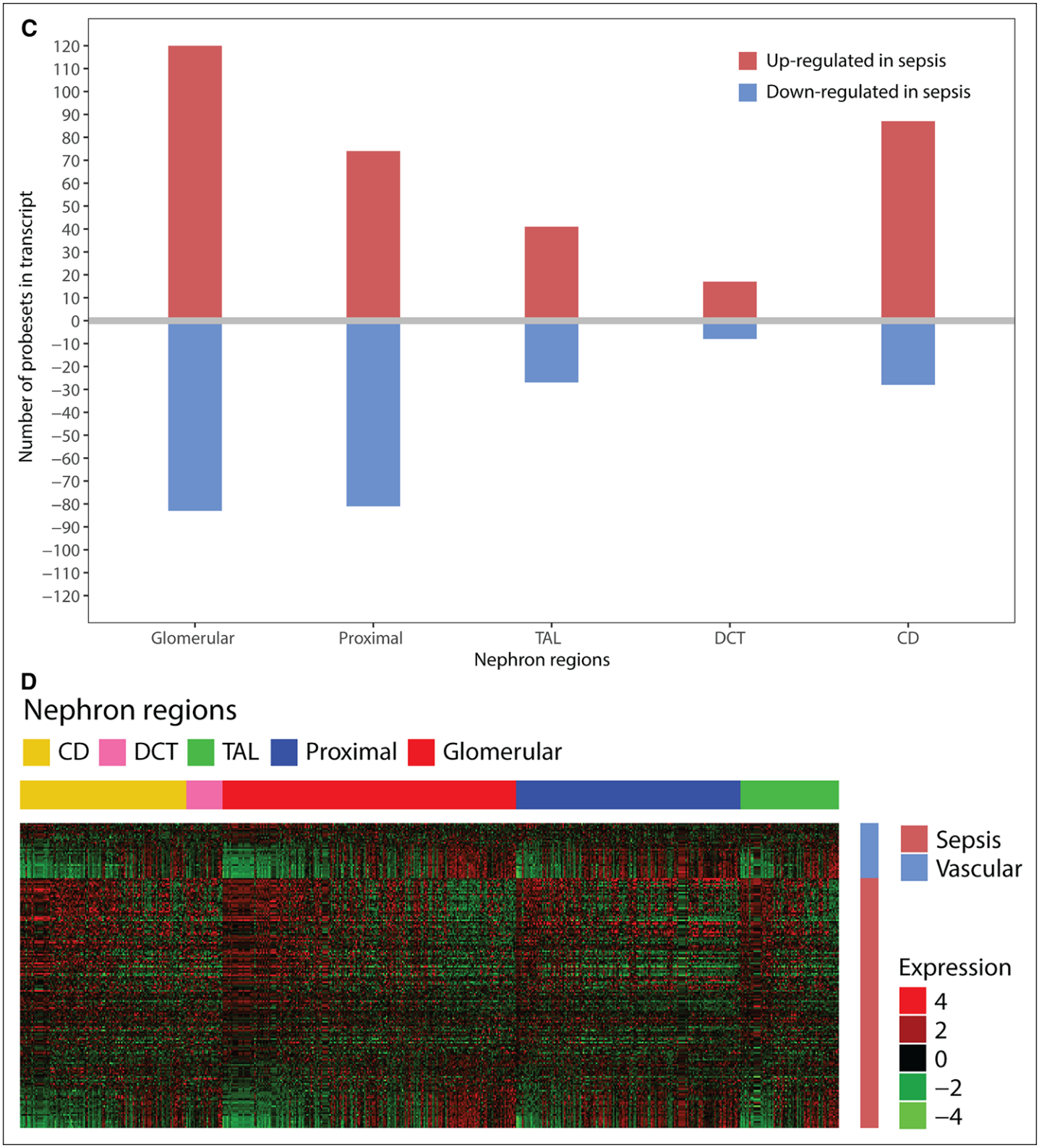

Figure 3. *(Continued)* **C**, Kidney cell deconvolution showing overall percentage differential regulation of kidney cell–specific markers (selected from the 472 genes selected from Chabardès-Garonne et al [28]; see Materials and Methods) between sepsis and control patients. This showed up-regulation of transcripts from the collecting ducts (CDs). **D**, Heatmap of kidney cell–specific/enriched markers in sepsis and vascular patients. No clear pattern was observed here. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. DCT = distal convoluted tubule, NK = natural killer, TAL = thick ascending limb.

TABLE 2. Performance of Selected Probe Sets on External Validation Data Using Support Vector Machine

No. of Probes	Area Under the Curve (95% CI)	Accuracy (95% CI)	F1 Score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	NPV (95% CI)
233	0.86 (0.77–0.93)	0.77 (0.70–0.88)	0.81 (0.71–0.90)	0.84 (0.72–0.95)	0.70 (0.51–0.86)	0.77 (0.67–0.91)	0.77 (0.64–0.92)
64	0.87 (0.80–0.93)	0.77 (0.66–0.85)	0.76 (0.66–0.86)	0.70 (0.55–0.84)	0.85 (0.70–0.95)	0.85 (0.72–0.96)	0.69 (0.53–0.84)
42	0.78 (0.67–0.88)	0.71 (0.61–0.82)	0.73 (0.61–0.83)	0.70 (0.55–0.84)	0.75 (0.61–0.93)	0.79 (0.66–0.93)	0.66 (0.51–0.83)

NPV = negative predictive value.

very strong classification of sepsis from control subjects in the external validation cohort using a modest number of genes, thus confirming our hypothesis that changes in urinary cells may reflect ongoing systemic processes discovered in transcriptomic analyses of blood samples (6, 54).

There are some general clinical differences between our discovery and validation cohorts.

This is expected as the validation cohort is a completely independent sample collected at a different time point compared with the discovery cohort. The robust performance of our ML models on this independent validation cohort proves that our models have succeeded in extracting a general set of features representing the disease state compared with some cohort-specific idiosyncrasies. This is one of the largest sepsis cohorts with complete clinical, immunologic, and molecular characterizations and long-term follow-up of the patients. We have applied rigorous methodology including the use of independent validation cohort to improve generalizability of our results and to overcome the limitation of a modest sample size.

Our study has limitations. The cost to prospectively enroll critically ill patients, obtain samples within 12 hours of sepsis onset, and analyze full-genome data was a determining factor for the limited size of the discovery and validation cohorts. We have taken multiple steps to ensure that class imbalance in the discovery cohort does not hinder the accuracy of our results including: 1) training using scikit-learn packages with class balancing ability wherein the ML model is biased toward the class with lower samples to a degree commensurate with the imbalance and 2) using a threshold that maximizes the Youden index rather than default threshold equal to 0.5 for calculating performance metrics. Although the current results are promising as they show dysregulation of gene expression of key cellular subsets in sepsis, additional comparisons will need to be made with patients with more pronounced systemic inflammatory syndrome. Furthermore, our future work will include evaluating the ability of urinary metabolomics compared with transcriptomics to differentiate sepsis from noninfected controls and to perform a complete comparison of pathways activated in urinary transcriptomics against those upregulated in blood transcriptomics of sepsis.

CONCLUSIONS

The whole-genome transcriptomic analysis of cellular RNA isolated from the urine samples of septic patients reveals changes in gene expressions unique to systemic immune and kidney-specific processes as early as within 12 hours of sepsis onset. Future studies need to confirm whether this approach can complement blood transcriptomic approaches for sepsis diagnosis and prognostication.

ACKNOWLEDGMENTS

We thank Jake Rubin, Matthew Ruppert, Justin Patton, Seth Williams, and Emel Bihorac for their help in urine sample collection and isolation, and George Omalay, Haleh Hashemighouchani, and Sidney Lowhar for their help in editing and submission procedures. We also thank all patients and their families and research coordinators for participating in this study.

Drs. Segal and Bihorac conceived the original idea for the study, and sought and obtained funding. Mr. Bandyopadhyay, Dr. Adhikari, Dr. Ozrazgat-Baslanti, Dr. Rashidi, and Dr. Bihorac had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Mr. Bandyopadhyay, Dr. Adhikari, Ms. Sautina, Ms. Lopez, Dr. Baker, Dr. Ozrazgat-Baslanti, Dr. Rashidi, and Dr. Bihorac participated in some analyses. The article was written by Mr. Bandyopadhyay, Dr. Lysak, Dr. Adhikari, Dr. Ozrazgat-Baslanti, and Dr. Bihorac with input from all coauthors. Drs. Ozrazgat-Baslanti, Rashidi, and Bihorac served as senior authors. Dr. Bihorac is guarantor for this article. All authors participated in the interpretation of data. All authors participated in critically revising the manuscript for important intellectual content and gave final approval of the version to be published.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

Supported, in part, by the National Institutes of Health/National Center for Advancing Translational Sciences Clinical and Translational Sciences Award to the University of Florida UL1 TR000064.

Drs. Adhikari, Ozrazgat-Baslanti, Rashidi, and Bihorac were supported, in part, by R01 GM110240 from the National Institute of General Medical Sciences. Ms. Sautina, Ms. Lopez, Dr. Efron, Dr. Brakenridge, Dr. Moldawer, Dr. Moore, Dr. Baker, Dr. Segal, Dr. Ozrazgat-Baslanti, and Dr. Bihorac were supported, in part, by the Sepsis and Critical Illness Research Center Award P50 GM-111152 from the National Institute of General Medical Sciences. Dr. Mohandas was supported, in part, by a K08 award from the National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute (K08:HL130945). Dr. Rashidi was supported, in part, by the grant NIH/National Institute of Biomedical Imaging and Bioengineering 1R21EB027344 and National Science Foundation CAREER 1750192. Dr. Ozrazgat-Baslanti was supported, in part, by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001427 and, in part, by the grant from Gatorade Trust (127900), University of Florida. The remaining authors have disclosed that they do not have any potential conflicts of interest.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Raw and normalized expression data for discovery and validation cohorts discussed in this article have been deposited in National Center for Bio-

technology Information's (NCBI's) Gene Expression Omnibus (Edgar R, Domrachev M, Lash AE: Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; 30:207–210) and are accessible through Gene Expression Omnibus series accession numbers GSE112100 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112100), GSE112099 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112099), and GSE112098 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112098).

For information regarding this article, E-mail: abihorac@ufl.edu

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810
2. Dellinger RP, Carlet JM, Masur H, et al: Surviving sepsis campaign guidelines for management of severe sepsis and septic shock. *Intensive Care Med* 2004; 30:536–555
3. Prescott HC, Angus DC: Enhancing recovery from sepsis: A review. *JAMA* 2018; 319:62–75
4. Liu Y, Hou JH, Li Q, et al: Biomarkers for diagnosis of sepsis in patients with systemic inflammatory response syndrome: A systematic review and meta-analysis. *Springerplus* 2016; 5:2091
5. Sweeney TE, Khatri P: Benchmarking sepsis gene expression diagnostics using public data. *Crit Care Med* 2017; 45:1–10
6. McHugh L, Seldon TA, Brandon RA, et al: A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: Discovery and validation in independent cohorts. *PLOS Med* 2015; 12:e1001916
7. White LE, Hassoun HT, Bihorac A, et al: Acute kidney injury is surprisingly common and a powerful predictor of mortality in surgical sepsis. *J Trauma Acute Care Surg* 2013; 75:432–438
8. Gómez H, Kellum JA: Sepsis-induced acute kidney injury. *Curr Opin Crit Care* 2016; 22:546–553
9. Decramer S, Gonzalez de Peredo A, Breuil B, et al: Urine in clinical proteomics. *Mol Cell Proteomics* 2008; 7:1850–1862
10. Skoberne A, Konieczny A, Schiffer M: Glomerular epithelial cells in the urine: What has to be done to make them worthwhile? *Am J Physiol Renal Physiol* 2009; 296:F230–F241
11. Szeto CC, Chan RW, Lai KB, et al: Messenger RNA expression of target genes in the urinary sediment of patients with chronic kidney diseases. *Nephrol Dial Transplant* 2005; 20:105–113
12. Achenbach J, Mengel M, Tossidou I, et al: Parietal epithelia cells in the urine as a marker of disease activity in glomerular diseases. *Nephrol Dial Transplant* 2008; 23:3138–3145
13. Oliveira Arcolino F, Tort Piella A, Papadimitriou E, et al: Human urine as a noninvasive source of kidney cells. *Stem Cells Int* 2015; 2015:362562
14. Kopetschke K, Klocke J, Grießbach AS, et al: The cellular signature of urinary immune cells in Lupus nephritis: New insights into potential biomarkers. *Arthritis Res Ther* 2015; 17:94
15. Abdulahad WH, Kallenberg CG, Limburg PC, et al: Urinary CD4+ effector memory T cells reflect renal disease activity in antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum* 2009; 60:2830–2838
16. Kim S, Kim HJ, Ahn HS, et al: Is plasma neutrophil gelatinase-associated lipocalin a predictive biomarker for acute kidney injury in sepsis patients? A systematic review and meta-analysis. *J Crit Care* 2016; 33:213–223
17. Honore PM, Nguyen HB, Gong M, et al; Sapphire and Topaz Investigators: Urinary tissue inhibitor of metalloproteinase-2 and insulin-like growth factor-binding protein 7 for risk stratification of acute kidney injury in patients with sepsis. *Crit Care Med* 2016; 44:1851–1860
18. Suthanthiran M, Schwartz JE, Ding R, et al; Clinical Trials in Organ Transplantation 04 (CTOT-04) Study Investigators: Urinary-cell mRNA profile and acute cellular rejection in kidney allografts. *N Engl J Med* 2013; 369:20–31
19. Ralla B, Stephan C, Meller S, et al: Nucleic acid-based biomarkers in body fluids of patients with urologic malignancies. *Crit Rev Clin Lab Sci* 2014; 51:200–231
20. Ramachandran K, Saikumar J, Bijol V, et al: Human miRNome profiling identifies microRNAs differentially present in the urine after kidney injury. *Clin Chem* 2013; 59:1742–1752
21. Loftus TJ, Mira JC, Ozrazgat-Baslanti T, et al: Sepsis and Critical Illness Research Center investigators: Protocols and standard operating procedures for a prospective cohort study of sepsis in critically ill surgical patients. *BMJ Open* 2017; 7:e015136
22. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med* 1992; 20: 864–874
23. Croft CA, Moore FA, Efron PA, et al: Computer versus paper system for recognition and management of sepsis in surgical intensive care. *J Trauma Acute Care Surg* 2014; 76:311–317; discussion 318–319
24. Gardner AK, Ghita GL, Wang Z, et al: The development of chronic critical illness determines physical function, quality of life, and long-term survival among early survivors of sepsis in surgical ICUs. *Crit Care Med* 2019; 47:566–573
25. Chomczynski P: A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 1993; 15:532
26. Irizarry RA, Hobbs B, Collin F, et al: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4:249–264
27. Abbas AR, Baldwin D, Ma Y, et al: Immune response in silico (IRIS): Immune-specific genes identified from a compendium of microarray expression data. *Genes Immun* 2005; 6:319–331
28. Chabardès-Garonne D, Mejéan A, Aude JC, et al: A panoramic view of gene expression in the human kidney. *Proc Natl Acad Sci U S A* 2003; 100:13710–13715
29. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; 3:Article3
30. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B (Methodological)* 1995; 57:289–300
31. Gu Z, Eils R, Schlesner M: Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; 32:2847–2849
32. Breiman L: Random forests. *Mach Learn* 2001; 45:5–32
33. Guyon I, Weston J, Barnhill S, et al: Gene selection for cancer classification using support vector machines. *Mach Learn* 2002; 46:389–422
34. Mccullagh P: Generalized linear-models. *Euro J Oper Res* 1984; 16:285–292
35. Kursa MB, Rudnicki WR: Feature selection with the Boruta package. *J Stat Softw* 2010; 36:1–13
36. Tibshirani R: Regression shrinkage and selection via the Lasso. *J R Stat Soc B-Methodol* 1996; 58:267–288
37. Stańczyk U, Jain LC: Feature Selection for Data and Pattern Recognition. Heidelberg, Springer, 2015
38. Winter DJ: rentrez: An R package for the NCBI eUtils API. *R J* 2017; 9:520–526
39. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011; 12:2825–2830
40. Zhang Q, Raoof M, Chen Y, et al: Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature* 2010; 464:104–107
41. Zhang J, Cheng Y, Duan M, et al: Unveiling differentially expressed genes upon regulation of transcription factors in sepsis. *3 Biotech* 2017; 7:46
42. Bauer M, Giamarellos-Bourboulis EJ, Kortgen A, et al: A transcriptomic biomarker to quantify systemic inflammation in sepsis—A prospective multicenter phase II diagnostic study. *Ebiomedicine* 2016; 6:114–125
43. Ma J, Chen C, Barth AS, et al: Lysosome and cytoskeleton pathways are robustly enriched in the blood of septic patients: A meta-analysis of transcriptomic data. *Mediators Inflamm* 2015; 2015:984825

44. Ge QM, Huang CM, Zhu XY, et al: Differentially expressed miRNAs in sepsis-induced acute kidney injury target oxidative stress and mitochondrial dysfunction pathways. *PLoS One* 2017; 12:e0173292
45. Sweeney TE, Azad TD, Donato M, et al: Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit Care Med* 2018; 46:915–925
46. Sweeney TE, Perumal TM, Henao R, et al: A community approach to mortality prediction in sepsis via gene expression analysis. *Nat Commun* 2018; 9:694
47. Davenport EE, Burnham KL, Radhakrishnan J, et al: Genomic landscape of the individual host response and outcomes in sepsis: A prospective cohort study. *Lancet Respir Med* 2016; 4:259–271
48. Burnham KL, Davenport EE, Radhakrishnan J, et al: Shared and distinct aspects of the sepsis transcriptomic response to fecal peritonitis and pneumonia. *Am J Respir Crit Care Med* 2017; 196:328–339
49. Maluf DG, Dumur CI, Suh JL, et al: The urine microRNA profile may help monitor post-transplant renal graft function. *Kidney Int* 2014; 85:439–449
50. Matignon M, Ding R, Dadhania DM, et al: Urinary cell mRNA profiles and differential diagnosis of acute kidney graft dysfunction. *J Am Soc Nephrol* 2014; 25:1586–1597
51. Mengual L, Lozano JJ, Ingelmo-Torres M, et al: Using gene expression from urine sediment to diagnose prostate cancer: Development of a new multiplex mRNA urine test and validation of current biomarkers. *BMC Cancer* 2016; 16:76
52. Tölle A, Jung M, Rabenhorst S, et al: Identification of microRNAs in blood and urine as tumour markers for the detection of urinary bladder cancer. *Oncol Rep* 2013; 30:1949–1956
53. Hultström M, Becirovic-Agic M, Jönsson S: Comparison of acute kidney injury of different etiology reveals in-common mechanisms of tissue damage. *Physiol Genomics* 2018; 50:127–141
54. Sweeney TE, Shidham A, Wong HR, et al: A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med* 2015; 7:287ra71