SUPPLEMENTAL DIGITAL CONTENT

TABLE OF CONTENTS

INTRODUCTION

Prior Weill Cornell Single Center Investigations. The most frequent type of acute rejection is acute T-cell-mediated rejection (TCMR), and the diagnosis of TCMR is established by the presence of tubulitis and interstitial infiltration in the kidney allograft biopsy. Noninvasive diagnosis of TCMR is preferable. TCMR in human kidney allografts is exemplified by intragraft presence of cytotoxic T cells.[1] We therefore investigated whether urinary cell levels of mRNA for cytotoxic attack molecules granzyme B and perforin are diagnostic of TCMR in human kidney allografts. We developed gene specific competitive quantitative PCR assays and quantified urinary cell levels of granzyme B and perforin. In our single center study of 85 kidney allograft recipients, we found that the levels of mRNA for granzyme B and perforin were higher in urine samples matched to TCMR biopsies than in urine samples matched to biopsies without rejection changes and urinary cell level of perforin mRNA predicted TCMR with 83% sensitivity and 83% specificity, while granzyme B mRNA level predicted TCMR with 79% sensitivity and 77% specificity.[2] Furthermore, urinary cell levels of perforin mRNA and granzyme B mRNA increased in sequential urine samples and identified those in whom TCMR developed.[2] In thematically linked urinary cell mRNA profiling studies at Weill Cornell, we found that: (i) urinary cell level of mRNA for serine proteinase inhibitor-9 (PI9), an endogenous antagonist of granzyme B, predicts TCMR with 76% sensitivity and 79% specificity;[3] (ii) mRNA for CD103 predicts TCMR with 59% sensitivity and 75% specificity;[4] (iii) mRNA for IP10 predicts TCMR with 100% sensitivity and 75% specificity,[5] and (iv) mRNA for CXCR3 predicts TCMR with 63% sensitivity and 83% specificity.[5]

The natural history of TCMR is heterogeneous, with some acute rejection episodes reversible with anti-rejection treatment and others nonreversible and leading to graft failure.[6,7] Development of noninvasive biomarkers predictive of TCMR outcome has the potential to lead to the development of personalized anti-rejection therapy.

We examined whether urinary cell mRNAs predict TCMR reversal with the use of an independent cohort of 83 kidney allograft recipients. Because FOXP3 mRNA encodes a specification factor essential for the development and maintenance of regulatory T lymphocytes (Tregs) critical for quenching immunity, including anti-allograft immunity,[8-13] we determined whether urinary cell level of FOXP3 mRNA is able to distinguish patients with a reversible episode of TCMR from those with a nonreversible episode of TCMR. Urinary cell mRNAs were measured in our study using preamplification enhanced real time quantitative PCR (customized PCR) assays developed in our laboratory.[14] In addition to replicating our earlier finding that urinary cell level of perforin mRNA is diagnostic of TCMR, we found that urinary cell levels of FOXP3 mRNA, CD3E mRNA, and CD25 mRNA are also higher in urine samples matched to TCMR biopsies than in urine samples matched to biopsies without rejection. Moreover, urinary cell FOXP3 mRNA level, but not the levels of mRNA for perforin, CD3E, or CD25, predicted reversal of an episode of TCMR.[14]

Formulation of Hypotheses. Data from our single center studies[2-5,14] led to the formulation of the following hypotheses: Urinary cell mRNA profiles are diagnostic of acute rejection; urinary cell mRNA profiles, ascertained in sequentially collected urine specimens, predict future development of TCMR; and urinary cell mRNA profiles predict TCMR reversal.

Clinical Trials in Organ Transplantation- 04 (CTOT-04) Study. In the multicenter CTOT-04 study, we replicated our single center findings that urinary cell levels of granzyme B, perforin, CD3E, IP10, CD103, and CXCR3 are diagnostic of TCMR.[15] We also found that the abundance of 18S rRNA in urinary cells are higher in urine specimens matched to TCMR biopsies than in urine specimens matched to biopsies without rejection, and that normalization of mRNA levels with 18S rRNA resulted in urinary cell levels of CD103 and CXCR3 being not associated with TCMR biopsy diagnosis. Very importantly, we developed and validated a 3-gene urinary cell signature of 18S rRNA normalized CD3E and IP10 mRNAs and 18S rRNA diagnostic of TCMR; the three gene signature was diagnostic of TCMR with 78% specificity (95% CI: 71 to 84) and 79% sensitivity (95% CI: 67 to 91) and the area under the receiver-operating-characteristic curve (AUC) was 0.85 (95% CI: 0.78 to 0.91; P<0.001).[15] The scores of the three-gene diagnostic signature in longitudinally collected urine specimens showed a marked increase during the 20-day period leading up to the TCMR biopsy diagnosis whereas the signature score remained relatively flat and well below the diagnostic threshold during the 270 days preceding a biopsy without histological features of rejection.

In the current investigation, we examined whether the CTOT-04 three-gene diagnostic signature predicts reversal of an episode of TCMR. In view of our single center finding that urinary cell FOXP3 mRNA level predicts TCMR reversal,[14] we determined whether urinary cell FOXP3 mRNA level predicts TCMR reversal in the CTOT-04 cohort. We considered it important to replicate our single center finding using and independent cohort of kidney graft recipients in light of the existing crisis in replicating scientific findings.[16,17]

METHODS

CTOT-04 Study. Of the 485 patients enrolled in the parent CTOT-04 study urinary cell cDNA from 480 patients were available for analysis in the current investigation. Among the 480 patients, 218 patients underwent at least one renal allograft biopsy and the remaining 262 patients did not undergo a biopsy. The characteristics of the recipients stratified by whether they underwent a kidney allograft biopsy or not in the CTOT-04 study have already been reported.[15]

The inclusion criteria for the CTOT-04 study were: (i) male and female recipients of all races, 0-80 years of age; (ii) patients undergoing primary or re-do living or deceased donor kidney transplantation; and (iii) ability to provide informed consent. The exclusion criteria were: (i) need for combined organ transplantation; (ii) recipients of previous non-renal solid organ and/or an islet cell transplant; (iii) infection with HCV or HIV; and (iv) inability or unwillingness to provide informed consent.

There was no selection or randomization process and enrollment was on a consecutive basis provided: (i) the patients were eligible based on inclusion and exclusion criteria; and (ii) the patients were willing to participate in the study.

An observational study design was used with each of the 5 participating transplant sites being able to use site-specific immunosuppression and infection protocols. The rationale for our study design included the consideration that the urine mRNA profiling results should be

generalizable to the kidney transplant population and not be restricted by the clinical therapeutic regimens used to treat kidney allograft recipients.

Quantification of mRNAs with the Use of Preamplification Enhanced Real Time Quantitative (customized) PCR Assays. Our protocols for the collection of urine samples from the kidney allograft recipients, preparation of urine cell pellets by centrifugation, and isolation of total RNA from the urinary cell pellet have been reported.[15] We reverse transcribed total RNA to cDNA using the TaqMan reverse transcription kit and reverse transcription was performed on the same day the total RNA was isolated. We designed gene specific oligonucleotide primers and TaqMan fluorogenic probes (hydrolysis probes) for the measurement of levels of mRNAs. The sequence and location of the gene specific oligonucleotide primer pairs and TaqMan probes used in this investigation to quantify transcript numbers are provided in Supplementary Table S1. PCR analysis was performed using our two-step process: a preamplification step followed by measurement of levels of mRNAs with an ABI Prism 7500/7900HT Fast detection system. A PCR-generated 73-bp mouse Bak amplicon was used to develop the standard curve and absolute levels of mRNAs were quantified, as previously described.[14] Additional details on the CTOT-04 study design have been reported.[15]

Recording of Kidney Allograft Biopsy Findings. Each on-site pathologist classified the biopsies using the updated Banff 97 diagnostic categories.[18] The biopsy results were recorded prospectively using the NIAID-Statistical and Clinical Coordinating Center Biopsy Form (Table S2).[15]

Acute T-Cell-Mediated Rejection (TCMR) Biopsy Group. Forty-three urine specimens matched to 43 acute T-cell-mediated rejection biopsies from 34 patients are included in this study. Nineteen of

the 43 biopsies were graded as mild acute T-cell-mediated rejection grade IA, 10 as mild acute T-cell-mediated rejection grade IB, 11 as moderate acute T-cell-mediated rejection grade IIA, 2 as moderate acute T-cell-mediated rejection grade IIB, and 1 as severe acute T-cell-mediated rejection grade III. Urinary cell mRNA levels in the 43 urine specimens matched to the 43 acute T-cell-mediated rejection biopsies were used to generate the violin plots shown in Figure 2. The median, 25^{th}, and 75^{th} percentiles of the $\log_{10}$-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA are reported in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

Reversible or Nonreversible Episodes of TCMR. Among the 43 biopsies with biopsy-matched urine specimens that demonstrated Banff acute T-cell-mediated rejection grade IA or higher, 39 TCMR biopsies from 33 patients were classifiable as either an TCMR episode that successfully responded to anti-rejection therapy (Reversible TCMR) or an TCMR episode that failed to respond to anti-rejection therapy (Nonreversible TCMR). An episode of TCMR was classified as reversible if the serum creatinine level returned to within 15% of the pre-rejection level within 4 weeks of initiation of anti-rejection therapy. The criterion used in the current study to classify an episode as reversible or nonreversible is identical to the criterion used to classify an episode of TCMR as reversible or nonreversible in our single center study that investigated the predictive value of urinary cell level of FOXP3 mRNA.[14] Four of 43 episodes were not classifiable as reversible or nonreversible because of missing post-treatment serum creatinine values (n=2), proximity to BKVN diagnosis or proximity to another episode of TCMR (n=1). The demographic and clinical characteristics of the kidney transplant recipients classified into the

group of subjects who contributed urine specimens matched to reversible TCMR episodes and subjects who contributed urine specimens matched to nonreversible TCMR episodes are summarized in Table 2. Kidney allograft biopsy related features are summarized in Table 3. The median, 25^{th}, and 75^{th} percentiles of the log_{10}-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA are reported in Table 4.

No Rejection Biopsy Group. Urinary cell mRNA levels in 162 urine specimens matched to 162 No Rejection biopsies from 126 patients are included in this study and were used to generate the violin plots shown in Figure 2. The No Rejection biopsy group did not include any biopsies that showed borderline changes or acute or chronic rejection changes. The median, 25^{th}, and 75^{th} percentiles of the log_{10}-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA are reported in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

Acute Antibody-Mediated Rejection (AMR) Biopsy Group. The AMR biopsy group consisted of 10 biopsies from 9 patients. Among the 10 AMR biopsies, 8 were for-cause biopsies and 2 were surveillance biopsies; among the 10 biopsies, 4 were graded as grade I-ATN and 6 were graded as grade II-capillary. Urinary cell mRNA levels in these 10 urine specimens matched to 10 AMR biopsies were used to derive the median, 25^{th}, and 75^{th} percentiles of the log_{10}-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA shown in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

Borderline Changes Biopsy Group. The group with Borderline Changes consisted of 19 biopsies from 17 patients. Among the 19 Borderline Changes biopsies, 14 were for-cause biopsies and 5 were surveillance biopsies Urinary cell mRNA levels in the 19 urine specimens matched to 19 Borderline biopsies from 17 patients were used to derive the median, 25^{th}, and 75^{th} percentiles of the $\log_{10}$-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

Other Findings Biopsy Group. The Other Findings biopsy group consisted of 9 urine specimens matched to 9 biopsies from 8 patients (Figure 1). All 9 biopsies were for-cause biopsies and included polyoma/BK virus nephropathy (n=7) or chronic allograft nephropathy (n=2). Polyoma/BK virus nephropathy was the most frequent biopsy diagnosis in the Other Findings biopsy group and 7 of the 9 biopsy matched urine samples were from 6 patients with BK virus nephropathy. Urinary cell mRNA levels in the 7 urine specimens from 6 patients were used to derive the median, 25^{th}, and 75^{th} percentiles of the $\log_{10}$-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

No Biopsy Group. There were 262 kidney transplant recipients who never had a recorded biopsy, with 199 of them meeting the following criteria to be classified as the Stable (no biopsy) group. The criteria are: (i) average of serum creatinine assessed at 6, 9, and 12 months post-transplantation less than or equal to 2.0 mg/dl (in 3 patients, 6-, 9-, and 12-month creatinine levels

were not available but their 24-month creatinine level was less than or equal to 2.0 mg/dl); (ii) no graft loss or death during the first 12 months following transplantation; (iii) no treatment for acute rejection; (iv) no evidence for cytomegalovirus or BK virus infection; and (v) did not require a biopsy. The 1524 urine specimens provided by these 199 stable patients were used to generate the violin plots shown in Figure 2. Urine samples from patients in the Stable (no biopsy) group were collected longitudinally on days 3, 7, 15, and 30 and in months 2, 3, 4, 5, 6, 9, and 12 post-transplantation. The median, 25^{th}, and 75^{th} percentiles of the $\log_{10}$-transformed ratios of mRNA copies per microgram of total RNA to 18S rRNA copies ($x10^{-6}$) per microgram of total RNA are reported in Table S3. The median, 25^{th}, and 75^{th} percentiles of absolute mRNA copies per total RNA are reported in Table S4.

Of the remaining 63 patients who did not have a biopsy, 47 had no serum creatinine values beyond 5 months post-transplantation, 4 had serum creatinine values at 6-, 9-, and/or 12-months that averaged >2.0 mg/dl, 9 were treated for CMV, BKV infection or both, 1 lost his/her graft within the first 12 months, and 2 patients died within the first 12 months. Urinary cell mRNA data from 237 urine specimens collected from these 63 patients were not included in the data analysis.

Statistics. Each mRNA measure was normalized by dividing its copy number per microgram of total RNA by 18S rRNA copy number ($\times 10^{-6}$) per microgram of total RNA. The distribution of each 18S-normalized mRNA measure exhibited considerable positive skewness, which was substantially reduced by use of the $\log_{10}$-transformation. Violin plots and the non-parametric Kruskal-Wallis and Mann-Whitney tests were used to examine group differences in each 18S-normalized mRNA measure across diagnostic categories (Figure 2 and

Table S3) and the Mann-Whitney test was used to examine group differences between the urine samples matched to reversible TCMR vs. nonreversible TCMR (Figure 3 and Table 4). Data analyses performed using absolute copy numbers of mRNAs are shown in Table S4.

Repeated Measures. In this study, several patients contributed more than one biopsy and more than one urine sample. We performed mixed effects analysis with random intercepts to account for the repeated measures, the estimated variance of the intercepts was zero, indicating that two observations from the same individual are no more similar to each other than observations from two different individuals. Similarly, we performed mixed effects analyses with random intercepts to account for repeated measures among patients with No Rejection biopsies and Stable patients and similarly found that the estimated variances of the intercepts were zero. Given these estimated between-patient variances of zero, the results of these mixed effects models are identical to the reported logistic regression analyses that do attempt to adjust for clustering due to repeated measures. Therefore, we have chosen to report the simpler analysis that does not attempt to adjust for clustering.

Fitting the Model. Logistic regression analyses predicting TCMR reversible status (Reversible TCMR episode vs. Nonreversible TCMR episode) from combinations of the 4 $\log_{10}$-transformed, 18S-normalized mRNA measures (FOXP3, CD3E, CD25, and Perforin mRNAs) obtained from the matched urine samples and clinical parameters (e.g., time from transplantation to biopsy, type of graft, donor age, recipient age) were performed using a stepwise model selection approach with a *P*-value of 0.10 for entry and exit a predictor into/from the model. Briefly, reversible status was regressed upon each predictor, one at a

time separately, with the most statistically significant predictor (based on the Wald test of significance) being selected for inclusion in the model, provided that its *P*-value was smaller than the 0.10 criterion specified for entry into the model). In the second step, reversible status was regressed, separately, upon each combination of the previously selected parameter plus one of the not-yet-selected parameters. As before, the most significant second parameter (based on the Wald test) was selected for inclusion in the model, provided that its *P*-value was smaller than the *P*-value required for entry into the model. At this point, the statistical significance of each predictor that had been selected for inclusion in the model was evaluated, and any predictor with a *P*-value larger than the *P*-value specified for exit from the model (also 0.10) was removed from the model. This iterative process of 1) evaluating all not-yet-selected parameters for the one that was most significant, controlling for those predictors already included in the model, for possible addition to the model, and 2) evaluating, after any new variable had been added, all included predictors for any that needed to be removed from the model was repeated until a parsimonious combination of predictors (each with P-value 0.10) had been selected from the set of all available parameters. We utilized log-likelihood ratio tests and analyses of area under the receiver-operating-characteristic (ROC) curve (AUC) as well as comparisons of AIC to evaluate the model fit. The regression estimates for the final model were used to define a signature whose ability to accurately discriminate reversible TCMR episodes from nonreversible TCMR episodes was then evaluated. The ability of this predictive signature to discriminate (i.e., correctly classify) the 24 reversible TCMR episodes from the 15 nonreversible TCMR episodes was evaluated by the area under the ROC curve of the fitted model, as well as the sensitivity and specificity. An AUC of 1.0 indicates perfect concordance; i.e., every reversible TCMR episode has a higher score on the

predictive signature than every nonreversible TCMR episode. An AUC of 0.50 would indicate that the ability of the predictive signature to differentiate a reversible TCMR episode from a nonreversible TCMR episode was no better than chance. Sensitivity and specificity were determined for the cutoff point on the ROC that maximized Youden's index (sensitivity + specificity – 1).[19]

Prospective Trajectories. To determine prospective longitudinal trajectories for FOXP3 mRNA levels and the CTOT-04 three-gene TCMR diagnostic signature, we included all QC-passed urine samples collected during the first 30 days post-biopsy. For both the reversible and nonreversible TCMR biopsy group, we estimated the pooled within-person average trajectory by fitting a loess (locally estimated scatterplot smoothing) model predicting either FOXP3 mRNA levels or the CTOT-04 three-gene TCMR diagnostic signature from the days after biopsy, with the patient as a covariate.

Kaplan-Meier Survival Analysis. Data for the endpoint – kidney allograft failure – on the 480 study subjects included in this study were obtained from the United Network for Organ Sharing (UNOS). For some comparisons, time to event was defined as the time from transplantation to graft failure, if graft failure occurred, or time from transplantation until the date that the patient was last known to have a functioning graft (i.e., date of censoring). For other analyses, restricted to those with a biopsy-diagnosed TCMR, time to event was defined as the time from the last TCMR biopsy to graft failure, if graft failure occurred, or the date of censoring. Patients were censored if they were lost to administrative follow-up or at the time of death. Time-varying covariates were not considered because the primary question being addressed was whether

urinary cell FOXP3 mRNA level, stratified at the -1.33 threshold (Youden's index for predicting TCMR reversal), predicts kidney allograft survival. Kaplan-Meier survival analysis was also used to investigate whether the following variables predict long-term graft outcome: an episode of TCMR, reversibility of an episode of TCMR, urinary cell mRNA level of CD25, CD3E, and perforin, all stratified by median value, serum creatinine level measured at the time of a first TCMR biopsy (also stratified at its median), and treatment of an episode of TCMR with anti-thymocyte globulin.

All analyses were performed using the SAS software, version 9.4 (SAS Institute), except for the repeated 10-fold cross-validation, performed using the *caret* R package (version 6.0-80; https://cran.r-project.org/web/packages/caret/index.html), and the creation of ROC curves using the *pROC* R package.

RESULTS

Urinary Cell mRNA Levels: Clinically Indicated Biopsies versus Surveillance Biopsies.

FOXP3 mRNA levels (P=0.425), CD25 mRNA levels (P=0.223), and CD3 mRNA levels (P=0.330) levels were not different between urines matched to clinically indicated No Rejection biopsies and surveillance No Rejection biopsies. Perforin mRNA levels (P=0.0124) were significantly higher in urines matched to clinically indicated No Rejection biopsies compared to levels in urines matched to surveillance No Rejection biopsies.

ROC Curve Analysis of Urinary Cell mRNA Levels. ROC curve analysis of urinary cell FOXP3 mRNA level yielded an area under the curve (AUC) of 0.600 (95% CI, 0.491 to 0.709, *P*=0.038, Supplementary Figure S1, Panel A) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.624 (95% CI, 0.518-0.731, *P*=0.001, Figure S1, Panel B) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.527 (95% CI, 0.479-0.575, *P*=0.245, Figure S1, Panel C) when the No Rejection biopsy group was compared with the Stable Graft Function group.

ROC curve analysis of urinary cell CD3E mRNA levels yielded an area under the curve (AUC) of 0.747 (95% CI, 0.665 to 0.828, *P*< 0.001, Figure S1, Panel D) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.719 (95% CI, 0.639-0.798, *P*<0.001, Figure S1, Panel E) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.545 (95% CI, 0.496-0.593, *P*=0.073, Figure S1, Panel F) when the No Rejection biopsy group was compared with the Stable Graft Function group.

ROC curve analysis of urinary cell CD25 mRNA levels yielded an area under the curve (AUC) of 0.596 (95% CI, 0.502 to 0.695, *P*=0.044, Figure S1, Panel G) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.623 (95% CI, 0.538-0.708, *P*=0.005, Figure S1, Panel H) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.528 (95% CI, 0.482-0.574, *P*=0.248, Figure S1, Panel I) when the No Rejection biopsy group was compared with the Stable Graft Function group.

ROC curve analysis of urinary cell perforin mRNA levels yielded an area under the curve (AUC) of 0.745 (95% CI, 0.684 to 0.826, *P*<0.001, Figure S1, Panel J) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.769 (95% CI, 0.699-0.839, *P*<0.001, Figure S1, Panel K) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.521 (95% CI, 0.474-0.568, *P*=0.286, Figure S1, Panel L) when the No Rejection biopsy group was compared with the Stable Graft Function group.

Urinary Tract Infection. Among the 39 TCMR biopsies classifiable as reversible or nonreversible, 5 occurred during a recorded infection. Among the 5 recorded infections, two occurred during nonreversible TCMR episodes (13.3% of nonreversible TCMR episodes), while the remaining three occurred during reversible TCMR episodes (12.5% of reversible TCMR episodes). Median log-transformed, 18S rRNA normalized FOXP3 level in urines matched to TCMR biopsies occurring during infections was -1.083, compared to a median level of -1.092 in urines matched to TCMR biopsies occurring in the absence of a recorded infection (*P*=0.356). Furthermore, FOXP3 mRNA remained significantly associated with TCMR reversal after adjustment for infection status.

Urinary Cell mRNA Levels and BKV Nephropathy. The median (IQR: interquartile range) log-transformed, 18S rRNA normalized FOXP3 level in urine matched to 7 BKVN biopsies was -1.178 (-1.690, -1.068); median (IQR) log-transformed, 18S rRNA normalized CD25, CD3E and perforin levels in urine matched to the same 7 BKVN biopsies were -0.119 (-0.626, 0.103), 0.364 (-0.297, 0.462), and 0.177 (-0.380, 0.441), respectively (Tables S3). Median FOXP3 levels were not significantly different between urines matched to biopsies showing TCMR and those

matched to biopsies showing BKV nephropathy (*P*=0.8268). Similarly, median CD3, CD25, and perforin levels were not significantly different between urines matched to biopsies showing TCMR and those matched to biopsies showing BKV nephropathy (Tables S3). The data from this small series of BKVN biopsies suggest that unlike bacterial urinary tract infection, BKVN is associated with a urinary cell mRNA signature that is similar to the TCMR signature. Indeed, we have reported that the urinary cell level of granzyme B in kidney allograft recipients with BKVN and subsequent graft dysfunction is similar to the level in urine matched to TCMR biopsies, and urinary cell granzyme B mRNA level in BKVN patients with stable graft function is significantly lower than the level in urine matched to TCMR biopsies. We would therefore recommend screening for BKV VP-1 mRNA in the cDNA used to measure mRNAs that are diagnostic and/or prognostic of TCMR.

A Composite Model for Predicting TCMR Reversal. Donor/Recipient age: Donor age (*P*=0.409) did not predict TCMR reversibility but recipient age did *(P*=0.033) by univariable analysis (Table 2). However, FOXP3 mRNA levels predicted TCMR reversibility after controlling for recipient's age in a multivariable logistic regression analysis (*P*=0.0090, Table 5). Serum creatinine measured at the time of biopsy: The median serum creatinine level measured at the time of allograft biopsy was 2.3 mg/dL in those with reversible TCMR and 2.5 mg/dL in those with nonreversible TCMR, and did not predict TCMR reversal in univariable analyses (*P*=0.315, Table 3). Furthermore, receiver-operating characteristic (ROC) curve analysis of serum creatinine measured at the time of biopsy showed that the 95% confidence interval for the area under the curve was 0.428 to 0.795 and included the 0.5 threshold for prediction by chance (*P*=0.308, Figure 3F). Time after Transplantation: The time from transplantation to TCMR

biopsy was significantly different by univariable analysis between those with reversible TCMR biopsy versus those with nonreversible biopsy with the median time from transplantation to biopsy being 117 days for reversible biopsies versus 269 days for the nonreversible TCMR biopsies (*P*=0.015, Table 3). However, FOXP3 mRNA levels independently predicted TCMR reversibility after controlling for time from transplantation to biopsy in a multivariable logistic regression analysis (*P*=0.0130, Table 5). Inclusion of protocol biopsies and clinically indicated biopsies making serum creatinine a predictor: Among the 43 TCMR biopsies included in this study, 39 biopsies were classifiable as reversible or nonreversible and 4 were not classifiable, as reported in the Results section. Among those 39 biopsies that were classified, 34 were for-cause biopsies, compared to only 5 performed as surveillance biopsies. An analysis restricted to the 34 for-cause biopsies showed that FOXP3 mRNA levels were significantly associated with reversibility (AUC: 0.775; 95% CI, 0.620-0.931; *P*=0.008) whereas serum creatinine was not a significant predictor of reversibility (AUC: 0.611; 95% CI, 0.415-0.806; *P*=0.360). The very small number of surveillance biopsies showing TCMR precluded an examination of the associations of FOXP3 mRNA levels and serum creatinine at the time of surveillance biopsy with TCMR reversibility. Nevertheless, FOXP3 mRNA levels were significantly associated with reversibility in for-cause TCMR biopsies alone (*P*=0.008) as well as in for-cause and surveillance TCMR biopsies combined (*P*=0.008), whereas serum creatinine at the time of biopsy was not significantly associated with reversibility in either the subset of for-cause TCMR biopsies alone (*P*=0.360) or the combination of both for-cause and surveillance TCMR biopsies (*P*=0.308, Figure 3F).

Additional factors, such as the use of thymoglobulin as anti-rejection treatment for TCMR (AUC=0.479; 95% CI, 0.321-0.637; *P*=0.792) and TCMR Banff grade (*P*=0.911, Table 3), were not associated with reversibility. Recipient gender, ethnicity, race, induction therapy, BMI and donor age, gender, ethnicity, race, and cause of death were also not associated with reversibility (Table 2), but the source of kidney allograft (living donor vs. deceased donor) was marginally associated (*P*=0.0550, Table 2) with reversibility. FOXP3 mRNA levels independently predicted TCMR reversibility after controlling for type of donor graft in a multivariable logistic regression analysis (*P*=0.0170, Table 5, previously table 3).

Equation for FOXP3 + clinical parameters:

$$1.419 + 1.441 * \log_{10}(\text{FOXP3/18S}) - 0.004 * Days\_from\_transplant\_to\_mRNA + 0.087 * Recipient\_age - 1.330 * Type\_of\_donor\_graft$$

Standard errors: intercept (2.474), FOXP3 (1.441), days from transplant to mRNA (0.003), recipient age (0.054), and donor type (0.935)

Equation for clinical parameters:

$$0.587 - 0.004 * Days\_from\_transplant\_to\_mRNA + 0.066 * Recipient\_age - 1.426 * Type\_of\_donor\_graft$$

Standard errors: intercept (2.203), days from transplant to mRNA (0.002), recipient age (0.043), and donor type (0.804)

DISCUSSION

Rationale for the Focus on FOXP3. FOXP3 is a specification factor for Tregs and is expressed primarily in CD4+CD25+ T cells.[8-10,20] Gain-of-function, over-expression studies, and analysis of Scurfy (*sf*) mice deficient in FOXP3, all show FOXP3 to be essential for the development and maintenance of Tregs.[21-24] Normal individuals possess FOXP3+ CD25+CD4+ T cells (about 5-10% of all CD4+ cells) with potent immunoregulatory properties. Genetic evidence and experimental models of autoimmunity demonstrate that deficiency or dysfunction of natural Tregs results in loss of self-tolerance and autoimmunity. The rare human X-linked neonatal diabetes mellitus, enteropathy, and endocrinopathy (IPEX) syndrome results from non-polymorphic mutations in the forkhead/winged-helix domain of the scurfin protein (encoded by FOXP3);[25-28] in *sf* mice, a frameshift mutation results in a product lacking the forkhead domain of FOXP3.

Rationale for the Focus on Tregs. A key role for natural Tregs or induced Tregs in promoting tolerance in pre-clinical models of transplantation has been reported.[11,12] Early data from clinical transplantation studies suggest that FOXP3 profiles may be informative of outcomes. In recipients of stem cell allografts, Zorn et al. found a decreased frequency of peripheral blood CD4+CD25+ T cells and reduced expression of FOXP3 mRNA in patients with chronic graft versus host disease (GVHD) compared to those without GVHD.[29] In liver allograft recipients, Li et al. found an increase in the frequency of peripheral blood CD4+CD25high+ cells in operationally tolerant liver graft recipients (n=12), compared to subjects on immunosuppressive therapy (n=19).[30] Baan and colleagues reported that FOXP3 mRNA levels were higher in

endomyocardial biopsies of cardiac allograft recipients with acute rejection than in those without rejection.[31]

***A Damage Control Role for Tregs in Allograft Rejection*.** How can we reconcile the seemingly contradictory observations that pre-clinical transplantation tolerance models are distinguished by high levels of Tregs, whereas pre-clinical as well as clinical studies of acute rejection including the current study demonstrate that acute rejection is associated with higher levels of FOXP3 expression compared to no rejection biopsies? We suggest that the explanation may reside in part in the models used to ascertain the role of Tregs and the different mechanisms by which the Tregs control effector cells. We suggest that Tregs regulate anti-allograft immunity not only by preventing an immune response but also by controlling the graft destructive effector arm. In this formulation, the immune repertoire during an episode of acute rejection includes the generation of both graft destructive cells such as CTLs and graft protective Tregs, and the balance between these two competing forces determines the outcome. Support for a local damage control role for Tregs is provided by the finding that the suppressor cells in the BDC2.5 T cell receptor of transgenic mice do not block autoreactive BDC2.5 effector cells by clonal deletion or clonal anergy mechanisms, but rather by constraining the antigen experienced effectors from causing islet damage.[32] This observation has been forcefully reinforced by the demonstration that FOXP3-dependent CD25+ Tregs prevent destructive insulitis by "reining in destructive T cells inside the islets, more than during the initial activation in the draining lymph nodes" in the BDC2.5/NOD FOXP3 scurfy mice.[33]

Rationale for the mRNAs Measured in our Study. We measured urinary cell level of mRNA for FOXP3, and the levels of mRNA for CD25, CD3E, and perforin. We selected these mRNAs for measurement since these were the same mRNAs that were measured in our earlier study demonstrating that urinary cell mRNA levels predict TCMR reversal.[14]

CONCLUDING REMARKS

The outcome following an episode of acute rejection is difficult to predict even with the invasive core needle biopsy procedure. Results from the current investigation validate our earlier observation that urinary cell levels of FOXP3 mRNA but not the levels of CD3E, CD25 and perforin mRNA predict reversal. A reproducibility crisis in replicating research findings including those published in high impact journals has been emphasized in the literature.[16,34] The current investigation, in replicating our single center findings in a multicenter study setting involving 480 kidney allograft recipients, provides additional support for our original findings and may stimulate investigation of the prognostic utility of FOXP3 in additional clinical settings.

REFERENCES

1. Strom TB, Tilney NL, Carpenter CB, Busch GJ. Identity and cytotoxic capacity of cells infiltrating renal allografts. *N Engl J Med.* 1975;292(24):1257-1263.
2. Li B, Hartono C, Ding R, et al. Noninvasive diagnosis of renal-allograft rejection by measurement of messenger RNA for perforin and granzyme B in urine. *N Engl J Med.* 2001;344(13):947-954.
3. Muthukumar T, Ding R, Dadhania D, et al. Serine proteinase inhibitor-9, an endogenous blocker of granzyme B/perforin lytic pathway, is hyperexpressed during acute rejection of renal allografts. *Transplantation.* 2003;75(9):1565-1570.
4. Ding R, Li B, Muthukumar T, et al. CD103 mRNA levels in urinary cells predict acute rejection of renal allografts. *Transplantation.* 2003;75(8):1307-1312.
5. Tatapudi RR, Muthukumar T, Dadhania D, et al. Noninvasive detection of renal allograft inflammation by measurements of mRNA for IP-10 and CXCR3 in urine. *Kidney Int.* 2004;65(6):2390-2397.
6. Madden RL, Mulhern JG, Benedetto BJ, et al. Completely reversed acute rejection is not a significant risk factor for the development of chronic rejection in renal allograft recipients. *Transpl Int.* 2000;13(5):344-350.
7. Opelz G, Dohler B, Collaborative Transplant Study R. Influence of time of rejection on long-term graft survival in renal transplantation. *Transplantation.* 2008;85(5):661-666.
8. Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol.* 2003;4(4):330-336.
9. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. *Science.* 2003;299(5609):1057-1061.
10. Ramsdell F. Foxp3 and natural regulatory T cells: key to a cell lineage? *Immunity.* 2003;19(2):165-168.
11. Wood KJ, Sakaguchi S. Regulatory T cells in transplantation tolerance. *Nat Rev Immunol.* 2003;3(3):199-210.
12. Cobbold SP, Castejon R, Adams E, et al. Induction of foxP3+ regulatory T cells in the periphery of T cell receptor transgenic mice tolerized to transplants. *J Immunol.* 2004;172(10):6003-6010.
13. Albert MH, Liu Y, Anasetti C, Yu XZ. Antigen-dependent suppression of alloresponses by Foxp3-induced regulatory T cells in transplantation. *Eur J Immunol.* 2005;35(9):2598-2607.
14. Muthukumar T, Dadhania D, Ding R, et al. Messenger RNA for FOXP3 in the urine of renal-allograft recipients. *N Engl J Med.* 2005;353(22):2342-2351.
15. Suthanthiran M, Schwartz JE, Ding R, et al. Urinary-cell mRNA profile and acute cellular rejection in kidney allografts. *N Engl J Med.* 2013;369(1):20-31.
16. McNutt M. Reproducibility. *Science.* 2014;343(6168):229.
17. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016;533(7604):452-454.
18. Racusen LC, Colvin RB, Solez K, et al. Antibody-mediated rejection criteria - an addition to the Banff 97 classification of renal allograft rejection. *Am J Transplant.* 2003;3(6):708-714.
19. Le CT. A solution for the most basic optimization problem associated with an ROC curve. *Stat Methods Med Res.* 2006;15(6):571-584.
20. Sakaguchi S. The origin of FOXP3-expressing CD4+ regulatory T cells: thymus or periphery. *J Clin Invest.* 2003;112(9):1310-1312.
21. Schubert LA, Jeffery E, Zhang Y, Ramsdell F, Ziegler SF. Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J Biol Chem.* 2001;276(40):37672-37679.
22. Smyk-Pearson SK, Bakke AC, Held PK, Wildin RS. Rescue of the autoimmune scurfy mouse by partial bone marrow transplantation or by injection with T-enriched splenocytes. *Clin Exp Immunol.* 2003;133(2):193-199.
23. Khattri R, Kasprowicz D, Cox T, et al. The amount of scurfin protein determines peripheral T cell number and responsiveness. *J Immunol.* 2001;167(11):6312-6320.
24. Sakaguchi S. Naturally Arising CD4+ Regulatory T Cells for Immunologic Self-Tolerance and Negative Control of Immune Responses. *Annu Rev Immunol.* 2004;22:531-562.
25. Bennett CL, Christie J, Ramsdell F, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet.* 2001;27(1):20-21.
26. Brunkow ME, Jeffery EW, Hjerrild KA, et al. Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat Genet.* 2001;27(1):68-73.

27. Wildin RS, Ramsdell F, Peake J, et al. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nat Genet.* 2001;27(1):18-20.
28. Chatila TA, Blaeser F, Ho N, et al. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. *J Clin Invest.* 2000;106(12):R75-81.
29. Zorn E, Kim HT, Lee SJ, et al. Reduced frequency of FOXP3+ CD4+CD25+ regulatory T cells in patients with chronic graft-versus-host disease. *Blood.* 2005;106(8):2903-2911.
30. Li Y, Koshiba T, Yoshizawa A, et al. Analyses of peripheral blood mononuclear cells in operational tolerance after pediatric living donor liver transplantation. *Am J Transplant.* 2004;4(12):2118-2125.
31. Baan CC, van der Mast BJ, Klepper M, et al. Differential effect of calcineurin inhibitors, anti-CD25 antibodies and rapamycin on the induction of FOXP3 in human T cells. *Transplantation.* 2005;80(1):110-117.
32. Gonzalez A, Andre-Schmutz I, Carnaud C, Mathis D, Benoist C. Damage control, rather than unresponsiveness, effected by protective DX5+ T cells in autoimmune diabetes. *Nat Immunol.* 2001;2(12):1117-1125.
33. Chen Z, Herman AE, Matos M, Mathis D, Benoist C. Where CD4+CD25+ T reg cells impinge on autoimmune diabetes. *J Exp Med.* 2005;202(10):1387-1397.
34. Baker M. Is there a reproducibility crisis? *Nature.* 2016;533(7604):452-454.

Table S1. Oligonucleotide primers and TaqMan probes used to measure levels of each mRNA.

Type of mRNA	GenBank Accession Number	Sequence[A]	Location
FOXP3	NM_014009.3	Sense: 5’-GAGAAGCTGAGTGCCATGCA-3’	939-958
		Antisense: 5’-GGAGCCCTTGTCGGATGAT-3’	1025-1007
		Probe: 5’-FAM-TGCCATTTTCCCAGCCAGGTGG-TAMRA-3’	962-983
CD25	NM_000417.2	Sense: 5’-GACTGCTCACGTTCATCATGGT-3’	245-266
		Antisense: 5’-AATGTGGCGTGTGGGATCTC-3’	326-307
		Probe: 5'-FAM-AGAGCTCTGTGACGATGACCCGCC-TAMRA-3'	282-305
CD3E	NM_000733	Sense: 5'-AAGAAATGGGTGGTATTACACAGACA-3'	131-156
		Antisense: 5'-TGCCATAGTATTTCAGATCCAGGAT-3'	233-209
		Probe: 5'-FAM-CCATCTCTGGAACCACAGTAATATTGACATGCC-TAMRA-3'	170-202
Perforin	M28393	Sense: 5'-GGACCAGTACAGCTTCAGCACTG-3'	492-514
		Antisense: 5'-GCCCTCTTGAAGTCAGGGTG-3'	587-568
		Probe: 5'-FAM-TGCCGCTTCTACAGTTTCCATGTGGTACAC-TAMRA-3'	526-555

[A] The fluorogenic TaqMan probes were labeled with 6-carboxy-fluorscein (FAM) at the 5' end and with 6-carboxy-tetramethylrodamine (TAMRA) at the 3' end. FAM functioned as the reported dye and TAMRA or MGB as the quencher.

Table S2. NIH-Statistical and Clinical Coordinating Center (SACCC) Biopsy Form.

Protocol/Study Number:	CTOT04/CTOT-04	Sponsor:	NIAID SACCC
Version Number:	FINAL 18.00	Site:	Wilmington
DCM Name (Description):	BIOP (Biopsy Form)	Server:	Caribou
QG Glib Domain	CTOT	Oracle Instance:	RTPOCP

Fields designated with an asterisk (*) are to be completed by the Lead CDM and Tech Ops together

DCM Subset #/ Name[3]	Clinical Planned Event	Question Group Name[1]	Save Defaults Question Group Codes	eCRF Question Prompt	Response Options (short and long labels)	Question Name	Question Domain	Question Data Type[2]	Length Decimals	Question Codes[3]	Default Value	DVG Name	DVG Domain	SAS Label / SAS Name (if different than Question Name)
1 / BIO	BIOPSY	BIO		**Biopsy Date: (dd/mon/yyyy)**	01-31	BIOPDTD	CTOT	C	2	M		DAY	STD CDISCV31	Day of Biospy Date
					JAN - DEC	BIOPDTM	CTOT	C	3	M		MON	STD CDISCV31	Month of Biospy Date
					1900-2006*	BIOPDTY	CTOT	C	4	M		YEAR	STD CDISCV31	Year of Biospy Date
				Reason for Biopsy:	2=Suspected Rejection, 3=Graft Dysfunction, 99=Other	BIOREAS	CTOT		2	M		***BIOREASss3***	CTOT	Reason For Biopsy
				If other, specify:		***BIOOTH***	CTOT		30					Other Biopsy Reason
				Comments		***BIOCOMM***	CTOT		200					Biopsy Comments
		BANF		**Was sample adequate for measurement?**	1=Yes; 0=No	BADQYN	CTOT	C	2	M		YNss2	STD CDISCV31	Adequate Sample for Measurement
				If 'Yes', record BANFF Classification or Other Findings										
				BANFF Classification: (Check One.)	Normal; Borderline; Antibody mediated rejection (Grade I-ATN); Antibody mediated rejection (Grade II Capillary); Antibody mediated rejection (Grade III-Arterial); Mild acute cellular rejection (Grade IA); Mild acute cellular rejection (Grade IB); Moderate acute cellular rejection (Grade IIA)| Moderate acute cellular rejection (Grade IIB)| Severe acute cellular rejection (Grade III); Mild chronic allograft nephropathy (Grade I); Moderate chronic allograft nephropathy (Grade II); Severe chronic allograft nephropathy (Grade III); N/A	BBANFF	CTOT	C	50			BBANFF	CTOT	Biopsy BANFF Classification
				Other Findings: (Check all that apply.)				C					CTOT	
					1=Acute tubular necrosis	BOTHF1	CTOT	C	2			BOTHFss1	CTOT	Other Biopsy Findings 1
					2=Calcineurin toxicity	BOTHF2	CTOT	C	2			BOTHFss2	CTOT	Other Biopsy Findings 2
					3=Baacterial infection/Pyelonephitis	BOTHF3	CTOT	C	2			BOTHFss3	CTOT	Other Biopsy Findings 3
					4=Cytomegalovirus infection	BOTHF4	CTOT	C	2			BOTHFss4	CTOT	Other Biopsy Findings 4
					5=Polynioma/BK nephropathy	BOTHF5	CTOT	C	2			BOTHFss5	CTOT	Other Biopsy Findings 5
					6=Recurrent Disease	BOTHF6	CTOT	C	2			BOTHFss6	CTOT	Other Biopsy Findings 6
					7=Obstruction	BOTHF7	CTOT	C	2			BOTHFss7	CTOT	Other Biopsy Findings 7
					8=Post-transplant Lymphoproliferative Disease (PLTD)	BOTHF8	CTOT	C	2			BOTHFss8	CTOT	Other Biopsy Findings 8
					9=Glomerulosclerosis	BOTHF9	CTOT	C	2			BOTHFss9	CTOT	Other Biopsy Findings 9
					10=Vascular narrowing	BOTHF10	CTOT	C	2			BOTHFss10	CTOT	Other Biopsy Findings 10
					11=Interstitial fibrosis	BOTHF11	CTOT	C	2			BOTHFss11	CTOT	Other Biopsy Findings 11
					12=Tubular atrophy	BOTHF12	CTOT	C	2			BOTHFss12	CTOT	Other Biopsy Findings 12
					13=Other	BOTHF13	CTOT	C	2			BOTHFss13	CTOT	Other Biopsy Findings 13
						OTHFTXT	CTOT	C	200					Specify Other Findings

1 If Repeating, # of repeats is specified in parentheses

2 Type Abbreviations: C=Character, N=Numeric, D=Date, T=time

TMS-related Question Abbreviations: (QS)=Question_Set, (TV)=Thesaurus_Validated

–Questions with data types of Date or Time are extracted into SAS with a Character format

–Numerics with decimals are extracted into SAS with a length of Oracle defined length + 1 (i.e., Oracle 4.1=SAS 5.1)

3 Question Codes:
D= Derived (Question is Derived from other variables)
NE=Not Enterable (Question is displayed but Not Enterable on eCRF)
ND=Not Displayed (Question is Not Displayed on eCRF) QQ=Qualifying Question
M= Mandatory (Question is Mandatory and will generate discrepancy if left blank)
DV= Default Value (Question is populated with Default Value)

Notes: –Use ***Bold italics*** for all objects ***NEW*** to the Global Library

COMMENTS:

Table S3. 18S rRNA normalized, log_{10}-transformed levels of mRNA in urinary cells across diagnostic categories.[A]

Type of mRNA	Acute T-Cell-Mediated Rejection (N = 43 samples, 34 patients)	No Rejection Group (N = 162 samples, 126 patients)	Stable (No Biopsy) Group (N = 1524 samples, 199 patients)[B]	*P*-Value[C]	Borderline (N = 19 samples, 17 patients)	Antibody Mediated Rejection (N = 10 samples, 9 patients)	BKVN (N = 7 samples, 6 patients)[D]		*P*-Value[E] TCMR Vs.	*P*-Value[E] No Reject Vs.	*P*-Value[E] Stable Vs.
FOXP3	-1.168 (-1.938, -0.517)	-1.628 (-2.061, -1.161)	-1.717 (-2.132, -1.230)	0.0124	-1.429 (-1.948, -0.796)	-1.940 (-2.124, -1.200)	-1.178 (-1.690, -1.068)	No Rejection	0.0453		
								Stable	0.0053	0.2624	
								Borderline	0.7621	0.2279	0.1158
								AMR	0.1321	0.5003	0.6764
								BKVN	0.8268	0.2711	0.1467
CD25	-0.335 (-0.867, 0.041)	-0.675 (-1.247, -0.122)	-0.734 (-1.307, -0.187)	0.0133	-0.932 (-1.229, 0.096)	-0.321 (-0.365, -0.168)	-0.119 (-0.626, 0.103)	No Rejection	0.0472		
								Stable	0.0060	0.2480	
								Borderline	0.2810	0.9323	0.7682
								AMR	0.9198	0.1873	0.0770
								BKVN	0.7427	0.2148	0.1371
CD3E	0.601 (-0.233, 0.995)	-0.504 (-1.071, 0.256)	-0.243 (-0.895, 0.343)	<0.0001	-0.024 (-0.703, 0.305)	0.022 (-0.306, 0.385)	0.364 (-0.297, 0.462)	No Rejection	<0.0001		
								Stable	<0.0001	0.1123	
								Borderline	0.0412	0.2412	0.3825
								AMR	0.0383	0.1787	0.3822
								BKVN	0.1261	0.2643	0.3729
Perforin	0.289 (-0.160, 0.817)	-0.689 (-1.215, 0.028)	-0.682 (-1.273, -0.061)	<0.0001	-0.327 (-0.612, 0.087)	-0.56 (-0.933, -0.164)	0.177 (-0.380, 0.441)	No Rejection	<0.0001		
								Stable	<0.0001	0.3067	
								Borderline	0.0139	0.1555	0.0798
								AMR	0.0053	0.7140	0.5771
								BKVN	0.6614	0.0792	0.0530

[A] Median (lower, upper quartiles) log-transformed ratio of copies per microgram of total RNA to 18S rRNA copies (x 10^{-6}) per microgram of total RNA is shown for each mRNA measure. Biopsy-matched urine samples (collected minus 3 days to plus 1 day of biopsy) were included for those who underwent a biopsy - Acute T-Cell Mediated Rejection, No Rejection, Borderline, Antibody-Mediated Rejection, and BKVN. Urine samples from patients in the Stable (no biopsy) group were collected longitudinally on days 3, 7, 15, and 30 and in months 2, 3, 4, 5, 6, 9, and 12 post-transplantation.

[B] Stable graft function is defined by virtue of meeting the following criteria: (i) average of serum creatinine assessed at 6, 9, and 12 months post-transplantation less than or equal to 2.0 mg/dl; (ii) no graft loss or death during the first 12 months following transplantation; (iii) no treatment for acute rejection; (iv) no cytomegalovirus or BK virus infection; and (v) no clinical indication for a biopsy.

[C] *P*-values are calculated using Kruskal-Wallis test for differences among the Acute T-Cell-Mediated Rejection biopsy group, No Rejection biopsy group, and Stable (no biopsy) group.

[D] BKVN category is the major subset of the Other Findings biopsy group shown in Figure 1. The Other Findings biopsy group consisted of 8 patients contributing 9 biopsy matched urine samples, and the 6 patients with BKVN contributed 7 of the 9 biopsy matched urine samples.

[E] *P*-values are calculated using Mann-Whitney test of pairwise differences between groups.

Table S4. Absolute levels of mRNA in urinary cells across diagnostic categories.[A]

Type of mRNA	Acute T-Cell-Mediated Rejection (N = 43 samples, 34 patients)	No Rejection Group (N = 162 samples, 126 patients)	Stable (No Biopsy) Group (N = 1524 samples, 199 patients)[B]	*P* value[C]	Borderline (N = 19 samples, 17 patients)	Antibody Mediated Rejection (N = 10 samples, 9 patients)	BKVN (N = 7 samples, 6 patients)[D]		P-Value[E] TCMR Vs.	P-Value[E] No Reject Vs.	P-Value[E] Stable Vs.
FOXP3	96 (13, 604)	13 (6, 13)	13 (6, 53)	<0.0001	13 (6, 179)	13 (6, 13)	100 (13, 182)	No Rejection	<0.0001		
								Stable	<0.0001	0.7377	
								Borderline	0.0787	0.1230	0.1160
								AMR	0.0143	0.9245	0.9994
								BKVN	0.8768	0.0171	0.0152
CD25	867 (217, 3440)	204 (13, 683)	191 (13, 757)	<0.0001	186 (13, 1604)	772 (192, 929)	470 (423, 2663)	No Rejection	<0.0001		
								Stable	<0.0001	0.9694	
								Borderline	0.0294	0.6502	0.5602
								AMR	0.2417	0.0978	0.0963
								BKVN	0.7797	0.0567	0.0563
CD3	8826 (2767, 19140)	356 (13, 1803)	509 (108, 2487)	<0.0001	944 (90, 3598)	867 (278, 1934)	4024 (573, 5607)	No Rejection	<0.0001		
								Stable	<0.0001	0.0321	
								Borderline	0.0016	0.2429	0.5722
								AMR	0.0046	0.1982	0.4811
								BKVN	0.1984	0.0721	0.1272
Perforin	4119 (1144, 11110)	223 (13, 1183)	207 (13, 999)	<0.0001	398 (194, 1417)	289 (156, 1387)	1595 (700, 5557)	No Rejection	<0.0001		
								Stable	<0.0001	0.8614	
								Borderline	0.0003	0.1633	0.1169
								AMR	0.0037	0.6565	0.6177
								BKVN	0.4255	0.0328	0.0241

[A] Median absolute copy number per microgram of total RNA (lower, upper quartiles) of each mRNA measure, without 18S rRNA normalization and without log-transformation are shown. Biopsy-matched urine samples (collected minus 3 days to plus 1 day of biopsy) were included for those who underwent a biopsy - Acute T-Cell-Mediated Rejection, No Rejection, Borderline, Antibody-Mediated Rejection, and BKVN. Urine samples from patients in the Stable (no biopsy) group were collected longitudinally on days 3, 7, 15, and 30 and in months 2, 3, 4, 5, 6, 9, and 12 post-transplantation.

[B] Stable graft function is defined by virtue of meeting the following criteria: (i) average of serum creatinine assessed at 6, 9, and 12 months post-transplantation less than or equal to 2.0 mg/dl; (ii) no graft loss or death during the first 12 months following transplantation; (iii) no treatment for acute rejection; (iv) no cytomegalovirus or BK virus infection; and (v) no clinical indication for a biopsy.

[C] P-values are calculated using Kruskal-Wallis test among the Acute T-Cell-Mediated Rejection biopsy group, No Rejection biopsy group, and Stable (no biopsy) group.

[D] BKVN category is the major subset of the Other Findings biopsy group shown in Figure 1. The Other Findings biopsy group consisted of 8 patients contributing 9 biopsy matched urine samples, and the 6 patients with BKVN contributed 7 of the 9 biopsy matched urine samples.

[E] P-values are calculated using Mann-Whitney test of pairwise differences between groups.

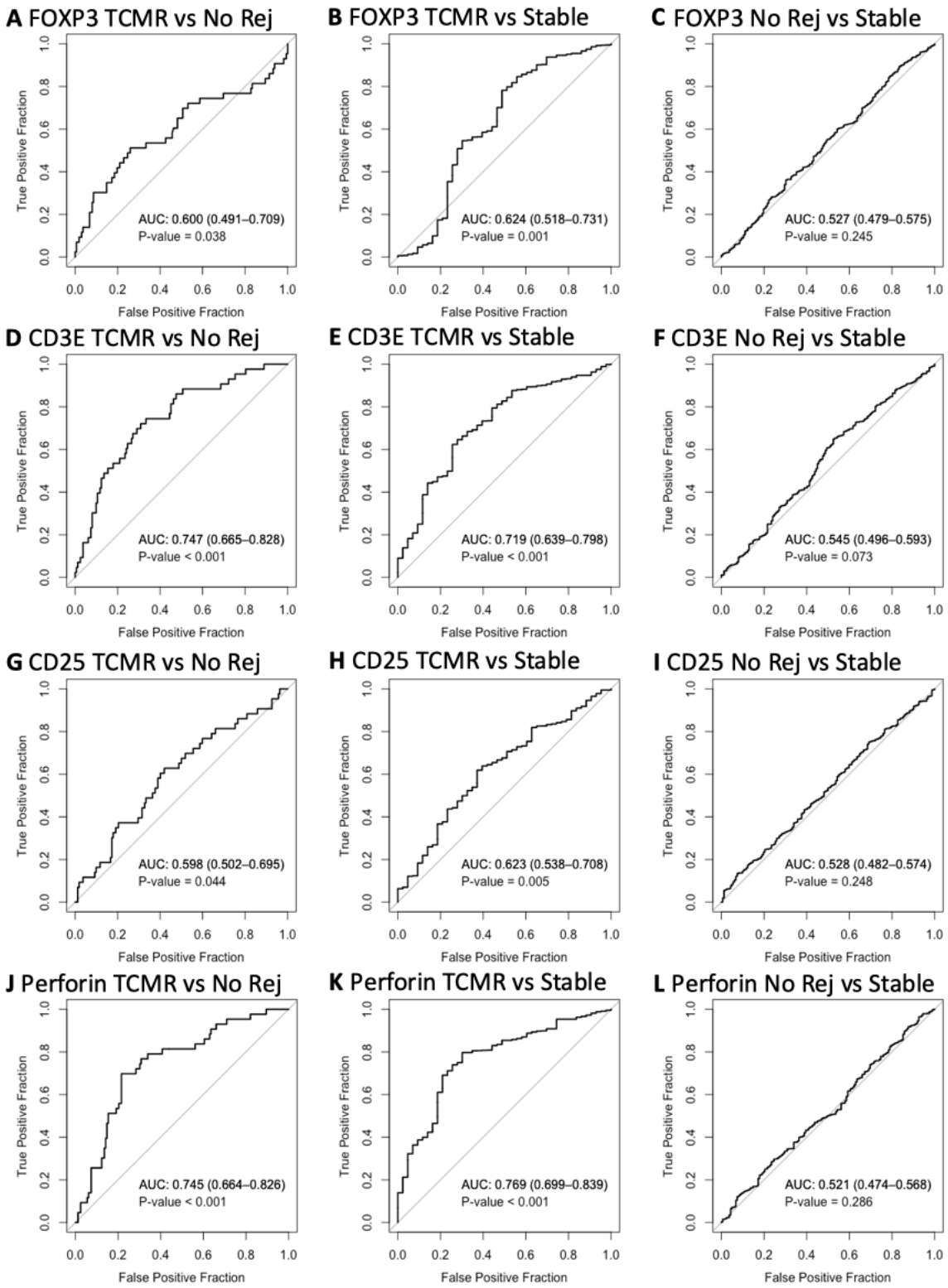
A FOXP3 TCMR vs No Rej
AUC: 0.600 (0.491–0.709)
P-value = 0.038
B FOXP3 TCMR vs Stable
AUC: 0.624 (0.518–0.731)
P-value = 0.001
C FOXP3 No Rej vs Stable
AUC: 0.527 (0.479–0.575)
P-value = 0.245
D CD3E TCMR vs No Rej
AUC: 0.747 (0.665–0.828)
P-value < 0.001
E CD3E TCMR vs Stable
AUC: 0.719 (0.639–0.798)
P-value < 0.001
F CD3E No Rej vs Stable
AUC: 0.545 (0.496–0.593)
P-value = 0.073
G CD25 TCMR vs No Rej
AUC: 0.598 (0.502–0.695)
P-value = 0.044
H CD25 TCMR vs Stable
AUC: 0.623 (0.538–0.708)
P-value = 0.005
I CD25 No Rej vs Stable
AUC: 0.528 (0.482–0.574)
P-value = 0.248
J Perforin TCMR vs No Rej
AUC: 0.745 (0.664–0.826)
P-value < 0.001
K Perforin TCMR vs Stable
AUC: 0.769 (0.699–0.839)
P-value < 0.001
L Perforin No Rej vs Stable
AUC: 0.521 (0.474–0.568)
P-value = 0.286
True Positive Fraction
False Positive Fraction
0.0
0.2
0.4
0.6
0.8
1.0

Fig. S1. ROC curve analysis of urinary cell mRNA levels. ROC curve analysis of urinary cell FOXP3 mRNA level yielded an area under the curve (AUC) of 0.600 (95% CI, 0.491 to 0.709, *P*=0.038, [A]) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.624 (95% CI, 0.518-0.731, P=0.001, [B]) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.527 (95% CI, 0.479-0.575, P=0.245, [C]) when the No Rejection biopsy group was compared with the Stable Graft Function group. ROC curve analysis of urinary cell CD3E mRNA levels yielded an AUC of 0.747 (95% CI, 0.665 to 0.828, *P*< 0.001, [D]) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.719 (95% CI, 0.639-0.798, P<0.001, [E]) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.545 (95% CI, 0.496-0.593, P=0.073, [F]) when the No Rejection biopsy group was compared with the Stable Graft Function group. ROC curve analysis of urinary cell CD25 mRNA levels yielded an AUC of 0.596 (95% CI, 0.502 to 0.695, *P*=0.044, [G]) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.623 (95% CI, 0.538-0.708, P=0.005, [H]) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.528 (95% CI, 0.482-0.574, P=0.248, [I]) when the No Rejection biopsy group was compared with the Stable Graft Function group. ROC curve analysis of urinary cell perforin mRNA levels yielded an AUC of 0.745 (95% CI, 0.684 to 0.826, *P*< 0.001, [J]) when the TCMR biopsy group was compared with the No Rejection biopsy group, 0.769 (95% CI, 0.699-0.839, P<0.001, [K]) when the TCMR biopsy group was compared with the Stable Graft Function group; and 0.521 (95% CI, 0.474-0.568, P=0.286, [L]) when the No Rejection biopsy group was compared with the Stable Graft Function group.

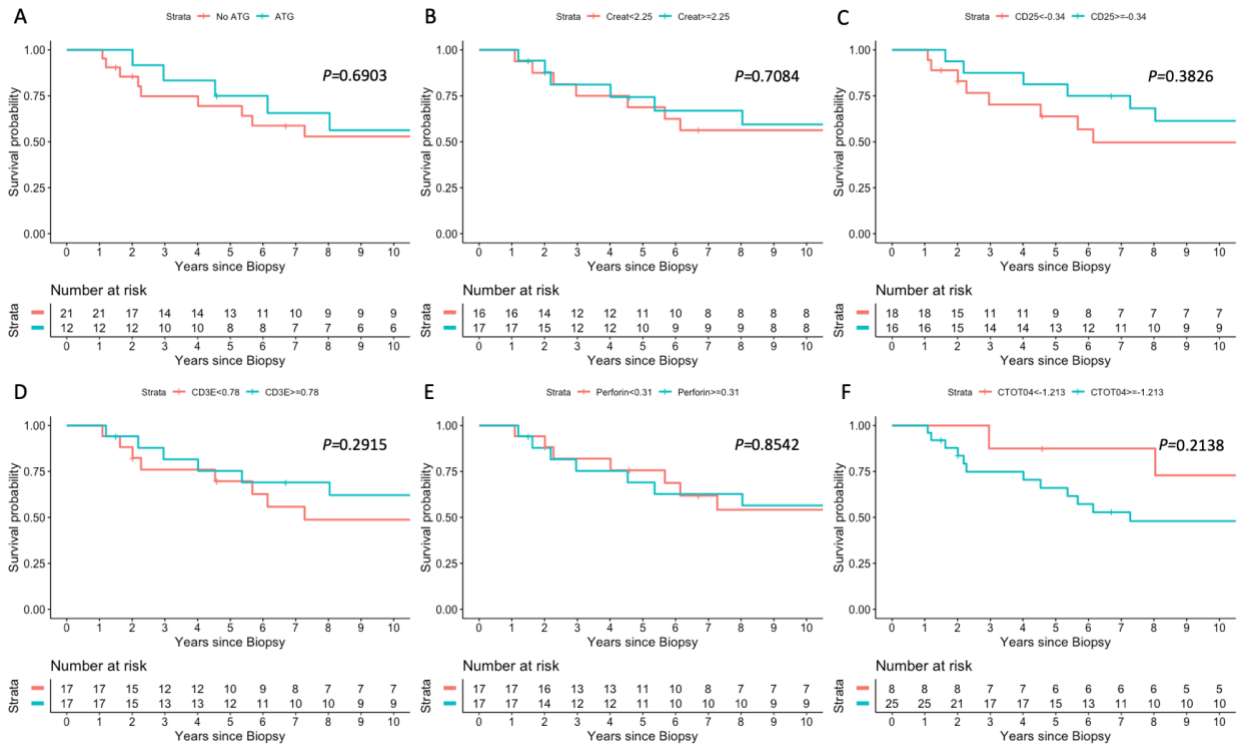

Figure S2. Kaplan-Meier kidney allograft survival curves. (A) Comparison of survival curves for the 33 patients with at least one episode of TCMR, stratified by anti-rejection therapy with thymoglobulin (Yes ATG) vs. ATG not used for anti-rejection therapy (No ATG) showed no significant difference based on treatment of TCMR with ATG (*P*=0.6903). (B) Comparison of survival curves for the same 33 patients, stratified by a median split of serum creatinine measured at time of first TCMR biopsy showed no significant difference ((*P*=0.7084). (C) Comparison of survival curves for the same 33 patients, stratified by a median split of CD25 mRNA level measured at time of first TCMR biopsy showed no significant difference (*P*=0.3826). (D) Comparison of survival curves for the same 33 patients, stratified by a median split of CD3E mRNA measured at time of first TCMR biopsy showed no significant difference (*P*=0.2915). (E) Comparison of survival curves for the same 33 patients, stratified by a median split of perforin measured at time of first TCMR biopsy showed no significant difference (*P*=0.8542). (F) Comparison of survival curves for the same 33 patients, stratified by the CTOT-

04 three-gene TCMR diagnostic signature cut-point of -1.213 showed no significant difference (*P*=0.2138). Time to event was calculated from date of TCMR biopsy (or the date of last TCMR biopsy for the 3 patients with multiple episodes) until graft failure or last follow-up date. Subjects were censored if they experienced death prior to graft failure or were lost to follow-up. *P*-values are based on log-rank tests. At-risk tables are shown in each plot, just above the X-axis.