



# A molecular classifier for predicting future graft loss in late kidney transplant biopsies

Gunilla Einecke,<sup>1,2</sup> Jeff Reeve,<sup>1</sup> Banu Sis,<sup>1</sup> Michael Mengel,<sup>1</sup> Luis Hidalgo,<sup>1</sup> Konrad S. Famulski,<sup>1</sup> Arthur Matas,<sup>3</sup> Bert Kasiske,<sup>4</sup> Bruce Kaplan,<sup>5</sup> and Philip F. Halloran<sup>1</sup>

<sup>1</sup>Alberta Transplant Applied Genomics Centre, University of Alberta, Edmonton, Alberta, Canada. <sup>2</sup>Department of Nephrology, Hannover Medical School, Hannover, Germany. <sup>3</sup>Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA. <sup>4</sup>Hennepin County Medical Center, University of Minnesota, Minneapolis, Minnesota, USA. <sup>5</sup>Department of Medicine, Nephrology Section, Arizona Health Science Centre, University of Arizona, Tucson, Arizona, USA.

**Kidney transplant recipients that develop signs of renal dysfunction or proteinuria one or more years after transplantation are at considerable risk for progression to renal failure. To assess the kidney at this time, a “for-cause” biopsy is performed, but this provides little indication as to which recipients will go on to organ failure. In an attempt to identify molecules that could provide this information, we used micorarrays to analyze gene expression in 105 for-cause biopsies taken between 1 and 31 years after transplantation. Using supervised principal components analysis, we derived a molecular classifier to predict graft loss. The genes associated with graft failure were related to tissue injury, epithelial dedifferentiation, matrix remodeling, and TGF- $\beta$  effects and showed little overlap with rejection-associated genes. We assigned a prognostic molecular risk score to each patient, identifying those at high or low risk for graft loss. The molecular risk score was correlated with interstitial fibrosis, tubular atrophy, tubulitis, interstitial inflammation, proteinuria, and glomerular filtration rate. In multivariate analysis, molecular risk score, peritubular capillary basement membrane multilayering, arteriolar hyalinosis, and proteinuria were independent predictors of graft loss. In an independent validation set, the molecular risk score was the only predictor of graft loss. Thus, the molecular risk score reflects active injury and is superior to either scarring or function in predicting graft failure.**

## Introduction

Kidney transplants that develop dysfunction or proteinuria after one year following transplantation are at considerable risk for progression to renal failure (1). Certain histopathologic features, particularly interstitial fibrosis and tubular atrophy (IFTA), correlate with graft dysfunction, treatment response, and risk of progression to failure in transplants as well as in native kidneys (2–8). This has led to the belief that late failure of kidney transplants is due to progressive nonspecific scarring, possibly related to calcineurin inhibitor toxicity. However, IFTA is common in kidney transplants, reflecting the burden of injury including donor death, organ harvest, and the transplantation process, and mostly develops in the first year (9). Recent analyses (1, 10) indicate that the main cause of late graft loss is not unexplained scarring or calcineurin inhibitor toxicity but specific disease entities, particularly late antibody-mediated injury and recurrent disease.

Identifying the molecules associated with graft failure could potentially lead to interventions that would slow the progression of organ failure. Some of the individual molecules that predict risk of failure in native proteinuric kidney disease include VEGF and molecules associated with activation of intracellular hypoxia response (11). In kidney transplant biopsies, many molecules show altered expression related to rejection or injury (12–16). However, no comprehensive analysis of the relationship between the transcriptome and allograft survival has been performed.

The emergence of microarrays permits a genome-wide survey of the transcripts associated with future failure in renal

allografts presenting with clinical indications for a biopsy, i.e., biopsies for cause (BFCs). The present study analyzed the relationship between gene expression in late BFCs in human kidney transplants and subsequent graft loss and assessed the predictive value of gene expression alone and in combination with histologic lesions and clinical variables. We evaluated the performance of these genes in an independent validation set and in a population of early biopsies that have a very low risk of subsequent graft failure.

## Results

### *Patient demographics and graft survival in the BFC population*

Because almost all failures occurred in patients who presented for a BFC after 1 year following transplantation (1), this group was selected for the analysis of risk prediction. The study population included 105 consecutive consenting patients who underwent BFCs between 1 and 31 years after transplantation (median, 57 months). Where more than 1 biopsy was available per patient, only the first biopsy was used for analyses. Median time to graft loss was 14 months, and the median follow-up after biopsy for patients without death or graft loss was 32 months.

We observed 30 graft failures during follow-up, and 4 patients died with a functioning graft. Demographics and clinical characteristics of all patients are outlined in Table 1. Grafts that subsequently failed had higher incidence of proteinuria and rapid deterioration in function before biopsy and lower glomerular filtration rate (GFR) at time of biopsy. There were no differences in primary disease, time after transplantation, maintenance immunosuppression, or incidence of anti-HLA antibodies between grafts that subsequently failed after biopsy

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Citation for this article:** *J Clin Invest.* 2010;120(6):1862–1872. doi:10.1172/JCI41789.



**Table 1**  
Demographics, immunosuppression, and indication for biopsy

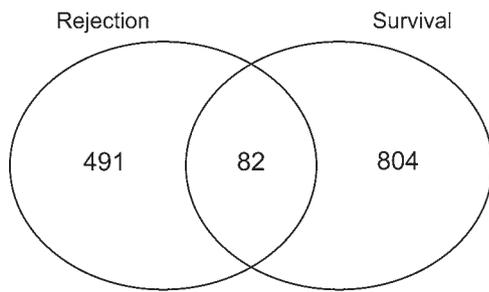
Patient demographics	All patients <i>n</i> = 105	Failed grafts <i>n</i> = 30	Non-failed grafts <i>n</i> = 75	<i>P</i>
Time of follow-up after biopsy (d)	774	404	922	3 × 10 <sup>-6</sup>
Time of biopsy after transplant (d)	1734	1,853	1,688	0.64
Diagnosis				
ABMR	11	5	6	0.19
TCMR	14	3	11	0.53
Borderline rejection	11	2	9	0.73
Mixed ABMR and TCMR	3	2	1	0.20
Glomerulonephritis	22	9	13	0.15
BK virus	1	0	1	1.0
Other diagnoses	43	9	34	0.15
Recipient sex, M/F	65/40 (62%/38%)	20/10 (67%/33%)	45/30 (60%/40%)	0.82
Race				
White	64 (61%)	18 (60%)	46 (61%)	0.99
Black	10 (10%)	4 (13%)	6 (8%)	0.70
Other	15 (14%)	4 (13%)	11 (15%)	0.98
Unknown	16 (15%)	4 (13%)	12 (16%)	0.98
Primary disease				
Diabetic nephropathy	13 (12%)	4 (13%)	9 (12%)	0.98
Hypertension/large-vessel disease	6 (6%)	1 (3%)	5 (7%)	0.80
Glomerulonephritis/vasculitis	51 (49%)	20 (67%)	31 (41%)	0.06
Interstitial nephritis/pyelonephritis	8 (8%)	1 (3%)	7 (9%)	0.58
Polycystic kidney disease	14 (13%)	1 (3%)	13 (17%)	0.16
Others	7 (7%)	2 (7%)	5 (7%)	1.00
Unknown etiology	6 (6%)	1 (3%)	5 (7%)	0.80
Previous transplant	9 (9%)	0 (0%)	9 (12%)	0.14
Donor sex, M/F	41/64 (39%/61%)	9/21 (30%/70%)	32/43 (43%/57%)	0.49
Deceased donor transplants	57 (54%)	18 (60%)	39 (52%)	0.76
Indication for biopsy				
Rapid deterioration of graft function	26 (25%)	14 (47%)	12 (16%)	0.004
Slow deterioration of graft function	39 (37%)	7 (23%)	32 (43%)	0.18
Stable impaired graft function	7 (7%)	1 (3%)	6 (8%)	0.69
Investigate proteinuria	15 (14%)	7 (23%)	8 (11%)	0.25
Follow-up from previous biopsy	6 (6%)	1 (3%)	5 (7%)	0.80
Others	6 (6%)	0 (0%)	6 (8%)	0.28
Indication unknown	6 (6%)	0 (0%)	6 (8%)	0.28
Estimated GFR (average ml/min ± SD)	47 ± 18	38 ± 20	50 ± 17	0.002
Proteinuria-positive	57 (54%)	25 (83%)	32 (43%)	<0.001
HLA antibody status				
PRA-positive	64 (61%)	21 (70%)	43 (57%)	0.49
PRA class I-positive	44 (42%)	12 (40%)	32 (43%)	0.97
PRA class II-positive	48 (46%)	16 (53%)	32 (43%)	0.61
DSA-positive	48 (46%)	18 (60%)	30 (40%)	0.18
DSA class I-positive	18 (17%)	7 (23%)	11 (15%)	0.57
DSA class II-positive	38 (36%)	15 (50%)	25 (33%)	0.28
Maintenance immunosuppressive regimens at biopsy				
MMF, tacrolimus, steroid	37 (35%)	11 (37%)	26 (35%)	0.98
MMF, tacrolimus	1 (1%)	1 (3%)	0 (0%)	0.28
MMF, cyclosporine, steroid	26 (25%)	7 (23%)	19 (25%)	0.98
MMF, steroids	3 (3%)	0 (0%)	3 (4%)	0.54
Azathioprine, cyclosporine, steroids	20 (19%)	5 (17%)	15 (20%)	0.93
Others	18 (17%)	6 (20%)	12 (16%)	0.89

Bold type indicates significant *P* values. MMF, mycophenolate mofetil; PRA, panel-reactive antibody.

and those that did not. As previously reported, the main disease diagnoses in biopsies from grafts that failed were antibody-mediated rejection (ABMR) (either C4d-positive or C4d-negative) and glomerulonephritis (1).

**Gene expression signature of future graft loss**

To identify all genes associated with graft loss, we performed a Cox regression on the entire dataset of 105 late biopsies, using all probesets that passed the interquartile range (IQR) filter



**Figure 1**

Overlap of genes associated with graft loss and those associated with rejection. We identified the genes associated with graft loss in a Cox regression model ( $P < 0.0001$ ) and compared them with those genes whose association with rejection versus nonrejection was significant ( $t$  test) at  $P < 0.0001$ . The number of genes that were unique to each list as well as the number of overlapping genes are shown.

( $n = 11,500$ ) (see Methods). 886 genes represented by 1,312 probesets were significantly associated with graft failure (598 positively and 288 negatively) at the 0.0001 level (uncorrected  $P$  value) (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI41789DS1).

We compared the genes associated with subsequent graft loss with those associated with rejection at the time of biopsy in the same dataset. Rejection-associated genes had been identified in a previous analysis (14) using the BioConductor software package limma (17). Rejection-associated genes and genes associated with future graft loss were derived from the same biopsies. Of the 886 genes associated with graft loss, only 82 (9%) overlapped with the gene list associated with rejection at the same  $P$  value cutoff (Figure 1). The transcripts associated with future graft loss were primarily those that had been annotated as associated with tissue injury, matrix remodeling, and epithelial dedifferentiation (Table 2), while the rejection-associated genes had been annotated as reflecting inflammation, i.e.,  $IFN-\gamma$  effects, infiltrating T cells, and macrophages (18).

**Developing a gene expression risk score for predicting graft loss**

Because we aimed to develop a biopsy-based risk prediction method, we built a gene-based classifier to predict graft loss. Classifier results were obtained using a multiple 10-fold cross-validation method (19). Details of the analysis are described in Methods. A diagram illustrating the process of building the classifier and validation steps is shown in Supplemental Figure 1.

The classifier used 2,748 probesets on at least one occasion, and 117 of the probesets were used in at least 50% of the classifiers. The 30 annotated genes used most frequently by the classifier to assign the risk of graft loss are shown in Table 3, ordered by the proportion of times each gene was used in all the cross-validation/resampling loops. Twenty-seven of the top 30 genes (90%) had previously been annotated in experimental transplantation systems as members of pathogenesis-based transcript sets (PBTs) reflecting tissue injury and matrix remodeling (see Methods and Supplemental Table 2).

We used the molecular classifier to assign a prognostic molecular risk score to each biopsy using supervised principal components analysis (PCA) (20). Based on the average risk score across the test sets in the 100 validation loops, we assigned patients into either the high- or low-risk group ( $n = 52$  high-risk,  $n = 53$  low-risk patients) (Figure 2). The range of risk scores for each biopsy across the validation loops is shown in Supplemental Figure 2. The mean risk score in those kidney grafts surviving to 1 year after biopsy was  $-0.31$  versus  $1.85$  in those grafts that failed by 1 year ( $t$  test,  $P = 9.3 \times 10^{-8}$ ). In the high-risk group, 25 of 52 patients progressed to graft loss after biopsy, compared with only 5 losses in the low-

risk group (log-rank test,  $P = 3 \times 10^{-7}$ ). Kaplan-Meier survival curves for the 2 risk groups are shown in Figure 3.

**Relationship between risk score and clinico-pathological features**

Table 4 shows the correlation between the molecular risk score and histologic and clinical variables. The risk score was negatively correlated with renal function (estimated GFR) at the time of biopsy and positively correlated with proteinuria, interstitial inflammation, atrophy, and fibrosis and less strongly with tubulitis. It was not correlated ( $P > 0.05$ ) with time of the biopsy after transplantation, glomerular or arterial changes, or arteriolar hyalinosis.

We examined the relationship between risk score and time to failure (Figure 4). Among the patients with graft loss, a higher risk score was associated with shorter time to failure ( $P = 0.0006$ ). Among patients censored only because of end of study, there

**Table 2**

PBT overrepresentation analysis for probesets associated with graft loss

Biological process	Transcript set <sup>A</sup>	P
Severe tissue injury	GST	$<10^{-14}$
Reversible tissue injury (intermediate phase)	IRIT(I)	$<10^{-14}$
TGF- $\beta$ signaling	TGFB(1)	$<10^{-14}$
TGF- $\beta$ signaling	TGFB(2)	$2.3 \times 10^{-14}$
Reversible tissue injury (late phase)	IRIT(L)	$1.7 \times 10^{-13}$
Endothelial cell activation	ENDAT	$1.7 \times 10^{-12}$
Fibroblast activation	FIBET	$5.5 \times 10^{-9}$
Severe tissue injury	CIST	$8.6 \times 10^{-9}$
Alternative macrophage activation	AMAT1	$5.3 \times 10^{-8}$
Tissue injury due to alloresponse	NIRIT	$6.8 \times 10^{-5}$
Macrophage activation by IL-4	AMAT	0.0002
Fibroblast activation	FIBTG	0.0005
Classical macrophage activation	IMAT	0.0013
$IFN-\gamma$ effects	GRIT2	0.0017
Parenchymal homeostasis	KT1	0.0024
Reversible tissue injury (early phase)	IRIT(E)	0.016
$IFN-\gamma$ effects	GRIT1	0.025
Parenchymal homeostasis	KT2	0.19
Macrophage infiltration/activation	QCMAT	0.43
T cell infiltration	QCAT	0.72
NK cell infiltration	NKST	0.88
B cell infiltration	BAT	0.94
T cell infiltration	CAT1	0.99
Plasma cell infiltration	IGT	1.00

<sup>A</sup>Transcript sets were defined in previous analyses and represent distinct biological processes in renal allograft dysfunction. They are explained in Methods.

**Table 3**

The 30 genes used most frequently by the classifier to assign risk of graft failure

Affymetrix ID	Gene <sup>A</sup>	Membership in transcript set <sup>B</sup>	Proportion of times gene was used in validation loops	<i>P</i> <sup>C</sup>	Expression in control kidneys	Expression in failed grafts	Expression in non-failed grafts	Ratio failed/censored grafts
215646_s_at	<i>VCAN</i>	FIBET NIRIT TGFB	1	$1.25 \times 10^{-10}$	33	184	54	3.423
202238_s_at	<i>NNMT</i>	NIRIT	1	$4.04 \times 10^{-10}$	94	570	170	3.353
207052_at	<i>HAVCR1</i>	IRIT(I)	0.999	$3.00 \times 10^{-12}$	48	97	38	2.526
208083_s_at	<i>ITGB6</i>	KT1	0.999	$1.30 \times 10^{-9}$	48	220	71	3.11
202018_s_at	<i>LTF</i>	IRIT(L)	0.986	$1.54 \times 10^{-7}$	153	1,620	431	3.761
201939_at	<i>PLK2</i>	AMAT	0.982	$3.39 \times 10^{-11}$	68	178	83	2.14
222486_s_at	<i>ADAMTS1</i>	IRIT(I)	0.982	$3.63 \times 10^{-11}$	131	204	98	2.08
204627_s_at	<i>ITGB3</i>	CIST GST	0.977	$2.20 \times 10^{-10}$	35	75	36	2.088
223218_s_at	<i>NFKBIZ</i>	AMAT	0.972	$4.90 \times 10^{-9}$	471	897	373	2.405
225681_at	<i>CTHRC1</i>	CIST	0.97	$5.54 \times 10^{-8}$	23	185	63	2.937
202376_at	<i>SERPINA3</i>	GST IRIT(I)	0.969	$3.96 \times 10^{-8}$	132	498	196	2.537
205466_s_at	<i>HS3ST1</i>	AMAT	0.962	$3.38 \times 10^{-9}$	37	104	47	2.183
203021_at	<i>SLPI</i>	AMAT	0.958	$1.27 \times 10^{-7}$	134	758	294	2.574
206336_at	<i>CXCL6</i>		0.954	$8.20 \times 10^{-7}$	13	77	26	2.965
206840_at	<i>AFM</i>	KT1	0.947	$1.01 \times 10^{-6}$	623	93	257	0.363
204124_at	<i>SLC34A2</i>		0.946	$1.03 \times 10^{-10}$	62	108	60	1.798
204259_at	<i>MMP7</i>	GST IRIT(L)	0.944	$5.75 \times 10^{-8}$	586	2,474	1,054	2.347
209774_x_at	<i>CXCL2</i>	AMAT CIST GST	0.941	$4.78 \times 10^{-9}$	116	72	36	2.021
209278_s_at	<i>TFPI2</i>	FIBET KT1	0.939	$7.73 \times 10^{-8}$	63	353	137	2.571
1569003_at	<i>TMEM49</i>		0.935	$1.26 \times 10^{-7}$	159	162	74	2.183
217767_at	<i>C3</i>	AMAT IRIT(I)	0.935	$3.80 \times 10^{-7}$	294	1,427	599	2.381
211981_at	<i>COL4A1</i>	FIBET IRIT(I) TGFB	0.929	$2.84 \times 10^{-11}$	267	650	378	1.717
211864_s_at	<i>MYOF</i>	IMAT	0.92	$2.03 \times 10^{-11}$	169	400	231	1.728
227697_at	<i>SOCS3</i>	CAT1	0.913	$2.62 \times 10^{-8}$	63	48	23	2.034
205729_at	<i>OSMR</i>	IRIT(I) GST	0.91	$2.09 \times 10^{-9}$	58	125	67	1.88
202404_s_at	<i>COL1A2</i>	IRIT(I) TGFB	0.908	$4.18 \times 10^{-7}$	104	779	323	2.414
212531_at	<i>LCN2</i>	AMAT	0.905	$8.60 \times 10^{-8}$	72	159	81	1.965
211161_s_at	<i>COL3A1</i>	IRIT(L) TGFB GST	0.905	$3.15 \times 10^{-7}$	141	867	360	2.407
226621_at	<i>FGG</i>	IRIT(E)	0.903	$4.46 \times 10^{-9}$	190	505	272	1.855
210764_s_at	<i>CYR61</i>	FIBET IRIT(I) GST	0.877	$1.07 \times 10^{-9}$	855	229	131	1.741

<sup>A</sup>Where there are multiple probesets per gene, only the most frequently used (highest "Proportion") is shown. <sup>B</sup>Transcript sets were previously defined and represent biological processes in allograft dysfunction; they are explained in Methods. <sup>C</sup>From Cox regression.

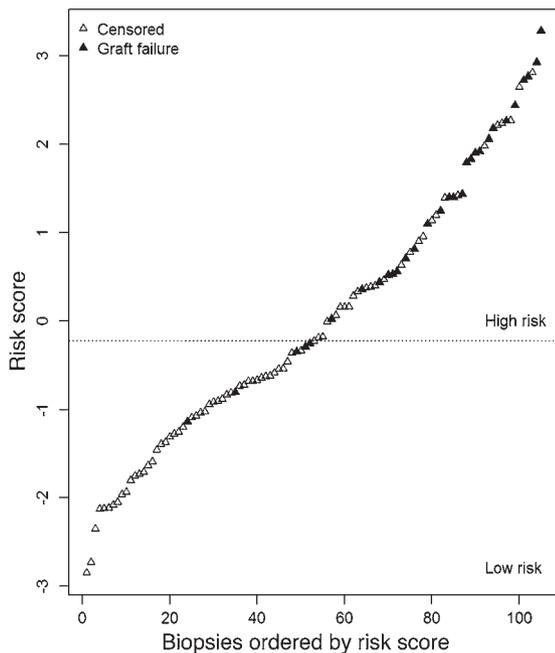
was no relationship between risk score and time to censoring. Among patients censored because of death, all 4 of whom died fewer than 400 days after biopsy, there was a significant correlation between risk score and time to death ( $P = 0.01$ ). This may reflect the effect of a failing kidney transplant on the risk of death, but the small number of observations precludes firm conclusions. Thus, we included deaths with functioning grafts as censored data in the classifier and Cox regression analyses. We conclude that the risk score predicts not only whether failure will occur but also time to failure.

The association of molecular, clinical, and histologic features with graft loss was assessed in a univariate Cox regression analysis (Table 5). In this analysis, the features significantly associated with graft loss were molecular risk score, proteinuria, interstitial fibrosis, tubular atrophy, peritubular capillary basement membrane multilayering (PTCML), mesangial matrix score, interstitial inflammation, glomerulitis, low GFR, and absence of arteriolar hyalinosis. Many variables were not associated with outcome, including time of biopsy after transplantation, the presence of HLA antibody, C4d staining, transplant glomerulopathy, capillaritis, and arterial changes.

To assess the independent associations of these variables, we performed a forward stepwise multivariate regression (Table 6) of all features that reached significance ( $P < 0.05$ ) in the univariate analysis. This analysis identified the independent predictors of graft loss as the molecular risk score, peritubular basement membrane multilayering, proteinuria, and arteriolar hyalinosis. Surprisingly, neither GFR nor IFTA was significantly associated with graft loss when the risk score was included in the multivariate analysis.

Since the presence of proteinuria and GFR were known before the biopsy, we repeated the multivariate analysis excluding these factors to determine which information arising from the biopsy (histology, microarray results) predicted failure. When GFR and proteinuria were excluded, the independent risk factors were the molecular risk score, PTCML, and absence of arteriolar hyalinosis.

When the receiver operating characteristic (ROC) curves for each of the features independently associated with graft loss in the multivariate analysis were compared (Figure 5A), the molecular risk score showed a greater area under the curve (AUC = 0.83) than the clinical or histologic features (AUC = 0.63–0.76). When the  $P$  values for the differences in AUCs were calculated by permutation test, they were not significant:  $P = 0.17$  for risk score versus pro-



**Figure 2**

Molecular risk scores in individual biopsies. We built a molecular classifier based on 105 kidney transplant biopsies taken for clinical indications 1 year or more after transplantation and used the result to assign a molecular risk score to each biopsy. Biopsies were sorted by risk score; each biopsy is represented by 1 triangle. Biopsies from patients with subsequent graft loss are indicated in black. Biopsies were assigned into high- or low-risk groups, with the threshold determined by the median risk score.

teinuria,  $P = 0.06$  versus PTCML, and  $P = 0.002$  versus arterial hyaline sclerosis. However, the maximum accuracy obtainable using the risk score was significantly greater than with proteinuria (Figure 5B) ( $P = 0.001$ ). In fact, using the presence of proteinuria as a threshold led to a lower accuracy (i.e., 0.69) than did the null distribution (accuracy, 0.71) (i.e., guessing that all samples survived). In contrast, all risk score thresholds greater than 0.0 produced higher accuracies than the null distribution.

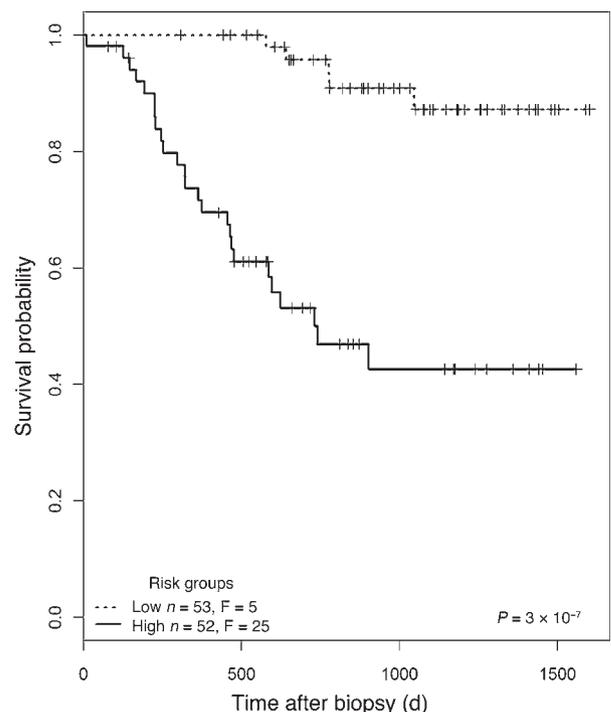
**Robustness of the classifier results**

*Single gene classifiers.* We assessed the strength of predictions from simple single gene classifiers, based solely on the expression values of the highest ranked and the 100th, the 500th, and the 5,000th ranked genes in each training set (Supplemental Figure 3). Predictions from single genes, even those that were not near the top of the ranked list of genes, were very similar to those using the full PCA model. Thus, the risk for graft loss can be predicted adequately by any of a great many genes. This is due to the highly coordinated changes in expression found in many thousands of genes.

*Application of the classifier to an independent validation set.* We assessed the performance of the classifier in an independent set of biopsies ( $n = 48$ ) taken for clinical indication more than 1 year after transplantation at the University of Minnesota, with patient characteristics and clinical and histologic features similar to those in the studies described above. This population experienced 11 graft losses and had a median follow-up time of 406 days. Risk scores were calculated by supplying the classifier built from our main dataset with the gene expression values from the Minnesota dataset. The relationship of the molecular risk score to graft loss, the Kaplan-Meier plots of high- versus low-risk groups (defined using the risk threshold in the main dataset), and the ROC curve illustrating the tradeoff between sensitivity and specificity (Figure 6, A–C) showed results similar to those in the original data set. In the univariate analysis in this validation set, only the molecular risk score was significantly associated

with subsequent graft loss ( $P = 0.0004$ ). Thus, the risk score was superior to classic predictors such as IFTA and low GFR.

*Application of the classifier to a patient population at low risk of graft loss.* To assess whether high risk scores always predict risk for graft loss, we applied the classifier algorithm (based on late biopsies) to a set of biopsies taken within the first year after transplantation, which had very little risk of graft loss during



**Figure 3**

Kaplan-Meier plots for the two risk groups. Biopsies were separated into high- and low-risk groups, as shown in Figure 2. F, failed.



**Table 4**  
Correlation of molecular risk scores with histologic and clinical features

Variable	Spearman correlation coefficient	P
Time after transplantation	-0.005	0.96
Renal function (estimated GFR)	-0.54	$3.0 \times 10^{-9}$
Proteinuria	0.30	0.002
Glomerulitis (g score)	0.16	0.10
Transplant glomerulopathy (cg score)	0.15	0.14
Interstitial inflammation (i score)	0.50	$4.1 \times 10^{-8}$
Interstitial fibrosis (ci score)	0.46	$7.7 \times 10^{-7}$
Tubulitis (t score)	0.28	0.004
Tubular atrophy (ct score)	0.49	$1 \times 10^{-7}$
Arteritis (v score)	0.07	0.46
Fibrous intimal thickening (cv score)	0.03	0.75
Arteriolar hyalinosis (ah score)	-0.11	0.25
Mesangial matrix (mm score)	0.14	0.17
Peritubular capillaritis (ptc score)	0.13	0.18
PTCML	0.21	0.06

Histologic lesions were graded according to the Banff classification (32, 33) as described in Methods.

follow-up in our study (Figure 7). Of 73 early biopsies (<1 year after transplantation), 44 (60%) had a risk score above the high-risk threshold from the late biopsy analysis (sensitivity, 1.00; specificity, 0.41; Table 7). However, only 2 of these grafts subsequently failed. Both of these were actually biopsied late in the first year, at 273 days and 360 days. Most of the remaining 42 early biopsies, which were classified as high risk but did not fail during the follow-up period, were diagnosed as T cell-mediated rejection (TCMR) ( $n = 8$ ), borderline TCMR ( $n = 4$ ), polyoma

virus nephropathy ( $n = 3$ ), or acute tubular necrosis ( $n = 21$ ). When comparing early and late biopsies with high-risk scores, we found differences in diagnostic categories: early biopsies with high-risk scores (but low incidence of failure) had mainly acute tubular necrosis and TCMR, while late biopsies with high-risk scores (and high incidence of failure) were often diagnosed as having ABMR or glomerulonephritis (Table 8).

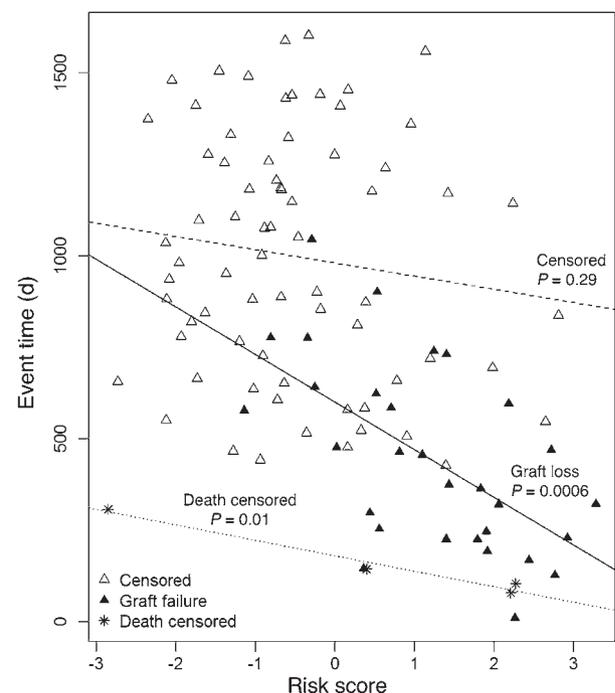
To exclude that the high rate of false-positive predictions in the early biopsy group was due to a statistical bias (based on the fact that the classifier was built on late biopsies and applied to the clinically different group of early biopsies), we built a separate classifier on the entire dataset (early and late biopsies). In this analysis, the molecular risk score continued to have a high predictive value for graft loss, but the difference between early and late biopsies in terms of positive predictive value remained (Table 7).

Thus reversible/treatable early injury and rejection can induce the same gene expression changes as the disease processes leading to graft loss in the late biopsies, without leading to graft failure.

## Discussion

The present study developed a gene expression-based classifier to predict kidney allograft failure after a late biopsy (>1 year after transplantation) for clinical indications. The molecular risk scores were strongly associated with graft loss. Using previously annotated transcript sets reflecting major biological events in renal transplants, we found that the transcripts used by the classifier to predict graft failure were those reflecting tissue injury, dedifferentiation of the epithelium, and tissue remodeling, but not those reflecting inflammation (*IFNG* effects and T cell or macrophage infiltration). The predictive ability of the classifier was high, with a sensitivity of 0.83 and a specificity of 0.63 at the high/low risk threshold. The gene expression risk score correlated with fibrosis and atrophy, interstitial inflammation, and glomerulitis but independently predicted graft failure in a mul-

**Figure 4**  
Relationship between risk score and failure/censoring time. Time to event (graft failure, patient death, or end of follow-up) is plotted against the molecular risk score for each biopsy. Each biopsy is represented by one symbol. Biopsies from patients with subsequent graft loss are represented as black triangles; biopsies from patients who died with a functioning graft are represented by asterisks. All other biopsies are represented by white triangles. Regression lines were drawn separately for patients censored for end of study, those censored for patient death with a functioning graft, and those with graft loss.





**Table 5**  
Univariate analysis of features associated with graft loss

Feature	Beta	r <sup>2</sup>	HR	P (likelihood ratio)
Molecular risk score	0.84	0.31	2.33 (1.78–3.05)	3.3 × 10 <sup>-10</sup>
Proteinuria	2.63	0.22	13.9 (3.28–58.7)	4.3 × 10 <sup>-7</sup>
GFR	-0.048	0.15	0.95 (0.93–0.98)	4.3 × 10 <sup>-5</sup>
Interstitial fibrosis (ci score)	0.74	0.10	2.09 (1.33–3.27)	0.0010
Arteriolar hyalinosis (ah score)	-1.29	0.09	0.28 (0.13–0.58)	0.0014
Tubular atrophy (ct score)	0.73	0.08	2.05 (1.28–3.27)	0.0025
PTCML	0.52	0.08	1.66 (1.11–2.48)	0.0091
Mesangial matrix (mm score)	0.46	0.05	1.59 (1.08–2.33)	0.018
Interstitial inflammation (i score)	0.50	0.05	1.65 (1.08–2.53)	0.021
Glomerulitis (g score)	0.44	0.04	1.56 (1.04–2.32)	0.044
Peritubular capillaritis (ptc score)	0.22	0.02	1.25 (0.92–1.69)	0.17
PRA	0.64	0.02	1.89 (0.71–5.02)	0.18
Fibrous intimal thickening (cv score)	-0.66	0.02	0.52 (0.21–1.27)	0.18
DSA	-0.071	0.05	0.93 (0.40–0.44)	0.30
Transplant glomerulopathy (cg score)	0.18	0.01	1.19 (0.86–1.66)	0.31
C4d staining	0.66	0.02	1.94 (0.52–0.73)	0.35
ABMR (C4d <sup>-</sup> or C4d <sup>+</sup> )	0.49	0.02	1.64 (0.80–3.35)	0.17
Time of biopsy after transplantation	0.079	0.002	1.08 (0.79–1.48)	0.62
Arteritis (v score)	-0.14	0.001	0.87 (0.38–1.99)	0.73
Tubulitis (t score)	0.063	0.0009	1.06 (0.72–1.58)	0.76

HR, hazard ratio.

tivariate analysis. The risk score was validated in an independent set of late BFCs.

The risk score is an indicator of active injury, whose significance depends on the disease state of the patient. Risk predictions were accurate when applied to the late biopsy population, where subsequent progression is considerable, but not when applied to early biopsies, where the diseases operating have a low risk of progression when treated. Early after transplantation the diseases causing injury are self-limited (acute tubular necrosis) or reversible by therapy (e.g., TCMR), but in late biopsies the diseases causing injury are progressive and unresponsive to treatment (ABMR, recurrence of primary disease). Thus, the probability of progression after a biopsy for clinical indications is conferred both by the presence of active injury (the risk score) and by the diseases inducing the active injury response and their potential for spontaneous resolution or response to therapy. The classifier and the individual genes predict risk not in absolute terms but in relation to the diseases or injury mechanisms disturbing the kidney and triggering the biopsy.

The transcripts predicting graft loss reflect an ongoing response to injury, including epithelial distress and dedifferentiation, with reexpression of developmental genes and loss of transcripts associated with differentiated epithelium as well as remodeling of the matrix. Many of the molecules used by the classifier to predict progression to renal failure are already known to be involved in responses to injury and play important roles in kidney development. For example, nicotinamide *N*-methyltransferase (NNMT) and versican (VCAN), both associated with cell migration in cancer (21–24), may represent dedifferentiation and epithelial-mesenchymal transition. ITGB6 is expressed by renal epithelium during stress and injury (25) and activates TGFB1 (26). HAVCR1 (also known as KIM1) is a well-established feature of the injured kidney (27). Interestingly, the collagen genes were not prominent in the list, suggesting that the genes predicting graft loss represent an active response to injury in the epithelium and matrix but

not fibrogenesis per se. In previous analyses, we have performed extensive comparisons between microarray results and RT-PCR, which showed excellent reproducibility of both methods and confirmed the robustness of microarray results (28).

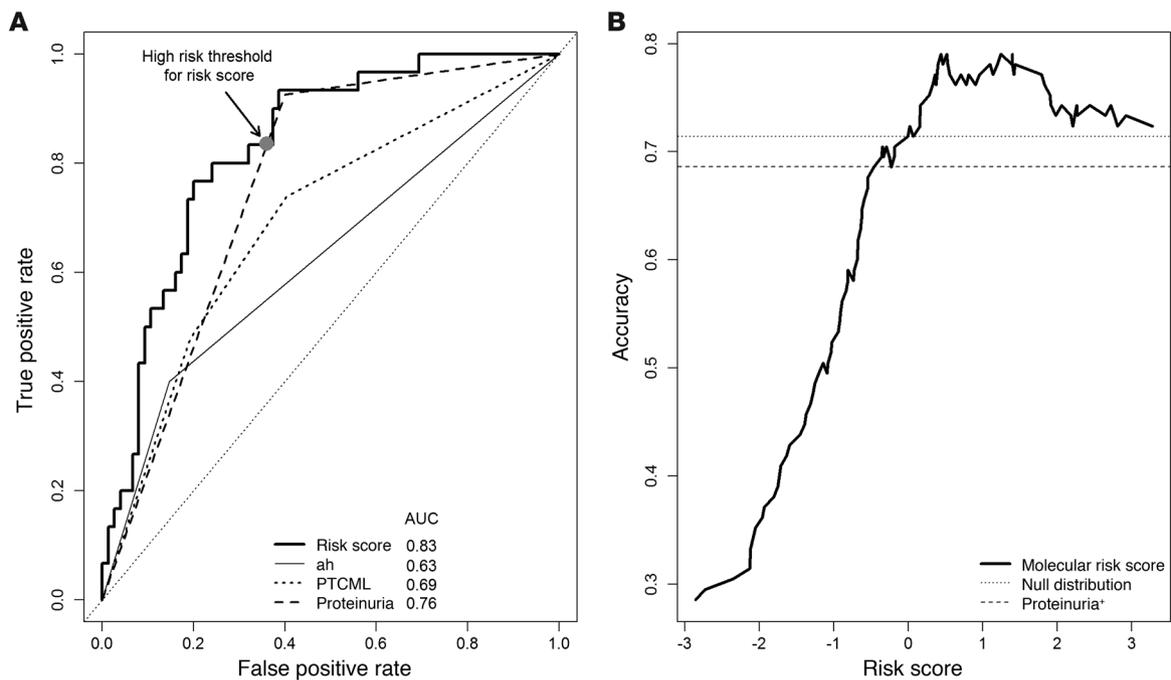
The genes associated with graft loss in late biopsies indicate that an active ongoing tissue response to injury is the final common pathway linking mechanisms of inflammation and noninflammatory disease states to parenchymal loss, dysfunction, and eventual kidney failure. The changes in the expression of these genes represent a stereotyped response of the tissue to injury, a structured program. Many transcripts were also found in mouse iso-grafts, which allowed us to map the time-dependent intrinsic responses of the nephron to a clearly defined and self-limited stress, indicating that these genes represent inherently reversible injury (29, 30), if the injury

mechanism is self limited or treatable. The gene list was similar to the transcript changes we recently found correlating with GFR disturbances in BFCs (31), which also reflect perturbations in the parenchyma and are indicators of an ongoing response to injury and thus potentially of a reversible epithelial injury repair process. This was further corroborated by observing high-risk scores in early biopsies from kidneys that did not progress to failure. Thus, the transcript changes reflect a perturbation in kidney biology that indicates ongoing remodeling and repair but not inevitable decline toward failure.

The risk score emerges as an independent predictor in multivariate analysis, indicating that transcriptome changes provide information beyond that derived from histology, demographics, GFR, or proteinuria. A molecularly based test has the potential to be more objective than a biopsy read by a pathologist, despite the standardization of the Banff criteria, because of intraobserver variation. In addition, the features of active injury are not captured by morphology but are reflected by gene expression changes. Thus, the risk score emerges as a more robust predictor of risk in the validation set than the classical factors: atrophy and fibrosis, low GFR, and proteinuria. The fact that not all grafts with a high-risk score failed, resulting in false positives,

**Table 6**  
Multivariate analysis of features associated with graft loss

Feature	Beta	HR	P
Molecular risk score	0.99	2.70 (1.82–4.00)	8.5 × 10 <sup>-7</sup>
PTCML	0.62	1.86 (1.16–2.97)	0.001
Arteriolar hyalinosis (ah score)	-1.39	0.25 (0.10–0.62)	0.003
Proteinuria	1.93	6.91 (1.56–30.56)	0.01

**Figure 5**

ROC curves of molecular risk score compared with clinical and histologic features. (A) To assess the sensitivity and specificity of the molecular classifier compared with that of clinical features, we compared the ROC curve derived from the molecular risk score with variables significantly associated with graft loss in a multivariate analysis. AUC is indicated for each parameter. (B) Accuracy (total correct predictions/total samples) for varying risk score thresholds for predicting graft failure. The dashed line represents the accuracy when the presence of proteinuria was used as the single threshold for predicting failure. The dotted line is the accuracy when all samples are predicted to survive. ah, arteriolar hyalinosis.

is not unexpected given the fact that patients were treated after the biopsy, resulting in reversal of the injury process in some cases. Since we did not include protocol biopsies performed after 1 year, we do not know whether there is a molecular signature of patients at high risk for graft failure in those without symptoms/signs of ongoing injury.

The superiority of the risk score to the classical features associated with progression to renal failure after a biopsy (atrophy, fibrosis, low GFR, proteinuria) indicates that these actually predict progression because they are correlated with an active injury response in biopsies for clinical indications. Atrophy, fibrosis, and low GFR indicate nephron loss, but the risk score subsumes this risk because it actually reflects nephrons in distress. This may offer hope that progression to renal failure is not inevitable if we can arrest the diseases causing progression. In other words, we find evidence here not of an overwhelming “point of no return” but of active injury that could be reversed if we find therapies that arrest the cause of the active injury.

## Methods

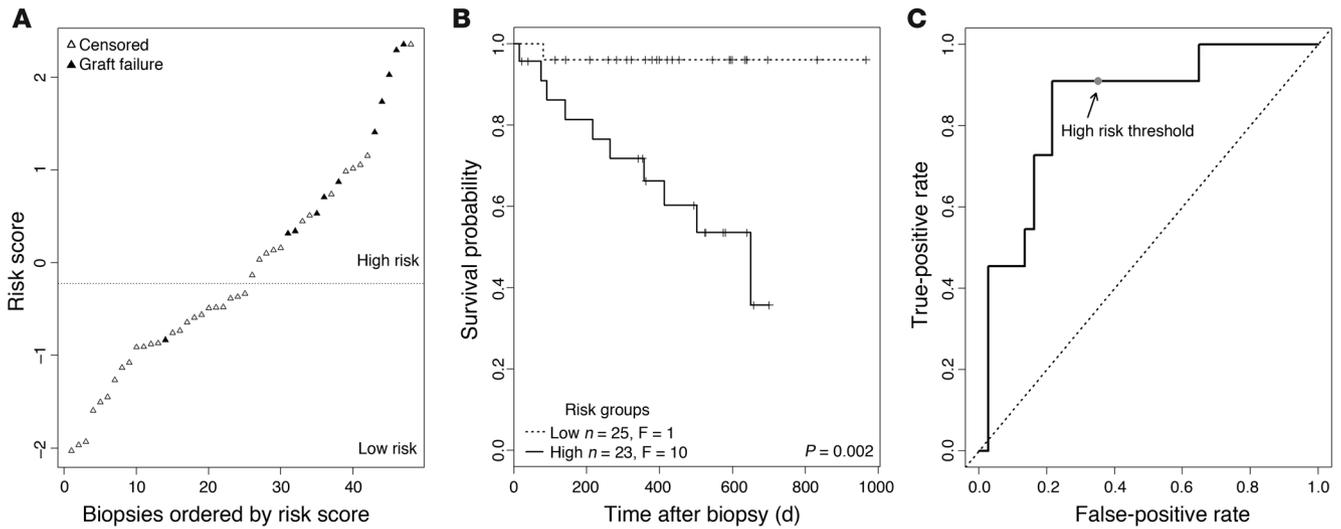
**Patient population, specimens, and data collection.** Written informed consent was obtained from all study patients. The study was approved by the institutional review boards of the University of Alberta (issue 5299), the University of Illinois, Chicago (protocol 2006-0544), the University of Minnesota (protocol HSC#0606 M87646), and the Hennepin County Medical Center (protocol HSR#06-2670). All consenting renal transplant patients undergoing BFCs as standard of care between September 2004 and October 2007 at the University of Alberta or between November 2006

and February 2007 at the University of Illinois were included in the analysis. In addition to our cross-validation analysis of the Edmonton dataset, biopsies obtained from Minnesota between September 2006 and September 2007 were used as an independent validation set.

Biopsies were obtained under ultrasound guidance by using spring-loaded needles (ASAP Automatic Biopsy, Microvasive). In addition to the cores required for standard histopathology, we collected one core for gene expression studies. The biopsy sample processing was performed as described in detail in our previous study (14).

**Histopathologic diagnosis.** All biopsies were assessed using the updated Banff 07 criteria (32, 33) by a pathologist who was blinded to the results of molecular studies. All biopsies had adequate cortical tissue for analysis by Banff criteria, with the exception of 2 biopsies that had no arteries. Diagnostic groups included TCMR, borderline TCMR, ABMR, IFTA not otherwise specified, glomerulonephritis, acute tubular necrosis, and BK virus nephropathy. In addition to ABMR with positive C4d staining (as specified in the Banff classification), we included the emerging category of C4d-negative ABMR (published in ref. 1), which is defined by the presence of circulating donor-specific HLA antibodies and evidence of microcirculation changes in the biopsy (presence of glomerulitis, transplant glomerulopathy, peritubular capillaritis, PTCML).

**PBTs.** PBTs represent biological processes during rejection and other types of injury in renal allografts. They serve as an annotation tool, with which genes identified in our analysis can be assigned to biological processes relevant to transplantation. PBTs were derived previously from mouse transplant models and in vitro human cell lines (15, 29, 30, 34–40). The definitions and algorithms for the PBTs referenced in the tables can be found in Supplemental Table 2. The PBTs included IFN- $\gamma$ -inducible



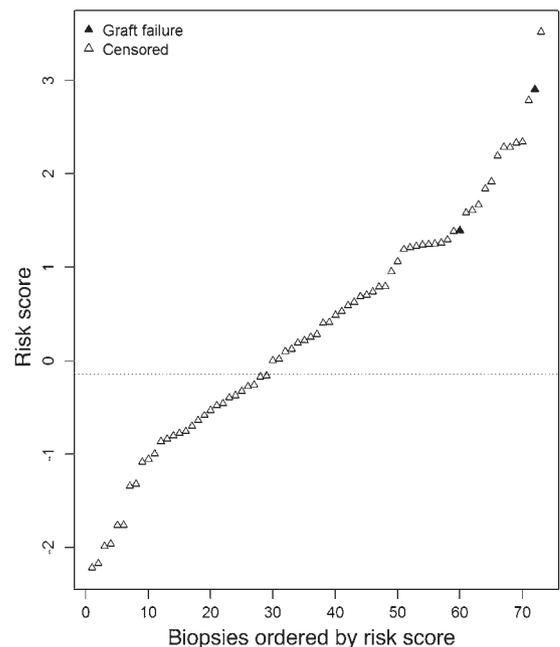
**Figure 6**  
 Performance of the classifier in an independent validation set. Based on the classifier results derived from the 105 biopsies in our dataset, we calculated risk scores for each sample in an independent set of biopsies ( $n = 48$ ) with similar clinical and histologic features. **(A)** Risk scores in relation to graft loss in the validation set, showing the risk threshold derived from the main ( $n = 105$ ) dataset. **(B)** Kaplan-Meier plots of the 2 risk groups in the validation set, using the threshold shown in Figure 6A. **(C)** ROC curves illustrating sensitivity/specificity of the risk scores in the validation set.

transcripts: GRIT1 and GRIT2 (34); cytotoxic T cell-associated transcripts: CAT1 (15) and QCAT (35); NK cell-associated transcripts: NKST (36); classical macrophage activation transcripts: IMAT (37); alternative macrophage activation transcripts: AMAT1 (38) and AMAT (IL-4-inducible transcripts); B cell-associated transcripts: BAT (39); immunoglobulin transcripts: IGT (39); endothelial transcripts: ENDAT (40); injury and repair transcripts: IRIT subsets IRIT(E), IRIT(I), and IRIT(L) (29); transcripts associated with severe tissue injury: GST and CISTS (38); kidney parenchymal transcripts: KT1 and KT2 (30); and transcript sets reflecting fibroblast activation (FIBET and FIBTG) or TGF- $\beta$  activation (TGFB[1] and TGFB[2]).

*Data analysis.* Because we aimed to develop a biopsy-based risk prediction method, graft survival was assessed as time between biopsy and graft failure/censoring, not time between transplantation and failure/censoring. Patients were censored for the end of study (July 26, 2009), death with functioning graft, or loss to follow-up. Graft failure was defined as return to dialysis ( $n = 30$ ).

Microarray data files were preprocessed using robust multichip averaging (RMA) in Bioconductor and subjected to variance-based filtering (41) as described previously (34). Nonspecific IQR filtering was used to eliminate probesets with low variation across the dataset. 11,500 of the original

**Figure 7**  
 Molecular risk scores in biopsies taken within 1 year after transplantation. The molecular risk score has high predictive value in patients at risk of graft loss. To assess whether the associated gene expression changes are unique to the population at risk, we applied the classifier results to a set of biopsies taken within the first year after transplantation.





**Table 7**  
Sensitivity and specificity of classifiers built on late biopsies only or early and late biopsies combined<sup>A</sup>

	Sensitivity	Specificity	Accuracy	PPV	NPV
<b>Classifier built on late biopsies only</b>					
Applied to early biopsies	1.00	0.41	0.42	0.05	1.00
Applied to late biopsies	0.83	0.63	0.69	0.47	0.90
<b>Classifier built on early and late biopsies</b>					
Applied to all biopsies	0.90	0.58	0.64	0.32	0.97
Applied to early biopsies	1.00	0.61	0.62	0.07	1.00
Applied to late biopsies	0.90	0.56	0.66	0.46	0.93

<sup>A</sup>The classifier was built either on late biopsies only or all biopsies (early plus late). For the calculations in this table, the threshold used for assigning a guess of failure versus non-failure was the median risk score.

54,675 probesets passed this filtering step and were retained for further analysis. Expression and phenotype data, as well the CEL files, are available at the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>; accession number GSE21374).

Classifier results were obtained using a multiple 10-fold cross-validation method (19). In each iteration, the data were divided into 10 roughly equal-sized subsamples. Nine of the subsamples were used to predict the risk scores of the remaining “left-out” subsample. This procedure was repeated 10 times, each time using a different left-out subsample, so that all biopsies received a single predicted risk score. Within each of the 10 folds of this algorithm, the genes used in the classifier were reselected based only on those samples not being left out, i.e., only using the training set for that left-out subsample. The supervised PCA method uses a shrinkage algorithm (20, 42), ordering the selected genes by their Cox regression *P* values within each training set. The shrinkage procedure itself selects for genes whose expression is most stable within samples from the same phenotypic class (e.g., within failures and within non-failures). One of the desirable consequences of this method is that many genes with similar *P* values, in terms of their association with graft loss, can be eliminated in such a way as to retain those with the most informative predictive content. No fixed *P* value cutoff is used. In each training set, the shrinkage threshold (equal to the number of genes used) was chosen by maximizing the predictive accuracy in that particular training set. This entire procedure was repeated 100 times (each time using a different random 10-fold data split, following the method used in ref. 19), and each biopsy’s average risk score over all 100 iterations was recorded. Based on the average risk score, patients were assigned to one of 2 risk groups (high and low), using the median across all biopsies as the cutoff between both groups. In the case of multiple biopsies from one patient, only the first biopsy was used. A diagram illustrating the process of building the classifier and validation steps is shown in Supplemental Figure 1.

Single gene classifiers were assessed by the same cross-validation method. In each training set, the *n*th ranked gene (*n* = 1, *n* = 100, *n* = 500, or 5,000) was determined, and its expression value in the corresponding test set was used as the risk score. Since a different *n*th ranked gene might occur in every training/test set split, all expression values were first standardized by subtracting that gene’s mean expression value and dividing by the gene’s standard deviation.

The full classifier derived from the Edmonton dataset was also applied to the gene expression values of an independent validation set from Minnesota, and risk scores were assigned.

Univariate associations between gene expression and graft survival were assessed by Cox regression analysis (*P* value cutoff = 0.0001). Multivariate regressions using the molecular risk score, pathology lesions, and clinical variables were built using a forward stepwise model using all features that reached significance (*P* < 0.05) in the univariate analysis. False discovery rates were estimated using the R package *fdrtool* (43). Unless otherwise specified, a *P* value of 0.05 was considered statistically significant.

### Acknowledgments

Special thanks go to Zija Jacaj for help with collection of the clinical data and to Vido Ramassar and Anna Hutton for technical support. This research has been supported

by funding and/or resources from Genome Canada, Genome Alberta, the University of Alberta, the University of Alberta Hospital Foundation, Roche Molecular Systems, Hoffmann-La Roche Canada Ltd., the Alberta Ministry of Advanced Education and Technology, the Roche Organ Transplant Research Foundation, the Kidney Foundation of Canada, and Astellas Canada. P. Halloran held a Canada Research Chair in Transplant Immunology until 2008 and currently holds the Muttart Chair in Clinical Immunology.

Received for publication November 18, 2009, and accepted in revised form April 26, 2010.

Address correspondence to: Philip F. Halloran, Director, Alberta Transplant Applied Genomics Centre, University of Alberta, 250 Heritage Medical Research Centre, Edmonton, Alberta T6G 2S2, Canada. Phone: 780.407.8880; Fax: 780.407.3417; E-mail: phil.halloran@ualberta.ca.

**Table 8**  
Diagnostic breakdown of high-risk biopsies based on the late-biopsy classifier

Diagnosis <sup>A</sup>	No. of biopsies <1 yr after Tx	No. of biopsies >1 yr after Tx
	(total <i>n</i> = 44; 2 failures)	(total <i>n</i> = 52; 25 failures)
TCMR	8	5 (1)
Borderline TCMR	4	2 (1)
C4d <sup>-</sup> antibody-mediated rejection	0	10 (6)
C4d <sup>+</sup> antibody-mediated rejection	1	9 (4)
Transplant glomerulopathy	0	3 (1)
Mixed TCMR and ABMR	1 (1)	5 (2)
Glomerulonephritis or recurrent disease	1	9 (4)
Isolated v lesion	2	1 (1)
BK virus nephropathy or probable BK	4 (1)	0
Calcineurin inhibitor toxicity	1	2 (2)
IFTA, not otherwise specified	0	3 (2)
Acute tubular necrosis	21	0
Miscellaneous other	1	3 (1)

<sup>A</sup>Diagnoses were defined according to the Banff classification, as described in Methods. Numbers in parenthesis indicate the number of failed grafts in the category. Tx, transplantation.



1. Einecke G, et al. Antibody-mediated microcirculation injury is the major cause of late kidney transplant failure. *Am J Transplant.* 2009;9(11):2520–2531.
2. Gaber LW, et al. Correlation between Banff classification, acute renal rejection scores and reversal of rejection. *Kidney Int.* 1996;49(2):481–487.
3. Colvin RB, et al. Evaluation of pathologic criteria for acute renal allograft rejection: reproducibility, sensitivity, and clinical correlation. *J Am Soc Nephrol.* 1997;8(12):1930–1941.
4. Haas M, Kraus ES, Samaniego-Picota M, Racusen LC, Ni W, Eustace JA. Acute renal allograft rejection with intimal arteritis: histologic predictors of response to therapy and graft survival. *Kidney Int.* 2002;61(4):1516–1526.
5. Agodoa L, et al. Assessment of structure and function in progressive renal disease. *Kidney Int Suppl.* 1997; 63:S144–S150.
6. Bohle A, Grund KE, Mackensen S, Tolon M. Correlations between renal interstitium and level of serum creatinine. Morphometric investigations of biopsies in perimembranous glomerulonephritis. *Virchows Arch A Pathol Anat Histol.* 1977;373(1):15–22.
7. Bohle A, Christ H, Grund KE, Mackensen S. The role of the interstitium of the renal cortex in renal disease. *Contrib Nephrol.* 1979;16:109–114.
8. Schainuck LI, Striker GE, Cutler RE, Benditt EP. Structural-functional correlations in renal disease. II. The correlations. *Hum Pathol.* 1970;1(4):631–641.
9. Nankivell BJ, Borrows RJ, Fung CL, O'Connell PJ, Allen RD, Chapman JR. The natural history of chronic allograft nephropathy. *N Engl J Med.* 2003; 349(24):2326–2333.
10. El Zoghby ZM, et al. Identifying specific causes of kidney allograft loss. *Am J Transplant.* 2009;9(3):527–535.
11. Rudnicki M, et al. Hypoxia response and VEGF-A expression in human proximal tubular epithelial cells in stable and progressive renal disease. *Lab Invest.* 2009;89(3):337–346.
12. Sarwal M, et al. Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. *N Engl J Med.* 2003;349(2):125–138.
13. Flechner SM, et al. Kidney transplant rejection and tissue injury by gene profiling of biopsies and peripheral blood lymphocytes. *Am J Transplant.* 2004;4(9):1475–1489.
14. Mueller TF, et al. Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets. *Am J Transplant.* 2007; 7(12):2712–2722.
15. Einecke G, et al. Expression of CTL associated transcripts precedes the development of tubulitis in T-cell mediated kidney graft rejection. *Am J Transplant.* 2005;5(8):1827–1836.
16. Strehlau J, et al. Quantitative detection of immune activation transcripts as a diagnostic tool in kidney transplantation. *Proc Natl Acad Sci U S A.* 1997;94(2):695–700.
17. Smythe GK. Limma: linear models for microarray data. In Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. *Bioinformatics and Computational Biology Solutions using R and Bioconductor.* New York, New York, USA: Springer Publishing; 2005:398–420.
18. Reeve J, et al. Diagnosing rejection in renal transplants: a comparison of molecular- and histopathology-based approaches. *Am J Transplant.* 2009; 9(8):1802–1810.
19. Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal.* 2009;53(11):3735–3745.
20. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biology.* 2004;2(4):E108.
21. Igarashi P, Shao X, McNally BT, Hiesberger T. Roles of HNF-1beta in kidney development and congenital cystic diseases. *Kidney Int.* 2005;68(5):1944–1947.
22. Wight TN, Merrill MJ. Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circ Res.* 2004;94(9):1158–1167.
23. Kinsella MG, Bressler SL, Wight TN. The regulated synthesis of versican, decorin, and biglycan: extracellular matrix proteoglycans that influence cellular phenotype. *Crit Rev Eukaryot Gene Expr.* 2004; 14(3):203–234.
24. Wu Y, Siadaty MS, Berens ME, Hampton GM, Theodorescu D. Overlapping gene expression profiles of cell migration and tumor invasion in human bladder cancer identify metallothionein 1E and nicotinamide N-methyltransferase as novel regulators of cell migration. *Oncogene.* 2008;27(52):6679–6689.
25. Trevillian P, Paul H, Millar E, Hibberd A, Agrez MV. A(v)beta(6) integrin expression in diseased and transplanted kidneys. *Kidney Int.* 2004;66(4):1423–1433.
26. Wipff PJ, Hinz B. Integrins and the activation of latent transforming growth factor beta 1 - An intimate relationship. *Eur J Cell Biol.* 2008;87(8–9):601–615.
27. Bonventre JV. Kidney Injury Molecule-1 (KIM-1): a specific and sensitive biomarker of kidney injury. *Scand J Clin Lab Invest Suppl.* 2008;241:78–83.
28. Allanach K, et al. Comparing microarray versus RT-PCR assessment of renal allograft biopsies: similar performance despite different dynamic ranges. *Am J Transplant.* 2008;8(5):1006–1015.
29. Famulski KS, et al. Transcriptome analysis reveals heterogeneity in the injury response of kidney transplants. *Am J Transplant.* 2007;7(11):2483–2495.
30. Einecke G, Broderick G, Sis B, Halloran PF. Early loss of renal transcripts in kidney allografts: relationship to the development of histologic lesions and alloimmune effector mechanisms. *Am J Transplant.* 2007;7(5):1121–1130.
31. Bunnag S, et al. Molecular correlates of renal function in human kidney transplant biopsies. *J Am Soc Nephrol.* 2009;20(5):1149–1160.
32. Racusen LC, et al. The Banff 97 working classification of renal allograft pathology. *Kidney Int.* 1999; 55(2):713–723.
33. Solez K, et al. Banff 07 classification of renal allograft pathology: updates and future directions. *Am J Transplant.* 2008;8(4):753–760.
34. Famulski KS, et al. Changes in the transcriptome in allograft rejection: IFN-γ induced transcripts in mouse kidney allografts. *Am J Transplant.* 2006; 6(6):1342–1354.
35. Hidalgo LG, et al. The transcriptome of human cytotoxic T cells: measuring the burden of CTL-associated transcripts in human kidney transplants. *Am J Transplant.* 2008;8(3):637–646.
36. Hidalgo LG, Einecke G, Allanach K, Halloran PF. The transcriptome of human cytotoxic T cells: similarities and disparities among allostimulated CD4(+) CTL, CD8(+) CTL and NK cells. *Am J Transplant.* 2008;8(3):627–636.
37. Einecke G, Mengel M, Hidalgo LG, Allanach K, Famulski KS, Halloran PF. The early course of renal allograft rejection: Defining the time when rejection begins. *Am J Transplant.* 2009;9(3):483–493.
38. Famulski KS, Sis B, Billesberger L, Halloran PF. Interferon-gamma and donor MHC class I control alternative macrophage activation and activin expression in rejecting kidney allografts: a shift in the Th1-Th2 paradigm. *Am J Transplant.* 2008; 8(3):547–556.
39. Einecke G, et al. Expression of B cell and immunoglobulin transcripts is a feature of inflammation in late allografts. *Am J Transplant.* 2008;8(7):1434–1443.
40. Sis B, Jhangri G, Bunnag S, Allanach K, Kaplan B, Halloran PF. Endothelial gene expression in kidney transplants with alloantibody indicates antibody-mediated damage despite lack of C4d staining. *Am J Transplant.* 2009;9(10):2312–2323.
41. Scholtens D, von Heydebreck A. Analysis of differential gene expression studies. In: Gentleman R, Carey VJ, Huber W, Irizarry R, Dudoit S, eds. *Statistics for Biology and Health: Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* New York, New York, USA: Springer Publishing; 2005:229–248.
42. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A.* 2002;99(10):6567–6572.
43. Strimmer K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics.* 2008;24(12):1461–1462.