Henry Ford Health
Henry Ford Health Scholarly Commons

Surgery Articles | Surgery

12-1-2021

Progressive decline of function in renal allografts with normal 1-year biopsies: Gene expression studies fail to identify a classifier

Walter D. Park

Dean Kim
Henry Ford Health, DKIM3@hfhs.org

Martin :. Mai

Kunam S. Reddy

Thomas Gonwa

See next page for additional authors

Follow this and additional works at: https://scholarlycommons.henryford.com/surgery_articles

Recommended Citation

Park WD, Kim DY, Mai ML, Reddy KS, Gonwa T, Ryan MS, Herrera Hernandez LP, Smith ML, Geiger XJ, Turkevi-Nagy S, Cornell LD, Smith BH, Kremers WK, and Stegall MD. Progressive decline of function in renal allografts with normal one year biopsies: Gene expression studies fail to identify a classifier. Clin Transplant 2021.

This Article is brought to you for free and open access by the Surgery at Henry Ford Health Scholarly Commons. It has been accepted for inclusion in Surgery Articles by an authorized administrator of Henry Ford Health Scholarly Commons.

Authors

Walter D. Park, Dean Kim, Martin :. Mai, Kunam S. Reddy, Thomas Gonwa, Margaret S. Ryan, Loren P. Herrera Hernandez, Maxwell L. Smith, Xochiquetzal J. Geiger, Sandor Turkevi-Nagy, Lynn D. Cornell, Byron H. Smith, Walter K. Kremers, and Mark D. Stegall

This article is available at Henry Ford Health Scholarly Commons: https://scholarlycommons.henryford.com/surgery_articles/536

Check for updates

Received: 30 April 2021 | Revised: 23 July 2021 | Accepted: 4 August 2021

DOI: 10.1111/ctr.14456

Clinical TRANSPLANTATION
The Journal of Clinical and Translational Research
WILEY

ORIGINAL ARTICLE

Progressive decline of function in renal allografts with normal 1-year biopsies: Gene expression studies fail to identify a classifier

Walter D. Park[1] | Dean Y. Kim[2] | Martin L. Mai[3] | Kunam S. Reddy[4] | Thomas Gonwa[3] | Margaret S. Ryan[4] | Loren P. Herrera Hernandez[1] | Maxwell L. Smith[4] | Xochiquetzal J. Geiger[3] | Sandor Turkevi-Nagy[5] | Lynn D. Cornell[1] | Byron H. Smith[1] | Walter K. Kremers[1] | Mark D. Stegall[1]

[1] Mayo Clinic Rochester, Rochester, Minnesota, USA

[2] Henry Ford Hospital, Detroit, Michigan, USA

[3] Mayo Clinic Florida, Jacksonville, Florida, USA

[4] Mayo Clinic Arizona, Phoenix, Arizona, USA

[5] University of Szegad, Szeged, Hungary

Correspondence
Walter D. Park, William J. von Liebig Transplant Center, Mayo Clinic, 200, 1st Street SW, Rochester, MN 55905, USA.
Email: park.walter@mayo.edu

Funding information
National Institutes of Health and the National Institute of Allergy and Infectious Diseases, Grant/Award Number: U01 096326

Abstract

Histologic findings on 1-year biopsies such as inflammation with fibrosis and transplant glomerulopathy predict renal allograft loss by 5 years. However, almost half of the patients with graft loss have a 1-year biopsy that is either normal or has only interstitial fibrosis. The goal of this study was to determine if there was a gene expression profile in these relatively normal 1-year biopsies that predicted subsequent decline in renal function. Using transcriptome microarrays we measured intragraft mRNA levels in a retrospective Discovery cohort (170 patients with a normal/minimal fibrosis 1-year biopsy, 54 with progressive decline in function/graft loss and 116 with stable function) and developed a nested 10-fold cross-validated gene classifier that predicted progressive decline in renal function (positive predictive value = 38 ± 34%%; negative predictive value = 73 ± 30%, c-statistic = .59). In a prospective, multicenter Validation cohort (270 patients with Normal/Interstitial Fibrosis [IF]), the classifier had a 20% positive predictive value, 85% negative predictive value and .58 c-statistic. Importantly, the majority of patients with graft loss in the prospective study had 1-year biopsies scored as Normal or IF. We conclude predicting graft loss in many renal allograft recipients (i.e., those with a relatively normal 1-year biopsy and eGFR > 40) remains difficult.

KEYWORDS
genomics, glomerular filtration rate, kidney (allograft) function / dysfunction

1 | INTRODUCTION

Improving long-term renal allograft survival is a major unmet need. The ability to identify grafts at risk for failure would improve our ability to counsel patients and to conduct clinical intervention trials. We and others have shown that in kidneys that survive a year, we can predict 5-year graft loss using a combination of clinical factors and histologic findings on 1-year surveillance biopsy.[1–3] However, we also have shown that these predictive factors are rare (almost 90% of patients have a low-risk profile) and that approximately half of all graft losses at 5 years occur in patients with a low risk profile.[1]

© 2021 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.

Renal allografts without overt inflammation on light microscopy show signs of inflammation by gene expression when compared to native kidneys suggesting that all renal allografts are inflamed.[4] This is likely due to early ischemia/reperfusion responses and innate immunity. Therefore, the goal of this study was to determine if at 1 year, in grafts whose biopsy was Normal or had Interstitial Fibrosis (IF), a gene expression profile might predict significant decline in graft function by 5 years post-transplant.

2 | METHODS

2.1 | Study design and patients

With IRB approval, this study was conducted in two parts in which a retrospective Discovery Cohort was first developed followed by a prospective multicenter validation study. All work on the Discovery Cohort including "locking" the classifier genes was completed prior to the start of the prospective multicenter validation study. The inclusion criteria were recipients of adult conventional solitary renal transplants with a 1-year surveillance biopsy. Repeated measures of renal function from 1 to 5 years post-transplantation with data and specimens were collected as standard of care.

2.2 | Discovery cohort

This cohort developed retrospectively from patients at Mayo Clinic Rochester as follows: (1) Patients transplanted from 1/7/2000 to 12/29/2005 who underwent a 1-year protocol biopsy (n = 675[5]); (2) Local pathologist Banff scores were available for 535 cases and 458 (86%) were scored as cg = 0, i = 0 and ci≥0 (No central pathology review was performed); and (3) From these, 170 biopsies were selected across the continuum of 1–5 year function (equal distribution across 1–5 year eGFR slope quintiles) with a 1-year estimated GFR (eGFR) by Modification of Diet in Renal Disease (MDRD)[6] of > 40 mL/min.

Each subject had a 1-year eGFR calculated using the MDRD equation, which was considered the baseline. The stability of eGFR was calculated using the average of all eGFR measurements within each of the eight 6-month intervals from 1 to 5 years post-transplant (i.e., 1-1.5, 1.5-2.0, 2.0-2.5, etc.).

"Progressors" were defined using the following data: (1) 1-year estimated GFR of > 40 mL/min; and (2) progressive decline in eGFR between 1 and 5 years. For each subject all of the following criteria were necessary to determine progressive decline: a minimum follow-up of 4 intervals post-1-year; a slope of ← 6.1% per year (i.e., slope of decline of renal function is > 6.1%); a > 20% decline in eGFR from 1-year post-transplant to latest follow-up within 4 years of the 1-year biopsy and at least one eGFR (MDRD) interval < 60 mL/min. Non-Progressors are subjects who did not meet these criteria. The specific slope cutoff was determine using data from a prior study which examined sequential eGFR measurements in renal transplant recipients from 1 to 5 years. In that study subjects in the lowest slope quintile (cutoff of −6.1%) accounted for 69% of the allograft failures after 2.5 years post-transplant.[5]

Of the 170 subjects in the Discovery cohort, 54 met the Progressor criteria and 116 were Non-Progressors with follow-up for study purposes truncated at 5 years in 78% of subjects with a mean total follow-up of 76 months post-transplant at the time of study design. The characteristics (demographics, transplant, follow-up, etc.) between the Discovery Cohort Progressors and Non-Progressors are contained in Supplemental Table 1.

2.3 | Prospective multicenter study—The validation cohort

Based on the Discovery set gene expression data, an online validation study size calculator[7,8] was used to estimate that 280 subjects would be necessary to validate the Discovery data (210 Non-Progressor and 70 Progressors). Briefly, sample sizes were calculated like a test of non-inferiority where the mean accuracy of a classifier, evaluated on an external dataset, must be contained within a tolerance interval of .1 of the training accuracy. After accounting for the rates of all inclusion criteria (80% 1-year histology, 84% eGFR < 40 and 87% not lost to follow-up overall) in a large retrospective cohort,[5] the overall enrollment for the study was set at 480. We then prospectively enrolled a total of 491 subjects in the multicenter IRB-approved study (ClinicalTrials.gov number: NCT01782586; NIH Protocol number: Gen04) since a number of enrolled subjects did not receive a 1 year surveillance biopsy (Figure 1). The multicenter study includes transplants between 02/2012 and 02/2015 at 4 sites (Henry Ford Hospital, Detroit, MI; Mayo Clinic Arizona, Phoenix, AZ, USA; Mayo Clinic Florida, Jacksonville, FL, USA; and Mayo Clinic Rochester, Rochester, MN, USA). Patient follow-up ended 03/2018. All subjects underwent 1-year surveillance biopsy as standard of care and were included in the Validation cohort if: (1) they had a eGFR > 40 mL/min at 1-year; and (2) had a biopsy without transplant glomerulopathy (cg = 0) and/or inflammation (i = 0). The determination of Progressor/Non-Progressor was determined in the same manner as the Discovery cohort. Patients who were markedly non-adherent (in the opinion of their study team) in the first year were not included in the recruitment.

See *Supplemental Methods* for further details regarding cohort development.

2.4 | Pathology consensus scoring for the multicenter study

Slides from all 1-year biopsies from the multicenter study were scanned using a whole slide scanner (Aperio AT2, Leica Biosystems, Buffalo Grove, IL, USA). When the central and local pathologist agreed on the biopsy phenotype necessary for validation of the molecular classifier (cg = 0, i = 0, ci≥0) no additional central pathologist was used. If the results differed, then a second central pathologist reviewed the case to adjudicate the results as described in our prior publication.[9] To

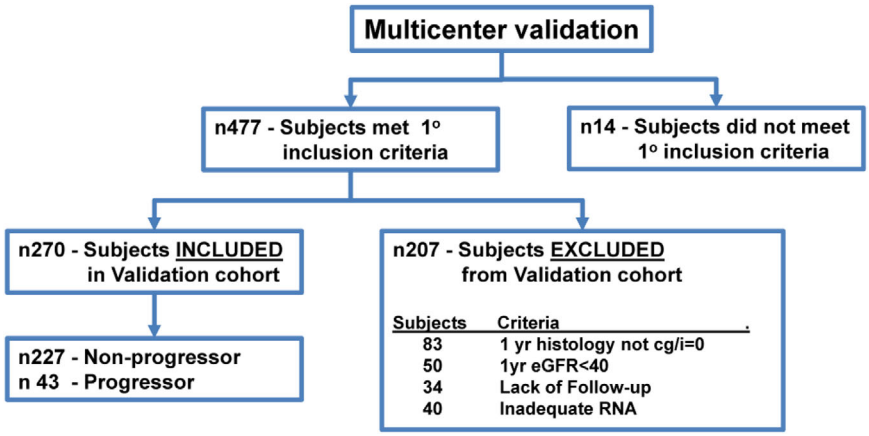


FIGURE 1 Validation cohort flowchart. The multicenter validation study contained 491 subjects, with 477 meeting primary inclusion criteria. Of those, 207 were excluded due to various criteria. The remaining 270 made up the Validation cohort with 227 subjects being Non-Progressors and were 43 Progressors

avoid center bias, the local and central pathologists were from different centers. For details see *Supplemental Methods*.

2.5 | RNA isolation, QC, and array core

RNA isolations were performed on the Discovery and multicenter study cohort biopsies as previously described[10–12]; *Supplemental Methods*). All specimens had a minimum of RNA Integrity Number > 6.9 and yield of 1 μg total RNA. Data quality, preprocessing, differential expression, and predictive modeling were conducted using the software R.[13] All array samples and data underwent multiple standard quality control assessments. The final microarray data were subjected to post-hybridization data processing that included quantile normalization[14] and plate correction. Microarray data are available on the GEO web site (https://www.ncbi.nlm.nih.gov/gds/; GEO: GSE181757).

2.6 | Differential expression

Both the Discovery and Validation cohorts underwent differential expression assessment between Progressors and Non-Progressors. These data were then overlaid onto previously annotated biologic pathways (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) and transplant related pathogenesis-based transcript sets.[15–17] The results were analyzed for over-representation by Fisher's exact test with Benjamini–Hochberg correction for multiple comparisons.[18]

2.7 | Classifier development and testing

The classifier was based upon the Discovery cohort and built using an elastic net penalty (90% L2 penalty and 10% L1 penalty) applied to a logistic regression setting to predict Progressor / Non-Progressor status. Only probes that had at least 50% detection in at least one of the Progressor groups were included. Ten-fold cross-validation was used to choose the tuning parameters. To calculate an unbiased estimate of error on a validation sample, a nested cross validation procedure was implemented. In this way, tuning parameters are identified using an internal 10-fold cross validation within each fold of an external 10-fold cross validation. The Validation cohort specimens were normalized to the same distribution that was used in the Discovery cohort calculating risk scores from the logistic model derived in from the Discovery cohort. The area under the receiver operating characteristic curve (AUC) was calculated based on this probability value. Youden's J Index was used to establish a threshold probability score for classification in to either Progressor or Non-Progressor after which the classifier assessed using positive predictive value (PPV), and negative predictive value (NPV). Detailed methods of classifier development are presented in the Supplemental Methods.

3 | RESULTS

3.1 | Discovery cohort differential expression

The Discovery cohort was developed retrospectively from patients at Mayo Clinic Rochester and included 54 patients with progressive decline in function/graft loss and 116 with stable function. After RNA extraction and hybridization, a total of 24 678 probes were considered detectable by at least 50% of the samples in the Progressor or Non-Progressor groups by the Illumina software. Using this set of probes, we performed a differential gene expression analysis for the Discovery cohort. A total of 1884 unique transcripts were identified as significantly altered between Progressor and Non-Progressors. To examine the pathways of transcripts that were significantly altered, a GeneSet Enrichment Analysis was performed using both Ingenuity Pathway and Pathogenesis-based transcript (PBT) gene lists. In the Discovery cohort dataset, the majority of the significantly enriched Ingenuity Pathways were related to pro-inflammatory processes such as antigen presentation, allograft rejection related to immune cells (T-cells, dendritic cells, natural killer cells, etc.) (Table 1). Similar findings were observed using PBTs with the significant enrichment of multiple immune-cell (i.e., T-cell and macrophage associated transcripts) and transplant-related PBTs. This included the PBT of genes positively associated with graft failure and genes negatively correlated with eGFR.

TABLE 1 Differential gene expression

		Discovery data			
	Genes per pathway	Altered genes per pathway	Ratio of altered expression	% Upregulated	*P*
Ingenuity Pathways					
Antigen Presentation Pathway	32	20	63%	100%	.0000
Allograft Rejection Signaling	31	19	61%	100%	.0000
Crosstalk between Dendritic Cells and Natural Killer Cells	62	29	47%	100%	.0000
Communication between Innate and Adaptive Immune Cells	50	22	44%	100%	.0009
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	37	18	49%	94%	.0009
Type I Diabetes Mellitus Signaling	94	34	36%	82%	.0009
Graft-versus-Host Disease Signaling	32	16	50%	100%	.0013
Dendritic Cell Maturation	114	38	33%	87%	.0017
TREM1 Signaling	43	19	44%	89%	.0017
CD28 Signaling in T Helper Cells	98	31	32%	87%	.0182
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	52	19	37%	100%	.0257
Pathogenesis Based Transcripts					
Immune cell-related					
CTL-associated transcripts (mouse)	209	66	37.5%	100%	.0000
CTL-associated transcripts 2 (human)	363	77	25.7%	94%	.0121
γ-IFN and rejection induced transcripts 1 (mouse)	29	15	55.6%	100%	.0005
γ-IFN and rejection induced transcripts 2 (mouse)	142	38	30.9%	95%	.0073
Macrophage-Associated Transcripts (human)	60	19	38.0%	95%	.0082
Quantitative CTL-Associated Transcript Set (human)	22	9	47.4%	100%	.0193
Transplant-related					
DSA Specific Transcripts	20	11	57.9%	100%	.0018
Transcripts correlated with eGFR decline (human)	125	33	28.0%	94%	.0414
Genes+ assoc with graft failure (human)	572	136	25.6%	99%	.0007
Human Antibody Mediated Rejection (Classifier)	19	9	50.0%	100%	.0137

Gene Set Enrichment Analysis using Ingenuity Pathway Analysis and transplant related Pathogenesis-Based Transcripts was performed using the Discovery cohort dataset. Significant enrichment ($P < .05$ by Benjamini-Hochberg corrected Fisher-Exact Test) is highlighted and bold.

3.2 | Classifier development

Initial classifier development revealed sensitivity and specificity of 78–80%. We hypothesized that the heterogeneous nature of the patients included in the study may be masking the true signal related to "Progression." Therefore, we chose to re-analyze the data after removal of patients with several criteria considered likely to influence intragraft expression and long-term renal function/survival. This included development of BK+ (n = 4), recurrent disease (n = 10), or DSA (n = 2) between transplant and 1-year biopsy or existence of pre-txp DSA (n = 15). A total of 31 patients/samples were removed from the Illumina dataset leaving 101 Non-Progressors and 38 Progressors used in the development of the molecular classifier (aka Restricted Discovery cohort).

The locked molecular classifier included 10 probes identified through the use of elastic net logistic regression using microarray data from the Discovery cohort (see Figure 2A). The Progressor-Prediction-Score for each participant is obtained by (1) multiplying the model coefficients for each probe by the log2-normalized probe expression values, (2) adding all of these products, (3) adding a value of 3.924532, and finally taking the logistic transform of this value. Model values ≥.7035 categorize Non-Progressors and values < .7035 categorize Progressors. This cutoff was chosen as it is the maximum sum of the sensitivity and specificity.

$$x = 3.924532 + \sum_{i=1}^{10} \text{Coefficient}_i \times \text{Probe}_i \quad y = \frac{1}{1+e^{-x}}$$

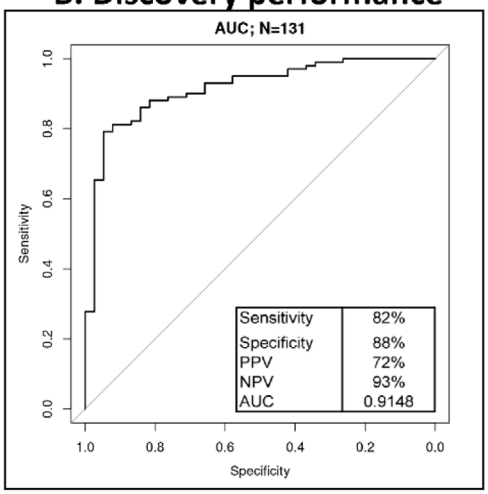
$$= \text{Probability of being a Progressor}$$

Using the model, the restricted Discovery cohort performance had an AUC of .9148, sensitivity (82%), specificity (88%), PPV (72%), and NPV (93%) (See Figure 2B). Based on the 10-fold cross-validation the mean AUC ± standard error is .5903 ± .0284 (sensitivity = 50% ± 39%, specificity = 62% ± 30%, PPV = 38% ± 34%, and NPV = 73% ± 30%)

A. Classifier Probes

	Symbol	Entrez Gene ID	Progressor Average	Progressor Stdev	Progressor % Detected	Non-Progressor Average	Non-Progressor Stdev	Non-Progressor % Detected	p-value
1	C18orf45	85019	6.78	0.21	94%	6.67	0.19	84%	0.0025
2	C5orf42	65250	6.47	0.14	50%	6.41	0.12	33%	0.0123
3	ELP3	55140	8.95	0.19	100%	9.07	0.19	100%	0.0003
4	KPNA2	3838	7.07	0.18	100%	6.94	0.17	100%	<.0001
5	NKRF	55922	7.77	0.14	100%	7.72	0.11	100%	0.0173
6	PDDC1	347862	6.78	0.10	98%	6.84	0.12	99%	0.0004
7	PHKA1	5255	7.01	0.13	100%	7.08	0.12	100%	0.0023
8	PPTC7	160760	8.11	0.23	100%	8.28	0.25	100%	<.0001
9	SLC25A28	81894	9.73	0.17	100%	9.82	0.17	100%	0.0021
10	TEX261	113419	7.46	0.19	100%	7.55	0.18	100%	0.0018

B. Discovery performance



C. Nested 10-fold Cross-validation performance

Sensitivity	50%
Specificity	62%
PPV	38%
NPV	73%
AUC	0.5903

FIGURE 2 Classifier development and performance. The list of 10 transcripts belonging to the molecular classifier are included in (A). For each transcript the average normalized expression value and percentage of samples considered to have detectable expression by the Illumina software is provided for the Progressor and Non-Progressor groups. (B) summarizes the performance of the Restricted Discovery Cohort with an Area Under the Curve (AUC) plot and other model performance metrics (sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV])

(Figure 2C). Importantly, all work on the Discovery Cohort and initial classifier development was completed prior to the prospective enrollment and gene expression testing of the Validation Cohort. That is the classifier was "locked" prior to testing samples from the Validation cohort.

3.3 | Validation cohort

The Validation cohort was developed from the four-site multicenter study (491). The subjects were followed from enrollment at 1-year surveillance visit until an average follow-up of 46 months post-transplant with a total of 18 confirmed graft losses, 26 patient deaths with function, and five subjects considered lost to follow-up (see Supplemental Table 2). As of March 2018, 428 grafts (89.7%) were still active (Supplemental Table 2). To determine 1-year histology, we used an adjudication scheme to score the biopsies. For the multicenter study, 19 biopsies originally read as Normal / IF by the local pathologist were "reclassified" as abnormal and 22 abnormal biopsies were reclassified as Normal / IF (Table 2). When histology and other exclusion criteria were assessed, a total of 207 multicenter subjects were excluded (83 had central 1- year histology as not Normal/IF; 50 had an eGFR < 40 mL/min at 1-year; 34 were lost to follow up and 40 had inadequate RNA). The remaining 270 constitute the Validation Cohort.

3.4 | Differences between the discovery and validation cohorts

Compared to Discovery Cohort, the Validation Cohort included more non-white recipients (20.4% vs 6.5%), was older at the time of transplantation (43.1 vs 39.8 years), lower rates of Thymoglobulin induction use (39.7% vs 85.3%), higher rates of steroid-free immunosuppression (28.1% vs 5.3%), fewer eGFR measurements (27.2 vs 33.9), and a higher percentage of biopsies with IF (48.9% vs 37.1%) (Table 3). Although the study phenotypes involved clinical data between 1 and 5 years post-transplant, the mean total follow-up was shorter in the Validation cohort (47 vs 77 months post-transplant). The percentage of patients defined as Progressors was lower in the Validation cohort (13.8% vs 31.8%), but the overall graft survival in these two cohorts was similar by Kaplan Meier (p = 0.7). None of the extended criteria variables were significantly different among the cohorts, including evdience of BK nephropathy, recurrent disease, or HLA Class I/II DSA from the time of transplant to 1 year.

3.5 | Characteristics of progressors

Many transplant and pre-1 year post-transplant clinical parameters were explored to determine the differences between Progressors and Non-Progressors in the Validation Cohort (Table 3). Only recipient gender (more females) and higher donor age were significantly associated with the Progressors, illustrating how difficult it would be to create a predictive model for progression using clinical parameters alone. In addition, the study database included data obtained from the clinic notes of every subject from 1 to 5 years (or last follow-up). The results show that Progressors are significantly more likely to have at least one > 1-year biopsy-proven acute cellular rejection (33% vs 10%, $P < .0001$), > 1-year recurrent disease (14% vs 4%, $P < .0150$), or > 1-year pyelonephritis (7% vs 2%, $P < .0485$) (Table 4). Although 67% of Progressors did experience at least one post-1 year complication, 33%

TABLE 2 Consensus histology

	1-year histology: Local vs Consensus							
		CONSENSUS Banff						
	Study Group	cg > 0	Normal / IF	IF+I	Inflamm	Missing: No Bx Scan	Missing: No gloms	Mixed	Total
LOCAL Banff	cg > 0	7	0	1	0			1	9
		78%	0%	11%	0%			11%	
	Normal / IF	2	**372**	7	2	1	1	6	**391**
		1%	95%	2%	1%	0%	0%	2%	
	IF+i	1	16	**39**	0			3	**59**
		2%	27%	66%	0%			5%	
	Inflamm	0	4	4	5			2	15
		0%	27%	27%	33%			13%	
	MISSING	**0**	**2**	**0**	**0**		**1**	**0**	3
		0%	67%	0%	0%		33%	0%	
		10	**394**	**51**	**7**	**1**	**2**	**12**	477

Each biopsy included in the multicenter study underwent additional Banff scoring until consensus was reached. Each biopsy was placed into a study group based on Banff scores: cg > 0, Normal/IF (cg & i = 0, ci > = 0), IF+I (cg = 0, ci & i > 0), Inflamm (cg & ci = 0, i > 0). A team of Central Pathologists re-scored using whole slide images of all original clinical slides created for light microscopy. If after one review the same Study Group was obtained, no additional review was performed. If not, then an additional Central Pathologist reviewed the case such that consensus was reached among at least two different pathologists for every case.

had a decline in renal function that could not be attributed to a known complication. Further, 50% of Non-Progressors had one or more post-1 year complication that did not result in a persistent decline in renal function.

3.6 | Classifier testing: the validation set

After normalization to the same quantile distribution as the Discovery dataset, the Progressor classifier was calculated for all qualifying subjects in the Validation cohort, including 43 Progressors and 227 Non-Progressors. The AUC of the entire Validation cohort (n270) was .5607 [95% CI: .4673–.6541). Due to the performance of the classifier, we performed a subset analysis involving subjects that met the same criteria used in the Discovery dataset. This Restricted Validation cohort included 223 subjects and has an AUC of .5821 (95% CI: .4837–.6804; Figure 3B) and lower model performance metrics when compared to the Discovery cohort (Figure 2BC). These metrics did not differ appreciably among the sites (Figure 3C).

A comprehensive list of clinical and molecular characteristics was compared between the specimens correctly and incorrectly identified by the classifier. The poorest performance appeared to be related to recipient race, type of transplant, and various reasons for transplant (see *Supplementary results*).

3.7 | Post-study follow-up

The multicenter validation study ended in 03/2018 and the dataset locked. The mean follow-up at this time point was 3.9 ± .8 years and only 20% of the Validation cohort had eGFR measures into the fifth year post-transplant compared to 78% in the Discovery cohort. The potential for this difference to impact the designations used for the primary endpoint (Progressor/Non-Progressors) caused the study team to seek additional follow-up and eGFR data from all subjects up to the fifth-year post-transplant interval. Thus, we performed post-study follow-up in October of 2019. This resulted in a mean follow-up of 4.7 ± .6 years with 69% of the Validation cohort having data in the fifth-year eGFR interval. At this later time point, six additional allografts reached failure (24 up from 18) with overall causes varying from recurrent glomerulonephritis, infections, and rejections (both antibody and cell mediated). The majority of these graft losses (58%) were in subjects with Normal / IF 1-year histology as we have previously described.[11,19] Most Validation cohort subjects had the same functional status (40 Progressor and 218 Non-Progressors), with nine Non-Progressors becoming Progressors, three Progressors becoming Non-Progressors, and seven subjects being added as sufficient data was now available to categorize them (five Non-Progressors and two Progressors). Unfortunately, despite the longer follow-up, the analyses did not improve the AUC for the Validation cohort (.5861).

4 | DISCUSSION

This study agreed with prior studies by showing that the majority of renal allografts with progressive decline in function or graft loss between 1 and 5 years after transplantation have a relatively normal surveillance biopsy at 1 year. In a retrospective, single-center Discovery cohort, we were able to identify a 10-gene classifier set that

TABLE 3 Comparisons between patient cohorts

		Discovery vs Validation			Validation phenotypes		
		Discovery	Validation		Non-Progressor	Progressor	
Type	Variable	170	270	*P*	227	43	*P*
PROGRESSOR							
	# Progressors	54 (31.8%)	43 (15.9%)				
	Henry Ford	–	2 (8.7%)			2 (8.7%)	
	Mayo Arizona	–	6 (14.6%)			6 (14.6%)	
	Mayo Florida	–	15 (24.6%)			15 (24.6%)	
	Mayo Rochester	54 (31.8%)	20 (13.8%)			20 (13.8%)	
RECIPIENT	Recip_Age	51.198 (12.683)	53.360 (13.092)	.089	53.102 (12.739)	54.725 (14.913)	.457
	Recip_Gender (%F)	70 (41.2%)	110 (40.7%)	.928	85 (37.4%)	25 (58.1%)	.011
	Recip_Ethicity (Hispanic)	0	15 (5.6%)	–	13 (5.7%)	2 (4.7%)	.911
	Recip_Race			< .001			.658
	Caucasian	159 (93.5%)	215 (79.6%)		181 (79.7%)	34 (79.1%)	
	African American	1 (.6%)	40 (14.8%)		33 (14.5%)	7 (16.3%)	
DONOR	Donor_Age	39.805 (12.078)	43.067 (13.524)	.011	42.053 (13.522)	48.419 (12.356)	.004
	Dono_Gender (%F)	75 (44.1%)	132 (48.9%)	.329	111 (48.9%)	21 (48.8%)	.994
	Donor_Ethicity (Hispanic)	0	11 (4.1%)		181 (79.7%)	37 (86.0%)	.315
	Donor_Race			.004			.640
	Caucasian	57 (96.6%)	235 (87.0%)		198 (87.2%)	37 (86.0%)	
	African American	0 (.0%)	20 (7.4%)		15 (6.6%)	5 (11.6%)	
TRANSPLANT	Txp_Type (Deceased)	31 (18.2%)	100 (37.0%)	< .001	81 (35.7%)	19 (44.2%)	.290
	Expanded criteria donors	0	16 (16.0%)		12 (14.8%)	4 (21.1%)	.504
	Total_Numb_Prior_Kid_Txps	.165 (.403)	0.167 (0.501)	.966	.167 (.487)	.163 (.574)	.956
	Total_Numb_Prior_NonKid_Txps	.100 (.320)	0.067 (0.278)	.249	.062 (.241)	.093 (.426)	.499
	InductionType			<.001			.300
	Campath	5 (2.9)	87 (32.6)		76 (33.8%)	11 (26.2%)	
	Simulect	15 (8.8)	72 (27)		55 (24.4%)	17 (40.5%)	
	Thymo	145 (85.3%)	106 (39.7%)		92 (40.9%)	14 (33.3%)	
	Thymo+IVIG	0 (.0%)	1 (.4%)		1 (.4%)	0 (.0%)	
	Thymo+Simulect	0 (.0%)	1 (.4%)		1 (.4%)	0 (.0%)	
	Immuno			<.001			.973
	Prograf-MMF	9 (5.3%)	76 (28.1%)		65 (28.6%)	11 (25.6%)	
	Prograf-MMF-Pred	133 (78.2%)	159 (58.9%)		132 (58.1%)	27 (62.8%)	
FOLLOW-UP	GraftStatus_POD	2316.40 (621.24)	1439.02 (279.63)	<.001	1436.06 (282.01)	1454.66 (269.36)	.690
	GraftStatus_POM	77.25 (20.68)	47.97 (9.33)	<.001	47.87 (9.41)	48.49 (8.96)	.692
	GraftStatus_FINAL			<.001			<.001
	Active	124 (72.9%)	257 (95.2%)		220 (96.9%)	37 (86.0%)	
	DWF	18 (10.6%)	7 (2.6%)		6 (2.6%)	1 (2.3%)	
	Failed	14 (8.2%)	5 (1.9%)		0 (.0%)	5 (11.6%)	
	Lost to Follow up	14 (8.2%)	1 (.4%)		1 (.4%)	0 (.0%)	
FUNCTION	MDRD_Intervals	8.341 (1.066)	7.078 (1.390)	<.001	7.101 (1.390)	6.953 (1.396)	.523
	MDRD_Measures	33.929 (22.037)	27.233 (20.116)	.001	24.934 (17.230)	39.372 (28.544)	<.001
	MDRD_1_1yrBx_-30_to_30d	57.917 (10.495)	59.444 (15.986)	.270	59.754 (16.552)	57.806 (12.618)	.465

(Continues)

TABLE 3 (Continued)

		Discovery vs Validation			Validation phenotypes		
		Discovery	Validation		Non-Progressor	Progressor	
Type	Variable	170	270	*P*	227	43	*P*
HISTOLOGY				.015			.254
	Interstitial Fibrosis (cg = 0, i = 0, ci > 0)	63 (37.1%)	132 (48.9%)		70 (31.4%)	19 (44.2%)	
	Normal (cg = 0, i = 0, ci = 0)	107 (62.9%)	138 (51.1%)		107 (48.0%)	16 (37.2%)	
EXTENDED CRITERIA							
	Txp-1 yr_BK_nephropathy	4 (2.4%)	10 (3.7%)	.432	8 (3.5%)	2 (4.7%)	.720
	Txp-1 yr_RecurrentDisease	10 (5.9%)	12 (4.4%)	.500	11 (4.8%)	1 (2.3%)	.462
	Pre-transplant DSA (Class I/II)	17 (16.5%)	33 (12.4%)	.091			
	Pre-transplant DSA (Class I/II; NV > 1K)	–	10 (3.7%)		10 (4.5%)	0 (.0%)	.163
	Txp-1 yr DSA (Class I/II)	2 (6.2%)	31 (12.1%)	.189			
	Txp-1 yr DSA (Class I/II; NV > 1K)	–	16 (6.2%)		12 (5.6%)	4 (9.8%)	.297

The demographics and follow-up is presented for the Discovery (n170) and Validation (n270) cohorts and the Validation cohort Progressor and Non-Progressors. For each variable, a statistical test was performed to determine if significant variation was observed among the cohorts. Variables bold/highlighted are significant ($P < .05$).

TABLE 4 Complications post-1 year

Documented Complications post-1 year			
Complication	Non-Progressor (n227)	Progressor (n43)	*P*
Biopsy-proven acute cellular rejection	10% (n23)	33% (n14)	***.0001***
Biopsy-proven antibody mediated rejection	2% (n0)	0% (n0)	.3805
Recurrent disease	4% (n10)	14% (n6)	***.0150***
Infection			
BK viremia	7% (n16)	0% (n0)	.0727
BK nephropathy	1% (n2)	5% (n2)	.0606
Cytomegalovirus	4% (n9)	5% (n2)	.8346
Pyelonephritis	2% (n4)	7% (n3)	***.0485***
Urinary Tract Infection	24% (n55)	35% (n15)	.1438
Malignancy	18% (n40)	7% (n3)	.0803
Any of the above	50% (n114)	67% (n29)	***.03800***

Post-1 year complications were obtained from the clinical notes during the study follow-up on all subjects. The rate of each complication was compared between the Progressors and Non-Progressors by chi-square test (Pearson). A *P*-value of < .05 was considered statistically significant (bold).

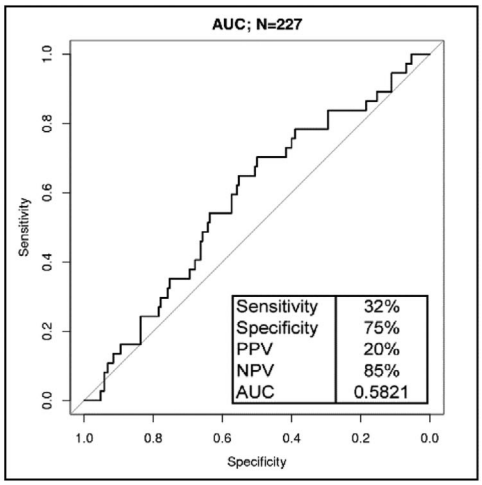
moderately predicted progressive decline in renal function. However, the classifier performed poorly in a large, multicenter Validation cohort.

While the results suggest lack of molecular classifier validation, the study does offer several important observations. It is often considered difficult, to enroll and monitor transplant patients for years following transplant, particularly after the third year when CMS reporting requirements stop. However, this study was fully enrolled within the expected timeframe, the rate of "lost to follow-up" was 2.1% and the number of serum creatinine lab tests during the study period averaged 27 per patient, despite some clear difference in how patients were followed at each center (range MCA 18 to Henry Ford 40). This suggests that it is possible to monitor transplant patients for years post-transplantation, as long as study design takes into account the standard practice of the Centers being included.

A. Classifier probes

	Symbol	Entrez Gene ID	Progressor Average	Progressor Stdev	Progressor % Detected	Non-Progressor Average	Non-Progressor Stdev	Non-Progressor % Detected	p-value
1	C18orf45	85019	6.72	0.24	95%	6.69	0.22	81%	0.5136
2	C5orf42	65250	6.43	0.14	35%	6.44	0.13	49%	0.9692
3	ELP3	55140	9.05	0.17	100%	9.05	0.18	100%	0.9730
4	KPNA2	3838	7.00	0.17	100%	6.96	0.18	96%	0.1448
5	NKRF	55922	7.72	0.12	100%	7.74	0.14	100%	0.4395
6	PDDC1	347862	6.83	0.12	95%	6.83	0.15	93%	0.9020
7	PHKA1	5255	7.03	0.17	100%	7.07	0.15	100%	0.1817
8	PPTC7	160760	8.23	0.22	100%	8.23	0.19	100%	0.9452
9	SLC25A28	81894	9.81	0.15	100%	9.79	0.14	100%	0.4401
10	TEX261	113419	7.52	0.16	100%	7.52	0.17	100%	0.8570

B. Model performance



C. Model performance metrics by site

	Henry Ford	Mayo Arizona	Mayo Florida	Mayo Rochester
Sensitivity	50%	40%	33%	28%
Specificity	78%	74%	76%	75%
PPV	20%	20%	31%	16%
NPV	93%	88%	78%	86%
AUC	0.5714	0.5000	0.5406	0.6100

FIGURE 3 Classifier validation. Differential expression analyses revealed that none of the 10 transcripts belonging to the classifier were significantly altered between Progressors and Non-Progressors in the Validation cohort (Panel A). For each classifier transcript the average normalized expression value and percentage of samples considered to have detectable expression by the Illumina software is provided for the Progressor and Non-Progressor groups. The overall model performance (sensitivity, specificity, PPV [Positive Predictive Value], NPV [Negative Predictive Value], and AUC (Area Under the Curve) for the Restricted Validation cohort is provided in Panel B. Panel C includes the model performance metrics by study site

Using clinical data, we have confirmed the original observation that eGFR progression from 1 to 5 years post-transplant does occur in grafts with relatively normal histology[19] and that these losses are most often preceded by progressive decline in renal function.[5] Our group has repeatedly shown that various Banff scores are associated with subsequent decreased graft survival such as transplant glomerulopathy and interstitial fibrosis + inflammation.[11,19] However, these features of overt injury are observed in only 15–20% of all 1-year protocol biopsies at Mayo Clinic Rochester. The majority of 1-year biopsies are considered normal or mild fibrosis and although the graft survival rate at 5 years exceeds 90%, the absolute number of graft losses by 5 years is similar to those with underlying pathology (transplant glomerulopathy, etc.).[11,19] These findings were verified at the three other participating centers, where the rate of Normal+IF ranged from 75% to 89% by consensus pathology and the majority of graft losses during study follow-up had this histology. This was despite clear differences in the demographics and transplant types among the centers (Supplemental Table 1). While the rate of "Progressor" was lower in the Validation study, no graft losses had a stable function phenotype. This suggests that a combined approach of histology and functional measurement is suitable for identifying subjects at highest risk for subsequent graft loss.

Our group has recently published that clinical factors alone can be used to assess risk at 1-year for subsequent graft loss by 5 years.[1] The "BirMayo" risk score is based on the use of demographic and 1-year laboratory information (eGFR, 24-hour urine protein, etc.) and histology (g and ci Banff scores). The score was found to be associated with future graft loss with good model performance (Concordance of .90). Despite this, for subjects with sufficient data to perform the calculation, BirMayo did not separate the Progressors and Non-Progressors in either the Discovery or Validation cohorts (*Data not shown*). This is likely because the study design omits grafts with many characteristics known to be associated with high graft loss risk such as low 1-year eGFR, poor histology, etc.

Another key concept identified in these studies is that despite the lack of overt histology on the 1-year biopsies, future progressive decline is associated with altered pro-inflammatory transcripts. The Progressors exhibited the enrichment of molecular pathways and genesets associated with inflammation despite having normal or fibrosis Banff scores. This included the Pathogenesis Based Transcript sets *Transcripts correlated with eGFR decline and Genes positively associated with graft failure* which were developed using for cause renal allograft biopsies and are associated with function and future graft failure.[16,17] This finding suggests that even among grafts with an excellent 5-year outcome, a subset do exhibit intragraft molecular changes at 1 year that are indicative of molecular inflammation.

A major challenge to any predictive classifier is that multiple clinical events which may influence the outcome can occur after the specimen collection/testing. All types of predictive modeling face this challenge. Florid non-adherence was rare in both cohorts and any patient lost to follow-up was not included in the Validation cohort. Thus, non-adherence after 1 year did not appear to be a major factor in progression. In our study, post 1-year complications were common in both Progressors (67%) and Non-Progressors (50%). While some complications may not impact the graft function (i.e., asymptomatic UTIs, skin malignancies), others like biopsy-proven acute rejection and

recurrent primary disease could switch a subject with a normal 1-year molecular signature to become a Progressor. The underlying causes of these complications are also likely multi-factorial (lifestyle choices, severity of primary disease, impact of chronic immunosuppression, non-adherence, etc.) and are beyond the scope of this report.

In addition to the time from testing/prediction to the endpoint, there are several potential explanations for the lack of validation of the classifier. The Discovery cohort was established using a retrospective cohort from a single center with a focus on following transplant recipients long-term. This meant that the majority of subjects had follow-up (graft status, labs, etc.) at 5 years and the population was more homogenous than the Validation cohort. This included such characteristics as recipient and donor race, type of transplant, induction, immunosuppression, etc. Several of these variables were associated with lower classifier accuracy (see Supplemental Table 3). For example, only 58% of the Non-Caucasian recipients were correctly predicted by the classifier, whereas Caucasians which made up 93.8% of the Discovery cohort were accurately predicted in 71% by the classifier. Taken together, it seems likely that the original classifier was created with clinical characteristics not likely represented sufficiently in the Discovery cohort and that limited the applicability of the classifier to the Validation cohort involving more diverse clinical phenotypes.

Another potential source of variability that was considered was the size of biopsy obtained for gene expression purposes. Banff provides specific guidance regarding the adequacy of tissue necessary to perform a diagnostic interpretation of a biopsy, which includes the sampling of at least seven glomeruli.[20,21] For studies involving morphometry our group has used a glomerular count cutoff of 4 and a cortex area cutoff of 2 mm^2 per section.[22] In this study, RNA yield and quality was used to assess adequacy and not size as we were dependent on the tissue supplied from each center that was collected as standard of care. For each biopsy, a pre-isolation visual assessment of the biopsy was recorded and larger biopsies had only a slightly better performance in the model (69% correct) than the smaller, less likely to be diagnostic, biopsies (64%) (Supplemental Table 4). This suggests that in smaller tissue pieces molecular testing may be sufficiently sensitive to avoid some of the sampling biases observed by histopathology or morphometry.

In the field of kidney transplantation, a number of molecular classifiers have been created.[23–26] Typically, these classifiers are designed for diagnosing acute events (i.e., rejection), when the transcriptome either in the graft, or circulating in the blood, is most perturbed. An example of predictive classifiers involving surveillance biopsies would be the GoCAR study.[27] Intragraft expression was measured in a month 3 surveillance biopsy to predict the development of a Chronic Allograft Damage Index score ≥ 2 at 12 months. None of the 13 GoCAR classifier transcripts in the final classifier were found to be significant by differential expression for either our Discovery or Validation cohorts. This could be because rather than predict a histologic endpoint, this study attempted to use the transcriptome of essentially normal biopsies to predict future function. While differential expression and enrichment analysis both showed evidence of increased pro-inflammatory transcripts in the Progressors, it does not appear that the magnitude of those changes is large/consistent enough to distinguish Progressors from Non-Progressors using a molecular classifier across multiple datasets. Future efforts could include expanding and refining the Discovery cohort to include a more comprehensive cohort that addresses several of the issues identified in the results of this study. Such changes should improve the feature selection used for the classifier and thereby improve the likelihood of future validation of a prognostic classifier for this important group of renal transplant patients.

5 | CONCLUSION

This study agrees with prior studies showing that approximately half of all renal allografts that experience progressive graft dysfunction or graft loss between 1 and 5 years after kidney transplantation have relatively normal 1-year surveillance biopsies. The current study also suggests that while enrichment of molecular inflammation may exist in these Progressors, a gene expression signature was not validated. Thus, predicting outcomes in this cohort remains difficult.

ACKNOWLEDGEMENTS

This study was supported with funding from the National Institutes of Health and the National Institute of Allergy and Infectious Diseases (U01 096326). This work was partially supported by the Renal Pathology Society's Research Collaborator Program (S T-N) and the content of this manuscript is the sole responsibility of the authors.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The clinical trial data is available on Immport, which is an NIH data repository for data sharing purposes (https://immport.niaid.nih.gov/home). Microarray data are available on the GEO web site (https://www.ncbi.nlm.nih.gov/gds/; GEO: GSE181757).

ORCID

Walter D. Park https://orcid.org/0000-0002-0536-9515
Byron H. Smith https://orcid.org/0000-0001-9318-2437
Mark D. Stegall https://orcid.org/0000-0002-1392-0792

REFERENCES

1. Gonzales MM, Bentall A, Kremers WK, Stegall MD, Borrows R. Predicting individual renal allograft outcomes using risk models with 1-year surveillance biopsy and alloantibody data. *J Am Soc Nephrol*. 2016;27(10):3165-3174.
2. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ*. 2019;366:l4923.
3. Bentall A, Smith BH, Gonzales MM, et al. Modeling graft loss in patients with donor-specific antibody at baseline using the Birmingham-Mayo (BirMay) predictor: implications for clinical trials. *Am J Transplant*. 2019;19(8):2274-2283.
4. Park W, Griffin M, Grande JP, Cosio F, Stegall MD. Molecular evidence of injury and inflammation in normal and fibrotic renal allografts one year posttransplant. *Transplantation*. 2007;83(11):1466-1476.

5. Park WD, Larson TS, Griffin MD, Stegall MD. Identification and characterization of kidney transplants with good glomerular filtration rate at 1 year but subsequent progressive loss of renal function. *Transplantation*. 2012;94(9):931-939.
6. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145(4):247-254.
7. Dobbin KK, Zhao Y, Simon RM. How large a training set is needed to develop a classifier for microarray data?. *Clin Cancer Res*. 2008;14(1):108-114.
8. Dobbin KK, Simon RM. Sample size planning for developing classifiers using high-dimensional DNA microarray data. *Biostatistics*. 2007;8(1):101-117.
9. Smith B, Cornell LD, Smith M, et al. A method to reduce variability in scoring antibody-mediated rejection in renal allografts: implications for clinical trials - a retrospective study. *Transpl Int*. 2019;32(2):173-183.
10. Taner T, Park WD, Stegall MD. Unique molecular changes in kidney allografts after simultaneous liver-kidney compared with solitary kidney transplantation. *Kidney Int*. 2017;91(5):1193-1202.
11. Park WD, Griffin MD, Cornell LD, Cosio FG, Stegall MD. Fibrosis with inflammation at one year predicts transplant functional decline. *J Am Soc Nephrol*. 2010;21(11):1987-1997.
12. Dean PG, Park WD, Cornell LD, Gloor JM, Stegall MD. Intragraft gene expression in positive crossmatch kidney allografts: ongoing inflammation mediates chronic antibody-mediated injury. *Am J Transplant*. 2012;12(6):1551-1563.
13. Team RC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016. http://www.R-project.org/
14. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-193.
15. Mueller TF, Einecke G, Reeve J, et al. Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets. *Am J Transplant*. 2007;7(12):2712-2722.
16. Bunnag S, Einecke G, Reeve J, et al. Molecular correlates of renal function in kidney transplant biopsies. *J Am Soc Nephrol*. 2009;20(5):1149-1160.
17. Einecke G, Reeve J, Sis B, et al. A molecular classifier for predicting future graft loss in late kidney transplant biopsies. *J Clin Invest*. 2010;120(6):1862-1872.
18. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*. 2003;19(3):368-375.
19. Cosio FG, Grande JP, Wadei H, Larson TS, Griffin MD, Stegall MD. Predicting subsequent decline in kidney allograft function from early surveillance biopsies. *Am J Transplant*. 2005;5(10):2464-2472.
20. Roufosse C, Simmonds N, Clahsen-Van Groningen M, et al. A 2018 reference guide to the banff classification of renal allograft pathology. *Transplantation*. 2018;102(11):1795-1814.
21. Haas M, Loupy A, Lefaucheur C, et al. The Banff 2017 Kidney Meeting Report: revised diagnostic criteria for chronic active T cell-mediated rejection, antibody-mediated rejection, and prospects for integrative endpoints for next-generation clinical trials. *Am J Transplant*. 2018;18(2):293-307.
22. Denic A, Lieske JC, Chakkera HA, et al. The substantial loss of nephrons in healthy human kidneys with aging. *J Am Soc Nephrol*. 2017;28(1):313-320.
23. 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.
24. Kurian SM, Williams AN, Gelbart T, et al. Molecular classifiers for acute kidney transplant rejection in peripheral blood by whole genome gene expression profiling. *Am J Transplant*. 2014;14(5):1164-1172.
25. Friedewald JJ, Kurian SM, Heilman RL, et al. Development and clinical validity of a novel blood-based molecular biomarker for subclinical acute rejection following kidney transplant. *Am J Transplant*. 2019;19(1):98-109.
26. Halloran PF, Reeve J, Akalin E, et al. Real time central assessment of kidney transplant indication biopsies by microarrays: the INTERCOMEX study. *Am J Transplant*. 2017:2851-2862.
27. O'connell PJ, Zhang W, Menon MC, et al. Biopsy transcriptome expression profiling to identify kidney transplants at risk of chronic injury: a multicentre, prospective study. *Lancet*. 2016;388(10048):983-993.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Park WD, Kim DY, Mai ML, et al. Progressive decline of function in renal allografts with normal 1-year biopsies: Gene expression studies fail to identify a classifier. *Clin Transplant*. 2021;35:e14456. https://doi.org/10.1111/ctr.14456