

Inflammatory macrophage-associated 3-gene signature predicts subclinical allograft injury and graft survival

Tej D. Azad,^{1,2} Michele Donato,^{1,2} Line Heylen,³ Andrew B. Liu,^{1,2} Shai S. Shen-Orr,⁴ Timothy E. Sweeney,^{1,2} Jonathan Scott Maltzman,⁵ Maarten Naesens,³ and Purvesh Khatri^{1,2}

¹Stanford Institute for Immunity, Transplantation and Infection and ²Division of Biomedical Informatics Research, Department of Medicine, Stanford University, Stanford, California, USA. ³Department of Microbiology and Immunology, KU Leuven – University of Leuven, Leuven, Belgium. ⁴Department of Immunology, Technion-Israel Institute of Technology, Haifa, Israel. ⁵Division of Nephrology, Department of Medicine, Stanford University, Stanford, California, USA.

Late allograft failure is characterized by cumulative subclinical insults manifesting over many years. Although immunomodulatory therapies targeting host T cells have improved short-term survival rates, rates of chronic allograft loss remain high. We hypothesized that other immune cell types may drive subclinical injury, ultimately leading to graft failure. We collected whole-genome transcriptome profiles from 15 independent cohorts composed of 1,697 biopsy samples to assess the association of an inflammatory macrophage polarization-specific gene signature with subclinical injury. We applied penalized regression to a subset of the data sets and identified a 3-gene inflammatory macrophage-derived signature. We validated discriminatory power of the 3-gene signature in 3 independent renal transplant data sets with mean AUC of 0.91. In a longitudinal cohort, the 3-gene signature strongly correlated with extent of injury and accurately predicted progression of subclinical injury 18 months before clinical manifestation. The 3-gene signature also stratified patients at high risk of graft failure as soon as 15 days after biopsy. We found that the 3-gene signature also distinguished acute rejection (AR) accurately in 3 heart transplant data sets but not in lung transplant. Overall, we identified a parsimonious signature capable of diagnosing AR, recognizing subclinical injury, and risk-stratifying renal transplant patients. Our results strongly suggest that inflammatory macrophages may be a viable therapeutic target to improve long-term outcomes for organ transplantation patients.

Introduction

Diverse subclinical graft injury processes, taking place over many years, result in chronic allograft failure (1, 2). Although immunosuppressive drugs have reduced acute rejection (AR) and improved 1-year survival rates in solid organ transplant, long-term graft failure rates remain stubbornly high (3). This observation strongly suggests the existence of additional allograft injury mechanisms that have not yet been identified. This incomplete understanding of graft failure pathophysiology limits the development of new treatments that may enable long-term transplant success (4). Further, the development of effective treatments is hindered because of a lack of early markers of graft injury and failure, which requires that patients must be followed long term to understand whether a new treatment is effective. In addition, a suitable biomarker for immune-mediated graft damage would help in identifying appropriate treatments on a patient basis.

Previous reports have implicated distinct gene expression patterns in acute cellular rejection (ACR) and antibody-mediated rejection (AMR) (5, 6). In contrast, using a multicohort analysis (7–9) of 13 independent data sets consisting of 1,164 graft biopsies from 4 transplanted organs, we identified and validated a common rejection module (CRM) consisting of 11 genes, which (a) is able to diagnose AR, (b) correlates with the extent of graft injury, and (c) predicts subclinical graft injury (7). We also demonstrated that the CRM is mechanistically involved in graft injury by identifying two FDA-approved drugs that reduced graft-infiltrating immune cells and extended graft survival in a mouse model of cardiac transplantation and electronic health records of renal transplant patients (7). Collectively, these results suggest the possibility of a gene signature common to both ACR and AMR (10–12).

Authorship note: T.D. Azad and M. Donato contributed equally to this work.

Conflict of interest: T.E. Sweeney and P. Khatri are stockholders of Inflammatrix Inc., which is focused on sepsis diagnosis.

Submitted: June 14, 2017

Accepted: December 12, 2017

Published: January 25, 2018

Reference information:

JCI Insight. 2018;3(2):e95659. <https://doi.org/10.1172/jci.insight.95659>.

Macrophage infiltration has been implicated in both AR (13) and chronic graft failure (10, 11), yet a precise role for this cell type in organ transplantation remains unclear (12). Macrophages take on diverse phenotypic states in response to environmental stimuli, with activation characterized by a spectrum of transcriptional reprogramming (14), ranging from an antiinflammatory phenotype in the presence of IL-4 or IL-10 to a proinflammatory phenotype in the presence of IFN- γ , referred to as M(IL-4), M(IL-10), and M(IFN- γ), respectively (15). Several studies have described transcriptional signatures associated with each of these activation states (14, 16). Here, we hypothesized that inflammatory macrophages, M(IFN- γ), may explain substantial part of the unidentified allograft injury mechanisms that are common to both ACR and AMR and could identify early biomarkers of graft injury and failure. We correlated literature-derived gene signatures for M(IFN- γ), M(IL-4), and M(IL-10) with the CRM. We found that M(IFN- γ) was highly correlated with the CRM across multiple independent renal transplant cohorts, whereas M(IL-4) and M(IL-10) showed moderate correlations. We identified a 3-gene signature as a subset of the M(IFN- γ) that accurately diagnosed subclinical and acute renal graft injury, correlated with the extent of graft injury, and predicted progressive graft injury and failure. We validated the signature in 3 independent cohorts of renal graft biopsies using RT-PCR and microarrays. Using 4 additional data sets, we extended our analysis to show that the 3-gene signature accurately distinguished AR samples in cardiac transplant but had lower accuracy in lung transplant. Our results strongly suggest an important role for inflammatory macrophages in early graft injury that leads to graft failure.

Results

Systematic search and preprocessing of allograft transcriptomes. We conducted a systematic search of the NIH Gene Expression Omnibus (GEO) to collect all available clinical transplant biopsy data sets. We identified 15 data sets matching the search criteria (for-cause or protocol biopsies obtained from organ transplant patients). The data sets comprised 1,697 biopsy samples from renal (11 data sets, 1,571 samples), cardiac (3 data sets, 94 samples), and lung transplants (1 data set, 32 samples; Table 1). Four of these data sets included only protocol biopsies (7, 17–20), six included only for-cause biopsies (21–26), and five included both protocol and for-cause biopsies (7, 27–29). We curated each data set to identify allograft biopsies from AR and stable (STA) patients. We removed biopsy samples corresponding to healthy controls from further analysis, as they were not biopsied for a clinical indication. For each study, we used the sample phenotypes as defined by the corresponding published study (see Table 1; for further details see Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.95659DS1>).

All renal transplant data sets, except GSE48581, were phenotypically defined using the Banff criteria available at the time of the study. GSE48581 was classified using the Alberta Transplant Applied Genomics Center Reference Standard Classification. All heart transplant biopsies were graded according to International Society for Heart & Lung Transplantation (ISHLT) criteria, where samples with ISHLT grade 0 were used as STA samples and those with grade 3A or higher were used as AR samples (29, 30). GSE6095 used bronchoalveolar lavage (BAL) samples that were selected to maximize potential differences in gene expression between AR and STA samples and to minimize bias from confounding factors as described previously (28). We divided the data sets in discovery and validation cohorts. As discovery cohorts, we selected 8 renal transplant data sets that were publicly available in GEO before October 2013. Although GSE25902 was publicly available prior to October 2013, we used it as a validation cohort because it was the only longitudinal data set. We used all renal transplant data sets published after October 2013 (GSE53605 and GSE76882). Finally, we used other cardiac and lung transplant data sets to extend our findings from renal transplant biopsies to other solid organ transplants.

Higher ratio of proinflammatory macrophage scores to antiinflammatory macrophage scores is associated with higher risk of rejection. We hypothesized that macrophage polarization plays an important role in graft injury and expression of proinflammatory macrophage polarization signature will be correlated with that of the CRM. To test this hypothesis, we used previously described macrophage polarization signatures in presence of IFN- γ , IL-4, and IL-10 (16, 31), from which we removed all overlapping genes with the CRM. We defined the M(IFN- γ) score, M(IL-4) score, and M(IL-10) score as the geometric mean of the corresponding signature genes in a sample. We found that, across 10 independent data sets of 1,499 renal graft biopsies, M(IFN- γ) scores were highly correlated with the CRM scores ($r = 0.95$, $P < 2.2 \times 10^{-16}$, whereas M(IL-4) and M(IL-10) scores had moderate correlations ($r = 0.77$ and 0.64 , respectively, $P < 2.2 \times 10^{-16}$; Supplemental Figure 1).

Table 1. Publicly available organ transplant data sets

GSE	Author, ref.	Transplanted organ	Phenotypes (no. of samples)	No. of AR	No. of STA	No. of RD	For-cause or protocol?	Date publicly available (month-YY)	Platform
Discovery data sets									
GSE1563	Flechner et al. (20)	Kidney	AR (5), STA (10), ATN (2), borderline (2), CNIT (2), FSGS (1)	5	10	7	Protocol	July-04	GPL96
GSE21374	Einecke et al. (25)	Kidney	AR (76), STA (206)	76	206	0	Cause	May-10	GPL570
GSE36059	Reeve et al. (24)	Kidney	AMR (65), ACR (35), mixed (22), STA (281)	122	281	0	Cause	February-13	GPL570
GSE50058	Khatri et al. (7)	Kidney	AR (43), STA (58)	43	58	0	Mixed	October-13	GPL570
GSE9493	Rödder et al. (26)	Kidney	AR (10), AR+CAN (7), CAN (25), borderline (4), STA (21)	17	25	0	Cause	February-09	GPL570
GSE47097	Rekers et al. (23)	Kidney	AR (36), STA (4)	36	4	0	Cause	May-13	GPL6883
GSE48581	Halloran et al. (22)	Kidney	STA (222), AMR (40), TCMR (32), mixed (6)	78	222	0	Cause	September-13	GPL570
GSE50084	Ó Broin et al. (21)	Kidney	AR (28), STA (33)	28	33	0	Cause	August-13	GPL6244
Validation data sets									
GSE25902	Naesens et al. (19)	Kidney	–	–	–	72	Protocol	December-11	GPL570
GSE53605	Maluf et al. (17)	Kidney	AR (13), STA (18), CNIT (14), IF/TA (10)	13	18	24	Protocol	June-14	GPL571
GSE76882	Modena et al. (39)	Kidney	AR (54), ADNR (40), IF/TA (71), IF/TA+I (10), IF/TA+AR (29), STA (99)	54	99	40	Mixed (114 surveillance, 120 for-cause)	January-16	GPL13158
GSE2596	Morgun et al. (29)	Heart	AR (16), Pre-AR (4), STA (27), Pre-Chagas (7), Chagas (8)	16	27	0	Mixed (monitoring for 6 months, additional were taken if rejection was suspected)	October-05	GPL1053
GSE4470	Morgun et al. (29)	Heart	AR (12), STA (13), Pre-AR (2)	12	13	0	Mixed (monitoring for 6 months, additional were taken if rejection was suspected)	March-06	GPL1053
GSE9377	Holweg et al. (30)	Heart	AR (7), STA (19)	7	19	0	Protocol	October-08	GPL887
GSE6095	Patil et al. (28)	Lung	AR (18), STA (14)	18	14	0	Mixed (22 surveillance, 10 clinically indicated)	January-08	GPL96

ACR, acute cellular rejection; ADNR, acute dysfunction with no rejection; AMR, antibody-mediated rejection; AR, acute rejection; ATN, acute tubular necrosis; CAN, chronic allograft nephropathy; CNIT, calcineurin inhibitor toxicity; FSGS, focal segmental glomerulosclerosis; IF/TA, interstitial fibrosis/tubular atrophy; IF/TA+I, interstitial fibrosis/tubular atrophy plus inflammation; STA, stable; TCMR, T cell-mediated rejection.

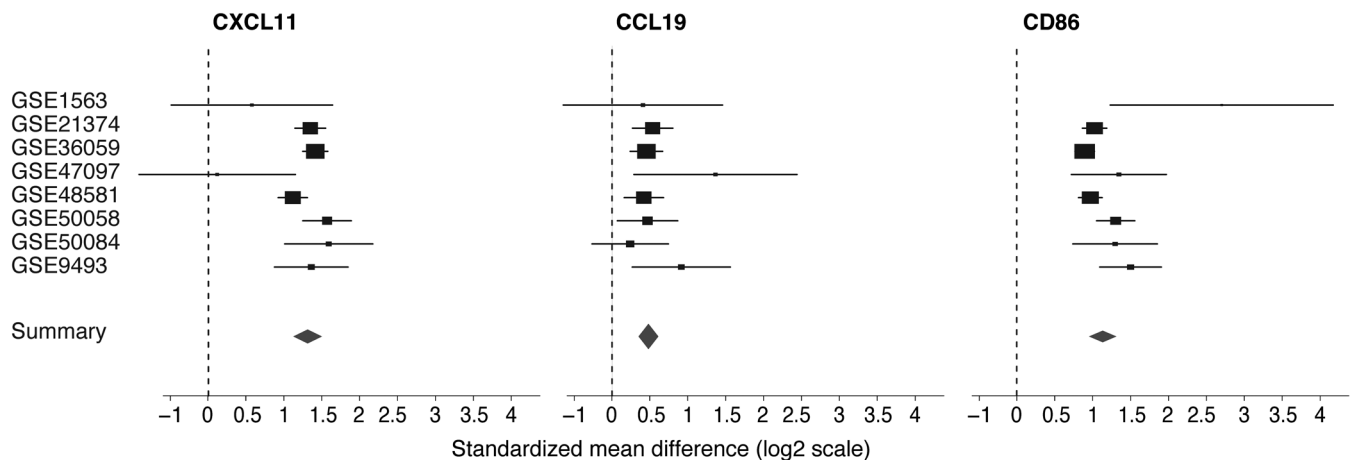


Figure 1. Three inflammatory macrophage polarization genes are overexpressed in AR patients across all renal transplant discovery cohorts. The x axis represents the standardized mean difference between AR and STA, computed as Hedges' g . The rectangle size is proportional to the standard error of the mean in the study. Whiskers denote 95% CI. Diamonds, overall, represent the combined mean difference for a given gene. Diamond width denotes 95% CI of overall mean difference.

Next, we performed a linear regression analysis with M(IFN- γ) score, M(IL-4) score, and M(IL-10) scores as independent predictors of the CRM score to investigate which of them explained more variance in the CRM score than the others. We found that these scores collectively explained 90% of variance in the CRM scores, of which the M(IFN- γ) score explained 60% ($P < 2.2 \times 10^{-16}$) of variance; this was more than twice the amount of variance explained by the M(IL-4) score (24%; $P < 2.2 \times 10^{-16}$) and almost 4 times the amount of variance explained by the M(IL-10) score (15.65%; $P < 2.2 \times 10^{-16}$).

It is possible that as the recipient immune system mounts response to the graft, the graft itself secretes antiinflammatory signals, indicating that both proinflammatory and antiinflammatory signals will be present in a transplant patient with AR. Therefore, we hypothesized that the patients with high proinflammatory macrophage scores compared with antiinflammatory macrophage scores will be at a higher risk of rejection. We performed a logistic regression analysis using the ratio of M(IFN- γ) score to M(IL-4) score as a predictor of AR. We found that higher ratio is associated with the higher risk of rejection (log likelihood of AR = 0.7; $P = 2.86 \times 10^{-10}$).

Collectively, these results demonstrate that proinflammatory macrophage polarization is strongly correlated with the CRM and explain the large amount of variance in the CRM. Our analysis further demonstrates that an increase in the ratio of proinflammatory to antiinflammatory macrophage in the graft significantly increases the risk of rejection.

Identification of a parsimonious proinflammatory macrophage gene signature. Our correlation analysis of the 25 M(IFN- γ)-stimulated macrophage polarization signature genes across 104 diseases using the MetaSignature database (8) showed that these genes are substantially correlated with each other (Supplemental Figure 2A), suggesting that the M(IFN- γ) signature could be further reduced. Therefore, we sought to identify a parsimonious M1 gene signature that can distinguish AR and STA allograft biopsies. We used a generalized linear regression model using the 25-gene M(IFN- γ) signature with binary classification of AR versus STA to identify a subset of these genes that was most discriminative in the discovery cohorts (Table 1). We identified 3 genes that were selected in the regression model in at least 4 of 8 discovery data sets (Supplemental Table 2): *CXCL11*, *CCL19*, and *CD86*. As expected, these genes were consistently expressed at a higher level in AR patient renal graft biopsies than in those of STA patients in the discovery cohorts (Figure 1). Importantly, despite the biological and technical heterogeneity in the discovery cohorts, there was little or no heterogeneity in these 3 genes. Finally, we found that across 25 immune cell types and epithelial cells, individually, each of the 3 genes was expressed in multiple cell types. *CD86* was expressed at similar levels in multiple myeloid cell types, whereas *CXCL11* and *CCL19* were highly expressed in inflammatory macrophages compared with other cell types (Supplemental Figure 2B). Similar to a previous report (5), we observed that *CXCL11* was also expressed in endothelial cells but at a substantially lower level than inflammatory macrophages. Importantly, we found that as a signature, these 3 genes are collectively enriched in inflammatory macrophages ($P = 4.87$

$\times 10^{-6}$, FDR = 0.013%; Supplemental Figure 2C). These results suggest that, although each of the 3 genes may be expressed in multiple cell types, only inflammatory macrophages predominantly expressed all 3 genes together and at significantly higher levels than the other cell types.

Validation of the 3-gene signature using microarrays and RT-qPCR. We defined the geometric mean of the expression of the 3 genes in the graft biopsy as the M1 score of an allograft biopsy. As expected, in the 8 discovery cohorts, the M1 score distinguished AR samples from STA samples with high accuracy (mean AUC = 0.79, range: 0.57–0.89; Figure 2A). In the 2 independent renal transplant data sets, which contained protocol and for-cause biopsies, M1 scores were higher in the AR group (Wilcoxon test $P = 6.6 \times 10^{-7}$ and $P = 6.2 \times 10^{-16}$ for GSE53605 and GSE76882, respectively; Supplemental Figure 3) and had AUC of 0.94 and 0.9 for GSE53605 and GSE76882, respectively (Figure 2B). Further, logistic regression using M1 score as an independent variable showed that each unit increase in the M1 score increased the odds of AR by 23- and 10-fold in GSE53605 and GSE76882, respectively. Maximized for sensitivity, the test characteristics across the validation data sets were (32) sensitivity = 95% and specificity = 63% for AR versus STA. Assuming 15% prevalence of AR, the corresponding negative predictive value is 99%.

Next, we explored whether M1 score performed differently in ACR and AMR. Two independent data sets, GSE36059 and GSE48581, identified whether a patient had AMR, ACR, or mixed (AMR and ACR) rejection. In both studies, we found that the M1 scores were higher in ACR ($P = 3.35 \times 10^{-20}$) and AMR ($P = 4.13 \times 10^{-17}$) than in STA patients (Supplemental Figure 4). Further, the M1 scores were higher in ACR and mixed rejection patients than in AMR patients.

Further, we measured the expression of these 3 genes using RT-qPCR in an independent cohort of 22 renal graft biopsies, consisting in 11 AR and 11 STA adult patients. The cohort was matched for age, serum creatinine, eGFR, and time after transplant (Table 2). Each of the genes was overexpressed during AR in this cohort (Supplemental Figure 5). Importantly, the M1 score also distinguished AR samples from STA samples with high accuracy (AUC = 0.89; Figure 2B) in this cohort, demonstrating its robustness to the technological platform used for measuring the 3-gene signature.

Finally, because the M(IFN- γ) score was highly correlated with the CRM score, which in turn is common across organ transplant rejections, we attempted to see if the 3-gene signature could also distinguish AR and STA samples in other organ transplants. We found that the M1 score distinguished AR from STA samples with high accuracy in 3 heart transplant data sets (mean AUC = 0.92, range: 0.88–0.97, Figure 2C; see Table 1 for details of the data sets) but with moderate accuracy in lung transplants (AUC = 0.68; Supplemental Figure 6). Collectively, these results suggest that, unlike the CRM score, the M1 score can distinguish AR and STA samples in renal and cardiac transplants but not across all solid organ transplants.

Intragraft M1 score is an early prognostic marker of subclinical allograft injury and graft failure. Next, we investigated whether the M1 score could also identify chronic histological damage in renal allografts, as defined by the chronic allograft damage index (CADI) (2). We used a longitudinal cohort of renal allograft biopsies (GSE25902, ref. 19) (Supplemental Figure 7); this cohort is referred to as the “longitudinal cohort” herein. None of the patients in the longitudinal cohort experienced Banff-grade acute T cell-mediated rejection within 2 years of transplantation. At 24 months after transplant, 12 of the 24 patients had greater histological progression, scored by the incremental CADI score. We defined these samples as progressors, and the remaining samples as nonprogressors. The M1 scores for the nonprogressor, progressor, and AR groups in the 72 patients were positively correlated with the extent of injury (Jonckheere-Terpstra trend test [JT], ref. 33, $P = 0.005$; Figure 3A). The M1 score was also positively correlated with different components of the Banff score (JT $P < 1 \times 10^{-6}$; Figure 3, B–E) in the longitudinal cohort, demonstrating that the M1 score correlates with various aspects of subclinical injury.

A repeated-measures analysis of variance of the longitudinal cohort showed that the M1 scores differed between the progressor and nonprogressor groups ($P = 3.76 \times 10^{-5}$, Figure 3F), where M1 scores increased over time for both groups ($P = 2.54 \times 10^{-7}$). Importantly, as time after transplant increased, the M1 scores for the progressor group increased at a faster rate compared with those for the nonprogressor group ($P = 0.017$). Interestingly, the M1 scores of nonprogressors were highly similar to those of STA groups (Figure 3F). Therefore, we asked if the M1 score for a 6-month protocol biopsy could predict future histological damage to the graft, as defined by CADI score at 24 months after transplant. M(IFN- γ) scores in 6-month biopsies predicted histological damage with 91.6% sensitivity and 83% specificity (AUC = 0.91; Figure 3G).

Finally, we explored whether the M1 score was associated with graft survival. In GSE21374, where time-to-event data was available, we divided renal allografts into two groups based on mean of M1 score:

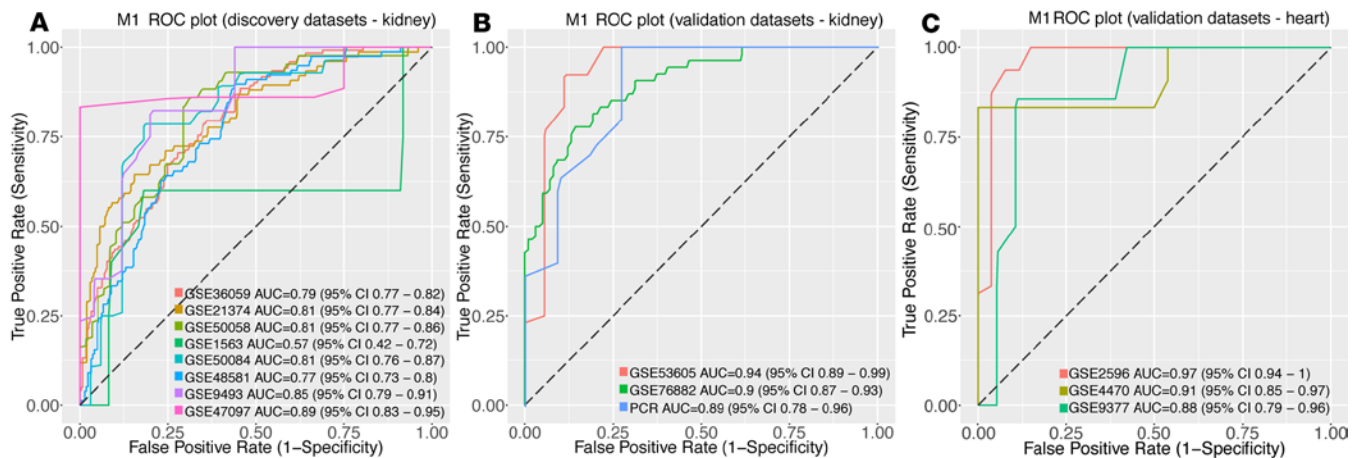


Figure 2. Three-gene M1 score can distinguish AR and STA samples with high accuracy in renal transplant patients. The 3-gene M1 score can distinguish AR and STA samples with high accuracy in renal transplant patients in (A) 8 discovery cohorts of 1,251 renal transplant biopsies, (B) 3 validation cohorts of 270 renal allograft biopsies, and (C) 3 validation cohorts of 94 heart allograft biopsies. The PCR cohort represents 22 renal allograft biopsy samples that profiled the M1 genes using RT-PCR; the other 2 validation cohorts were publicly available after October 2013.

high M1 score, which included samples with M1 score greater than the mean M1 score, and low M1 score, which included samples with M1 score less than or equal to the mean M1 score. Cox proportional hazard analysis showed that graft survival was higher in the low M1 group (hazard ratio = 2.79, 95% CI: 1.56–5.01, $P = 5.5 \times 10^{-4}$, Figure 4). When we corrected for time since transplant to account for ongoing subclinical graft injury, the hazard ratio for graft failure increased to 3.24 (95% CI: 1.8 – 5.8, $P = 9 \times 10^{-5}$).

Discussion

Long-term graft failure rates continue to be unacceptably high despite the development of immunosuppressive drugs, underscoring the unmet need for robust prognostic biomarkers of allograft injury and failure. Here, we hypothesized that an inflammatory macrophage-specific gene signature would identify robust diagnostic and prognostic signatures of graft injury and failure. Despite the biological and technical heterogeneity in existing data sets, we identified a parsimonious 3-gene signature enriched in inflammatory macrophages that is (a) highly correlated to the common rejection module across all solid organ transplants, (b) overexpressed during AR, (c) correlated with the extent of graft injury as defined by Banff and CADI, and (d) able to identify allografts at the risk of failure irrespective of time after transplant.

We found a near-perfect correlation between the inflammatory macrophage, M(IFN- γ), gene signature and the CRM signature (7) in renal transplants that was substantially higher than that of the antiinflammatory macrophage, M(IL-4) or M(IL-10), gene signatures. This result suggests innate and adaptive immune responses functioning in tandem in renal and heart transplants. Further, our analysis found that 60% of the variance observed in the CRM is explained by M(IFN- γ), and the ratio of M(IFN- γ) score to that of antiinflammatory macrophage polarization scores, M(IL-4) and M(IL-10), is positively associated with increased AR risk. However, the 3-gene set is different from the 11-gene CRM. A predominant source of the 3-gene set is enriched for innate immune response, M(IFN- γ), macrophages, whereas the CRM score largely reflects the adaptive immune response, with many of the 11 genes in the CRM highly expressed in T cells, including *LCK*, *CD6*, *CD7*, and *NKG7*. Collectively, this result suggests that dysregulation of balance in the innate immune system between proinflammatory and antiinflammatory macrophage polarization may result in AR and subclinical chronic allograft injury through the adaptive immune system, which may be difficult to detect in protocol biopsies. AR itself may be cell mediated (ACR) or antibody mediated (AMR). Compared with STA biopsies, we found that the M(IFN- γ) score was higher in both ACR and AMR, suggesting that inflammatory macrophages contribute to both types of AR.

Other groups have investigated intra-graft macrophages, though many investigations have focused on murine models of cardiac transplantation (34–36). Wu et al. interrogated macrophage polarization in transplant rejection (35) and found that graft-infiltrating macrophages in a murine model of chronic cardiac transplant rejection demonstrated an M2 (IL-4) phenotype; they reported that blocking the recep-

Table 2. Details of independent RT-qPCR data set used for validation, comprising 22 renal allograft biopsies (11 AR, 11 STA)

	Stable (n = 11)	Acute rejection (n = 11)	P value (t test)
Male	7	8	
Female	4	3	
Protocol biopsy	10	8	
Indication biopsy	1	3	
Age in years (mean ± SD)	53.9 ± 15.6	53.0 ± 13.41	0.85
Serum creatinine (mean ± SD)	2.2 ± 1.2	2.0 ± 0.63	0.54
eGFR (mean ± SD)	36.5 ± 19.3	39 ± 15.82	0.75
Days after transplant (mean ± SD)	91 ± 10.8	99 ± 47.31	0.59

eGFR, estimated glomerular filtration rate.

tor P2X7, which may be upregulated in M2 macrophages, improved long-term cardiac allograft survival. This group also provided evidence that mTOR deletion both suppresses M2 polarization and may prevent chronic rejection. An important limitation of the initial study is the use of single markers of polarization state (iNOS, M1, and CD206, M2), contrary to current experimental guidelines (15). Studies of acute and chronic rejecting human allografts with panels of macrophage polarization markers are necessary to extend the understanding of macrophage polarization in transplant rejection. Toki et al. contributed to this objective by staining for two markers of M2 macrophages (CD163 and CD206) in 46 human renal allograft biopsies (10). While the majority of CD68-positive cells coexpressed CD206 in ACR (98.7%) and AMR (73.5%), transcriptional profiles of the 12-month biopsies revealed that inflammatory and IFN- γ signaling genes were associated with interstitial fibrosis/tubular atrophy.

Instead of starting with the whole-transcriptome analysis, we started with a targeted set of gene lists, in which genes are significantly overexpressed in M(IFN- γ), M(IL-4), and M(IL-10) macrophages. Because of this, the 3-gene signature is enriched in inflammatory macrophages compared with other immune cell types and endothelial cells. However, despite this difficulty, in concordance with a previous report (5), we observed that epithelial cells also expressed *CXCL11*. In addition, we observed that M(IFN- γ) expressed significantly higher levels of *CXCL11*. Further, no cell types other than M(IFN- γ) expressed all 3 genes. These observations point to a difficulty in definitively associating inflammatory macrophages as a predominant source, and further work is needed to definitively determine the source of the 3-gene signature that is outside the scope of this manuscript.

Given the robust and consistent correlation of inflammatory macrophage polarization with the CRM and its diagnostic and prognostic power, our results suggest that therapeutic inhibition of inflammatory macrophages may reduce graft injury and extend graft survival. Previously, we used a HLA-mismatched mouse model of cardiac transplant to show that treatment with atorvastatin reduced F4/80⁺ macrophages in the allograft but cyclosporine A treatment did not (7). Further, statins can reduce macrophage proliferation by inactivating the small G protein/p38 MAPK pathway (37). Statin pretreatment is shown to reduce renal damage as well as renal dysfunction in ischemia/reperfusion injury by inducing heme oxygenase-1 (*HMOX1*) in infiltrating macrophages in rat models (38). Importantly, graft survival was increased in the atorvastatin treatment group compared with untreated mice. We also observed reduced graft failure with statin treatment in a cohort of more than 2,500 renal transplant patients (7). Collectively, these results suggest that further mechanistic studies should be performed to better understand the mechanisms by which statins may reduce graft-infiltrating macrophages and their proliferation and improve graft survival.

More importantly, by identifying patients with subclinical injury and those with higher risk of graft failure substantially earlier than clinical manifestation in a longitudinal cohort of interstitial fibrosis/tubular atrophy patients, our results provide strong evidence that the M1 score could allow clinicians to treat high-risk patients before significant damage is done to the allograft. The M1 score could potentially represent a marker for endpoint for graft loss and injury, as it could identify patients at risk of injury prior to clinical manifestation and those at higher risk of graft loss. If validated in additional prospective cohorts, these results strongly suggest that the M1 score can help enrich enrollment in clinical trials by removing patients

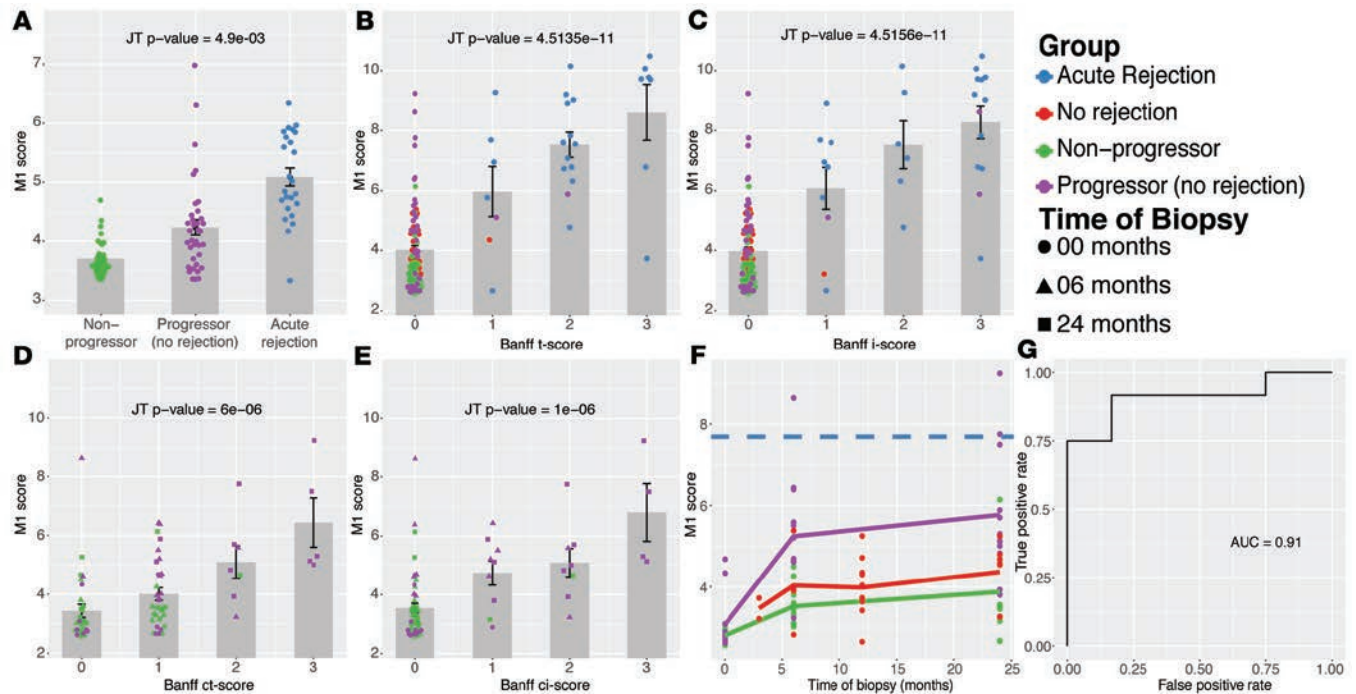


Figure 3. M1 score is diagnostic and prognostic in longitudinally collected protocol biopsies from renal transplant patients. (A) M1 scores in renal biopsies from nonprogressors, progressors, and AR transplant patients in GSE25902. (B and C) Comparison of M1 scores (y axis) with Banff tubulitis (t-score) and interstitial inflammation (i-score) scores in renal transplant patients, respectively. (D and E) Comparison of M1 scores with ct-score and ci-score in STA renal transplant patients who did not have AR episodes in the first 2 years after transplant. (F) Change in M1 scores over time in STA renal transplant patients. Patients were divided into 2 groups: progressors ($n = 12$, CADI ≥ 6) and nonprogressors ($n = 12$, CADI < 6) 2 years after transplant. Protocol biopsies were obtained from each patient at 3 time points: the time of transplant, 6 months after transplant, and 24 months after transplant. Repeated-measures analysis of variance was used to analyze CRM scores between progressors and nonprogressors over the time points. The dotted blue line represents the mean M1 score in AR biopsies; solid red line represents the mean M1 score in STA biopsies. (G) Six-month protocol biopsy M1 scores were used to predict patients with severe histological damage ≥ 2 years after transplant. A JT trend test was used to compute P values for correlations between M1 scores and Banff t- and i-scores as well as ct- and ci-scores. Error bars indicate the standard error of the mean.

whose grafts are unlikely to fail regardless of treatment and, hence, not show any potential improvement by the therapy. The M1 score could also serve as a surrogate endpoint, reducing the length of a trial by markedly reducing the interval between intervention and ability to determine efficacy.

We found that the M1 score in protocol biopsies 6 months following transplant predicted histological damage at 24 months after transplant with sensitivity of 91.6% and specificity of 83%. This result, combined with its high negative predictive value of 99% at 15% prevalence suggest that the M1 score be most applicable as a “rule-out test” for long-term graft loss. That is, patients with low M1 scores may be candidates for less aggressive immunosuppression, limiting treatment-related side effects.

Our results are in concordance with those of recent studies and improve upon them by identifying a parsimonious gene signature of graft injury and failure in specific immune cells. A recent study has shown association of immune response and inflammation-related gene networks with AR and graft loss, even in patients without histological evidence of inflammation (39). O’Connell et al. described a 13-gene set for predicting chronic allograft injury (40). This gene set demonstrated high predictive ability for the development of fibrosis at 1 year and early allograft loss. In GSE21374 (25), the 13-gene set had a hazard ratio of 2.72 for graft loss, as determined by the two principle components of their gene set that were significantly associated with graft loss. In the same data set, the M1 score had a virtually identical hazard ratio of 2.79 for graft loss. However, the mechanistic role of the 13-gene set is not clear, although 7 of 13 genes seem overexpressed in fibroblasts. In contrast, our 3-gene signature is more parsimonious, is enriched for inflammatory macrophage polarization, and provides a possible mechanistic explanation of the allograft injury that should be further investigated in follow-up studies.

Our study has several limitations. First, while we validated the 3-gene set using RT-qPCR in an independent cohort of renal transplant biopsies, it should be prospectively validated in a larger cohort. Sec-

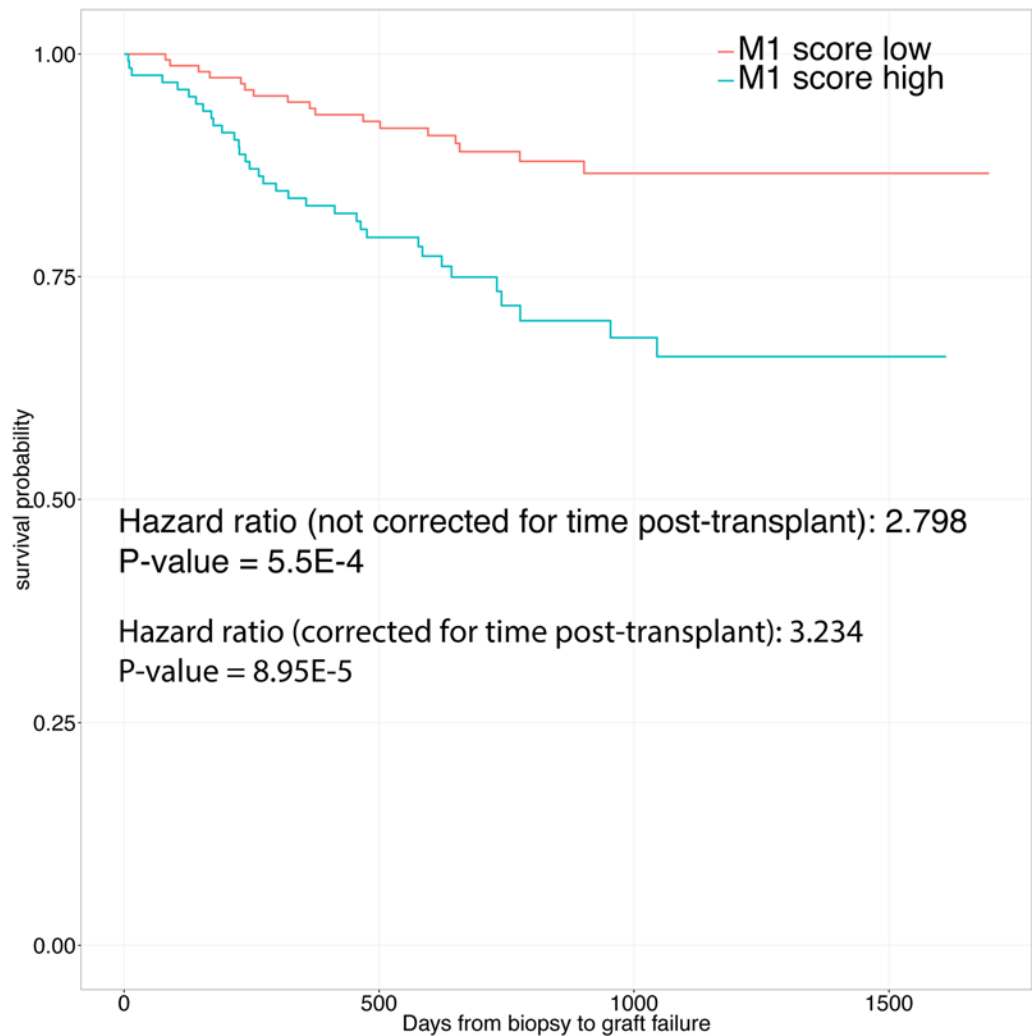


Figure 4. High M1 score increases the risk of graft failure. Renal transplant biopsies in GSE21374 were divided into two groups based on the mean M1 score within the data set. Biopsies with M1 scores higher than the mean M1 score in GSE21374 were assigned to the high M1 score group; the rest were assigned to the low M1 score group.

ond, the prognostic value of the M1 score also needs to be validated in a prospective cohort. Third, the M1 score had lower accuracy in distinguishing AR and STA patients among the lung transplant patients than in distinguishing those from renal and heart transplant patients. The reasons for this remains unclear. It is possible that the lower AUCs were due to smaller sample sizes in the lung data set compared with heart and renal transplant cohorts. It is also possible that inflammatory macrophage polarization has lower association with the CRM in lung transplant patients, potentially due to the contribution of organ-specific responses that differentiate these tissues. However, it is also possible that the BAL samples from lung transplant patients used in our analysis are not representative of biology of rejection in lung biopsy, as BAL may reflect different constituent cell populations and functional and phenotypic differences within similar cell types (41). Fourth, as mentioned above, further studies are needed to definitively identify the sources of the 3 genes in AR.

Overall, our analysis identified a parsimonious 3-gene signature with high diagnostic and prognostic power in 8 discovery and 8 validation cohorts comprising more than 1,700 graft biopsies from 4 different organs. Optimizing a clinical assay for this gene set to get results within a window of clinical relevance should be feasible, as demonstrated by RT-PCR validation in an independent cohort. Additional prospective studies are needed to further confirm our findings and to investigate the cell types and molecular pathways these genes are involved in.

Methods

Definition of proinflammatory macrophage, antiinflammatory macrophage, M(IFN- γ) score, and CRM scores. The M(IFN- γ) and M(IL-4) macrophage signatures used in this study were described previously as proinflammatory (M1) and antiinflammatory (M2) signatures by Krausgruber et al. (31). The M(IL-10) signature was drawn from the results of a transcriptome analysis of IL-10-stimulated macrophages (16).

To assess the association between macrophage signatures and risk of rejection, we correlated their gene expression levels with those of the 11-gene CRM (7), representing a common immunological and inflammatory response during AR in different transplanted organs.

All gene expression scores used in the manuscript [M(IL-4) score, M(IL-10) score, M(IFN- γ) score, 3-gene M1 score, and CRM score] were calculated as a geometric mean of corresponding constituent signature genes for each patient. M(IFN- γ), M(IL-4), and M(IL-10) signatures consist of 25, 26, and 18 genes, respectively, whereas the CRM consists of 11 genes (7).

Relative importance of regressors in linear regression models. Macrophage scores for each patient were used in a linear regression model to assess their contribution to the CRM. We used a method based on averaging sequential sums of squares over orderings of regressors to compute relative importance of gene scores in linear regression models (42, 43).

Gene-gene expression correlations among genes in the inflammatory macrophage polarization signature. Gene-gene expression correlation was computed using MetaSignature (8). MetaSignature allows computation of correlations of gene pairs across expression profiles of more than 600 publicly available data sets, covering more than 100 diseases.

Selection of minimal gene set for diagnosing AR. We applied LASSO (44), a penalized regression method, to build a logistic regression model using a 25-gene signature (31) for M(IFN- γ) polarization to distinguish between AR and STA patients. We applied the regression model to 8 renal transplant data sets that were publicly available before October 2013 (the time of publication of the CRM). LASSO allows identification of sparse regression models by altering the model fitting process to select a subset of covariates for use in the final model, rather than using all of them. LASSO-based regression results in coefficients for some features (genes) to be exactly 0, allowing identification of the most important features in the model. We selected genes with non-0 coefficients in at least half of the cohorts (≥ 4), such that they were either positive or negative across all cohorts.

Immune cell-type enrichment tests. We performed immune cell-type enrichment as described previously (45, 46). Briefly, we searched GEO for gene expression profiles of clinical samples of relevant immune cell types. We limited our search to samples run on Affymetrix platforms to ensure platform effect homogeneity. A total of 19 data sets matched our criteria: GSE3982 (47), GSE5099 (48), GSE8668 (49), GSE11292 (50), GSE12453 (51), GSE13987 (52), GSE14879 (53), GSE15743 (54), GSE16020 (55), GSE16836 (56), GSE24759 (57), GSE28490 (58), GSE28491 (58), GSE31773 (59), GSE34515 (60), GSE38043 (34), GSE39889 (35), GSE42519 (36), and dGSE49910 (61). We downloaded all data sets in raw format and normalized each one using gcRMA (62). For each sample, the mean of multiple probes mapping to the same gene was taken as the gene value. We discarded genes that were not present in all samples. For each cell type, we created a single vector by computing mean expression for each gene across all samples. To obtain a Z score for a gene set in each cell-type vector, we computed the geometric mean of expression levels of the 3 genes. These scores were standardized across all cell types, such that the score represented the number of SDs from the group mean. This thus represents the enrichment of a given gene set in each cell type, relative to other cell types.

RT-quantitative PCR. We performed RT-qPCR for the 3 genes in renal allograft biopsies using high-throughput RT-qPCR (LightCycler 96, Roche Diagnostics). RNA was isolated with Allprep DNA/RNA/miRNA Universal kits (Qiagen) per the manufacturer's instructions. 2.5 μ g cellular RNA was reverse transcribed into cDNA using SuperScript II reverse transcriptase and random hexamer primers (Invitrogen Life Technologies). The PCR reaction was carried out in a mixture that contained appropriate sense and antisense primers in TaqMan assays of the 6 M(IFN- γ) genes and the TaqMan Fast Universal PCR Master Mixture (Applied Biosystems). *GAPDH* was used as housekeeping gene. Each sample was assayed in duplicate in a MicroAmp optical 96-well plate. The $\Delta\Delta$ Ct-method was used to determine relative gene expression levels. The IDs of the assays used in PCR are as follows: *GAPDH* (Hs02758991_g1), *CXCL11* (Hs04187682_g1), *CCL19* (Hs00171149_m1), and *CD86* (Hs01567026_m1).

Statistics. The Wilcoxon test was utilized to compare two continuous variables, the JT test was used to determine the significance of trends. The Wald test was used to determine the significance of parameters in regression models. All tests were 2 tailed. *P* values of less than 0.05 were considered significant.

Author contributions

PK, TDA, and MD conceived the study. TDA and MD collected data and performed statistical analysis. LH and MN performed RT-qPCR experiments. TDA and MD wrote the first draft of the manuscript. TDA, MD, LH, ABL, SSS, TE, JSM, MN, PK interpreted data, edited the manuscript, and approved the final version of the manuscript.

Acknowledgments

TDA is funded by the TL1 Clinical Research Training Program (TL1 TR 001084). PK is supported by grants from the Bill & Melinda Gates Foundation and NIAID grants R01 AI125197-01, 1U19AI109662, and U19AI057229.

Address correspondence to: Purvesh Khatri, 279 Campus Drive, Beckman Center B235A, Stanford, California 94305-5415, USA. Phone: 650.497.5281; Email: pkhatri@stanford.edu.

1. Baboolal K, Jones GA, Janezic A, Griffiths DR, Jurewicz WA. Molecular and structural consequences of early renal allograft injury. *Kidney Int.* 2002;61(2):686–696.
2. Yilmaz S, et al. Protocol core needle biopsy and histologic chronic allograft damage index (CADI) as surrogate end point for long-term graft survival in multicenter studies. *J Am Soc Nephrol.* 2003;14(3):773–779.
3. Lamb KE, Lodhi S, Meier-Kriesche HU. Long-term renal allograft survival in the United States: a critical reappraisal. *Am J Transplant.* 2011;11(3):450–462.
4. Lo DJ, Kaplan B, Kirk AD. Biomarkers for kidney transplant rejection. *Nat Rev Nephrol.* 2014;10(4):215–225.
5. Venner JM, Hidalgo LG, Famulski KS, Chang J, Halloran PF. The molecular landscape of antibody-mediated kidney transplant rejection: evidence for NK involvement through CD16a Fc receptors. *Am J Transplant.* 2015;15(5):1336–1348.
6. Venner JM, Famulski KS, Badr D, Hidalgo LG, Chang J, Halloran PF. Molecular landscape of T cell-mediated rejection in human kidney transplants: prominence of CTLA4 and PD ligands. *Am J Transplant.* 2014;14(11):2565–2576.
7. Khatri P, et al. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J Exp Med.* 2013;210(11):2205–2221.
8. Haynes WA, et al. Empowering multi-cohort gene expression analysis to increase reproducibility. *Pac Symp Biocomput.* 2017;22:144–153.
9. Sweeney TE, Haynes WA, Vallania F, Ioannidis JP, Khatri P. Methods to increase reproducibility in differential gene expression via meta-analysis. *Nucleic Acids Res.* 2017;45(1):e1.
10. Toki D, et al. The role of macrophages in the development of human renal allograft fibrosis in the first year after transplantation. *Am J Transplant.* 2014;14(9):2126–2136.
11. Pilmore HL, Painter DM, Bishop GA, McCaughan GW, Eris JM. Early up-regulation of macrophages and myofibroblasts: a new marker for development of chronic renal allograft rejection. *Transplantation.* 2000;69(12):2658–2662.
13. Bergler T, et al. Infiltration of macrophages correlates with severity of allograft rejection and outcome in human kidney transplantation. *PLoS ONE.* 2016;11(6):e0156900.
12. Salehi S, Reed EF. The divergent roles of macrophages in solid organ transplantation. *Curr Opin Organ Transplant.* 2015;20(4):446–453.
14. Xue J, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity.* 2014;40(2):274–288.
15. Murray PJ, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity.* 2014;41(1):14–20.
16. Lurier EB, et al. Transcriptome analysis of IL-10-stimulated (M2c) macrophages by next-generation sequencing. *Immunobiology.* 2017;222(7):847–856.
17. Maluf DG, et al. Evaluation of molecular profiles in calcineurin inhibitor toxicity post-kidney transplant: input to chronic allograft dysfunction. *Am J Transplant.* 2014;14(5):1152–1163.
18. Bohne F, et al. Intra-graft expression of genes involved in iron homeostasis predicts the development of operational tolerance in human liver transplantation. *J Clin Invest.* 2012;122(1):368–382.
19. Naesens M, et al. Progressive histological damage in renal allografts is associated with expression of innate and adaptive immunity genes. *Kidney Int.* 2011;80(12):1364–1376.
20. 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.
21. Ó Broin P, et al. A pathogenesis-based transcript signature in donor-specific antibody-positive kidney transplant patients with normal biopsies. *Genom Data.* 2014;2:357–360.
22. Halloran PF, et al. Potential impact of microarray diagnosis of T cell-mediated rejection in kidney transplants: The INTERCOM study. *Am J Transplant.* 2013;13(9):2352–2363.
23. Rekers NV, et al. Increased metallothionein expression reflects steroid resistance in renal allograft recipients. *Am J Transplant.* 2013;13(8):2106–2118.
24. Reeve J, et al. Molecular diagnosis of T cell-mediated rejection in human kidney transplant biopsies. *Am J Transplant.* 2013;13(3):645–655.
25. Einecke G, et al. A molecular classifier for predicting future graft loss in late kidney transplant biopsies. *J Clin Invest.* 2010;120(6):1862–1872.
26. Rödder S, et al. Renal allografts with IF/TA display distinct expression profiles of metzincins and related genes. *Am J Trans-*

- plant.* 2009;9(3):517–526.
27. Modena BD, et al. Gene expression in biopsies of acute rejection and interstitial fibrosis/tubular atrophy reveals highly shared mechanisms that correlate with worse long-term outcomes. *Am J Transplant.* 2016;16(7):1982–1998.
 28. Patil J, Lande JD, Li N, Berryman TR, King RA, Hertz MI. Bronchoalveolar lavage cell gene expression in acute lung rejection: development of a diagnostic classifier. *Transplantation.* 2008;85(2):224–231.
 29. Morgun A, et al. Molecular profiling improves diagnoses of rejection and infection in transplanted organs. *Circ Res.* 2006;98(12):e74–e83.
 30. Holweg CT, et al. Identification and classification of acute cardiac rejection by intra-graft transcriptional profiling. *Circulation.* 2011;123(20):2236–2243.
 31. Krausgruber T, et al. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. *Nat Immunol.* 2011;12(3):231–238.
 32. Kester AD, Buntinx F. Meta-analysis of ROC curves. *Med Decis Making.* 2000;20(4):430–439.
 33. Jonckheere AR. A distribution-free K-sample test against ordered alternatives. *Biometrika.* 1954;41(1/2):133–145.
 34. Huen NY, et al. Up-regulation of proliferative and migratory genes in regulatory T cells from patients with metastatic castration-resistant prostate cancer. *Int J Cancer.* 2013;133(2):373–382.
 35. Malcolm KC, et al. Mycobacterium abscessus induces a limited pattern of neutrophil activation that promotes pathogen survival. *PLoS One.* 2013;8(2):e57402.
 36. Rapin N, et al. Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. *Blood.* 2014;123(6):894–904.
 37. Senokuchi T, et al. Statins suppress oxidized low density lipoprotein-induced macrophage proliferation by inactivation of the small G protein-p38 MAPK pathway. *J Biol Chem.* 2005;280(8):6627–6633.
 38. Gueller F, et al. Statins attenuate ischemia-reperfusion injury by inducing heme oxygenase-1 in infiltrating macrophages. *Am J Pathol.* 2007;170(4):1192–1199.
 39. Modena BD, et al. Gene expression in biopsies of acute rejection and interstitial fibrosis/tubular atrophy reveals highly shared mechanisms that correlate with worse long-term outcomes. *Am J Transplant.* 2016;16(7):1982–1998.
 40. O'Connell PJ, 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.
 41. Ward C, et al. Endobronchial biopsy and bronchoalveolar lavage in stable lung transplant recipients and chronic rejection. *Am J Respir Crit Care Med.* 1998;158(1):84–91.
 42. Lindeman RH, Merenda PF, Gold RZ. *Introduction to Bivariate Multivariate Analysis.* Glenview, IL; Scott Foresman & Co:1980.
 43. Kruskal W. Relative importance by averaging over orderings. *Am Stat.* 1987;41(1):6–10.
 44. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol.* 1996;58(1):267–288.
 45. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med.* 2016;4(3):213–224.
 46. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med.* 2015;7(287):287ra71.
 47. Jeffrey KL, et al. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol.* 2006;7(3):274–283.
 48. Martinez FO, Gordon S, Locati M, Mantovani A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *J Immunol.* 2006;177(10):7303–7311.
 49. Radom-Aizik S, Zaldivar F, Leu SY, Galassetti P, Cooper DM. Effects of 30 min of aerobic exercise on gene expression in human neutrophils. *J Appl Physiol.* 2008;104(1):236–243.
 50. He F, et al. PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells. *Mol Syst Biol.* 2012;8:624.
 51. Giefing M, et al. Hodgkin-Reed-Sternberg cells in classical Hodgkin lymphoma show alterations of genes encoding the NADPH oxidase complex and impaired reactive oxygen species synthesis capacity. *PLoS ONE.* 2013;8(12):e84928.
 52. Meyers JA, Su DW, Lerner A. Chronic lymphocytic leukemia and B and T cells differ in their response to cyclic nucleotide phosphodiesterase inhibitors. *J Immunol.* 2009;182(9):5400–5411.
 53. Eckert S, et al. Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to Hodgkin lymphoma. *Leukemia.* 2009;23(11):2129–2138.
 54. Stegmann KA, et al. Interferon-alpha-induced TRAIL on natural killer cells is associated with control of hepatitis C virus infection. *Gastroenterology.* 2010;138(5):1885–1897.
 55. Vinh DC, et al. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood.* 2010;115(8):1519–1529.
 56. Ancuta P, et al. Transcriptional profiling reveals developmental relationship and distinct biological functions of CD16+ and CD16- monocyte subsets. *BMC Genomics.* 2009;10:403.
 57. Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell.* 2011;144(2):296–309.
 58. Allantaz F, et al. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS ONE.* 2012;7(1):e29979.
 59. Wu C, et al. Graft-infiltrating macrophages adopt an M2 phenotype and are inhibited by purinergic receptor P2X7 antagonist in chronic rejection. *Am J Transplant.* 2016;16(9):2563–2573.
 60. Frankenberger M, et al. Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. *Eur J Immunol.* 2012;42(4):957–974.
 61. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics.* 2013;14:632.
 62. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc.* 2004;99:909–917.