Application of the Banff Human Organ Transplant Panel to kidney transplant biopsies with features suspicious for antibody-mediated rejection

see commentary on page 423

OPEN

Jack Beadle[1,2,5], Artemis Papadaki[1,5], Frederic Toulza[1], Eva Santos[3], Michelle Willicombe[1,2], Adam McLean[2], James Peters[1] and Candice Roufosse[1,4]

[1]Centre for Inflammatory Disease, Department of Immunology and Inflammation, Faculty of Medicine, Imperial College London, London, UK; [2]Imperial College Renal and Transplant Centre, Imperial College NHS Trust, London, UK; [3]H&I Laboratory, North West London Pathology, London, UK; and [4]Department of Cellular Pathology, North West London Pathology, London, UK

The Banff Classification for Allograft Pathology includes the use of gene expression in the diagnosis of antibody-mediated rejection (AMR) of kidney transplants, but a predictive set of genes for classifying biopsies with 'incomplete' phenotypes has not yet been studied. Here, we developed and assessed a gene score that, when applied to biopsies with features of AMR, would identify cases with a higher risk of allograft loss. To do this, RNA was extracted from a continuous retrospective cohort of 349 biopsies randomized 2:1 to include 220 biopsies in a discovery cohort and 129 biopsies in a validation cohort. The biopsies were divided into three groups: 31 that fulfilled the 2019 Banff Criteria for active AMR, 50 with histological features of AMR but not meeting the full criteria (Suspicious-AMR), and 269 with no features of active AMR (No-AMR). Gene expression analysis using the 770 gene Banff Human Organ Transplant NanoString panel was carried out with LASSO Regression performed to identify a parsimonious set of genes predictive of AMR. We identified a nine gene score that was highly predictive of active AMR (accuracy 0.92 in the validation cohort) and was strongly correlated with histological features of AMR. In biopsies suspicious for AMR, our gene score was strongly associated with risk of allograft loss and independently associated with allograft loss in multivariable analysis. Thus, we show that a gene expression signature in kidney allograft biopsy samples can help classify biopsies with incomplete AMR phenotypes into groups that correlate strongly with histological features and outcomes.

Kidney International (2023) **104,** 526–541; https://doi.org/10.1016/j.kint.2023.04.015

Correspondence: *Jack Beadle, Centre for Inflammatory Disease, Department of Immunology and Inflammation, Faculty of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. E-mail: jlb106@ic.ac.uk*

[5]These authors contributed equally.

Received 10 October 2022; revised 7 March 2023; accepted 14 April 2023; published online 11 May 2023

KEYWORDS: B-HOT; NanoString; pathology; rejection; transcriptomics; transplantation

Lay Summary

The leading cause of transplant loss is antibody-mediated rejection (AMR). This is diagnosed by kidney transplant biopsy according to strict diagnostic criteria, but not all biopsies meet the diagnostic threshold, and in these cases, transplant outcomes are worse than those in biopsies without rejection. We have developed and validated a gene expression model to predict AMR on archived formalin-fixed paraffin embedded kidney transplant biopsies and have applied this to help classify biopsies that do not meet the full criteria for AMR diagnosis. This model can be used to identify biopsies at a higher risk of graft loss.

Antibody-mediated rejection (AMR) is the leading cause of late kidney transplant loss.[1] This occurs when donor-specific antibodies (DSAs) are generated that bind to the endothelium, triggering transcriptomic and phenotypic changes, activation of the complement cascade, and inflammatory cell recruitment.

The standard of care for diagnosing kidney transplant pathology is histological assessment of biopsies. The Banff Classification of Allograft Pathology was developed to standardize and grade histological lesions in biopsies[2] and requires 3 elements to diagnose AMR: criterion 1: histological features of AMR; criterion 2: evidence of the interaction between DSA and the endothelium; and criterion 3: evidence of DSA.[3,4] However, histology has limitations: biopsies are prone to sampling error, potentially leading to "false-negative" diagnoses; in addition, there is significant interobserver variability with grading of histology.[5,6] Biopsies with histological features of AMR, but without other criteria, have incomplete phenotypes and cannot be diagnosed as AMR.

The Banff classification was modified to include the use of "thoroughly validated" transcripts in the biopsy that are "associated with AMR" as a surrogate for either evidence of DSA interaction (criterion 2) or in place of evidence for DSA (criterion 3).[3]

There is no "gold standard" for diagnostic transcript expression in biopsies. Discovery studies have consistently identified groups of genes that are differentially expressed in transplant biopsies with rejection pathologies[7] and that biopsies with indistinguishable histological features have different transcript profiles that correlate with outcomes.[8] In biopsies with AMR, identification of "pathogenesis-based transcripts,"[9] such as endothelial-cell transcripts, DSA-specific transcripts, and natural killer–associated transcripts, have influenced our understanding of the pathophysiology[10–12] and have been described in different platforms, such as microarray,[9–15] quantitative polymerase chain reaction,[16,17] and NanoString.[18,19]

High-throughput gene expression analysis can be performed on formalin-fixed paraffin-embedded (FFPE) biopsy material leftover from routine histological diagnosis by using multiplexed color-coded probe-based gene expression analysis with NanoString nCounter technology.[18,20] NanoString has a similar sensitivity to quantitative polymerase chain reaction[18,20] and microarray,[21] both of which require a separate core of biopsy tissue in an RNA preservative. Use of leftover FFPE tissue for molecular analysis carries no additional risks for patients and permits the use of archived samples for investigation. As analysis is performed on the same tissue that histological grading has been performed on, direct comparisons between histological features and transcript expression can be performed. Studies show that NanoString has a strong correlation between gene expression and histological features on allograft biopsies than does quantitative polymerase chain reaction.[20]

To enable the validation of comparable gene expression assays across multiple centers for potential clinical use, the Banff Molecular Diagnostics Working Group identified a list of validated transplant-related genes of interest from a literature search of peer-reviewed microarray and NanoString publications on transplanted organs. In conjunction with NanoString, they developed the Banff-Human Organ Transplant (B-HOT) codeset of 758 transplant-related genes of interest and 12 internal reference genes[22] (Supplementary Table S1).

The aims of this study were to (i) identify an AMR-associated gene score derived from the B-HOT Panel by comparing biopsies with and without AMR in a retrospective cohort of kidney transplant biopsies and (ii) to test the context of use for the gene score as defined in the most recent Banff classification by reclassifying biopsies with incomplete phenotypes for AMR.

METHODS

Study cohort

We assembled a continuous retrospective cohort of indication adult kidney allograft biopsies performed at Imperial College between March 1, 2016 and December 31, 2017. Patients who were transplanted outside the United Kingdom or had no available tissue-typing data were excluded.

Kidney allograft biopsies were obtained using an 18-G spring-loaded biopsy needle under ultrasound guidance. Human samples used in this project (R14094) were obtained from the Imperial College Healthcare Tissue and Biobank, approved by Wales REC 3 to release human material for research (22/WA/2836). Patients were followed up until April 2022.

Histopathological features

Biopsies were classified according to the 2019 Banff criteria[4] by using histological and serological assessment but excluding the results of gene expression analysis. Biopsies were divided into 3 groups on the basis of the fulfilment of the Banff criteria for active/chronic active AMR.

Biopsies were classified as "AMR" if they fulfilled all criteria for active or chronic active AMR: criterion 1: histological evidence of tissue injury; criterion 2: antibody interaction with the endothelium; and criterion 3: evidence of circulating DSA, or C4d staining. Biopsies that did not meet criterion 1 were defined as "No-AMR." Biopsies with C4d deposition without any other features of rejection were included in the No-AMR group.

Cases that met criterion 1 for active or chronic active AMR only, but no additional criteria, were categorized as suspicious for AMR ("Susp-AMR").

Histological evidence of tissue injury was defined as glomerulitis (g) or peritubular capillaritis (ptc) $\geq$1 (in the absence of glomerulonephritis for g and of borderline or T cell–mediated rejection [BLTCMR or TCMR] or infection for ptc); evidence of intimal arteritis (v) $\geq$1; and/or thrombotic microangiopathy (TMA) on light microscopy in glomeruli or arterioles. Acute tubular injury on its own was not considered a criterion for AMR. Chronic active AMR required the presence of glomerular capillary double contours (cg) $\geq$1a or peritubular capillary basement membrane multilayering (>7 layers of capillary basement membrane multilayering in at least 1 peritubular capillary, plus $\geq$5 layers in at least 2 others on electron microscopy). Electron microscopy was not performed on all biopsies routinely.

DSA detection

DSAs were assessed using LABScreen mixed beads (One Lambda, Inc.), and, if positive, the anti–human leukocyte antigen (HLA) antibody specificity was identified using LABScreen single antigen beads. Patients in whom an antibody was detected in $\geq$1 sample, with a mean fluorescence intensity of >500, were considered DSA-positive. Patients with detectable DSA within 3 months prebiopsy or 1-month postbiopsy were classed as DSA-positive at biopsy.

RNA extraction and quantification

Six consecutive 12 μm paraffin curls were cut from each FFPE block using a Leica Microtome and placed into 1.5 ml RNase-free Eppendorf tubes. Microtome blades were replaced between blocks and equipment cleaned with RNaseZap (ThermoFisher). RNA was extracted from paraffin curls using the Rneasy FFPE kit (Qiagen) and deparaffinization solution (Qiagen), according to the manufacturer's instructions, and eluted in the minimum volume of RNase-free water. RNA purity and concentration were measured using a NanoDrop 2000c spectrophotometer (ThermoFisher).

NanoString gene expression preprocessing

The B-HOT codeset[22] was used to analyze gene expression in FFPE biopsy samples using NanoString technology. The codeset comprises reporter and capture probes that hybridize to target sequences of RNA. One hundred nanograms of RNA from each biopsy were added to the NanoString codeset in hybridization buffer and incubated at 65 °C for 18 hours. Gene expression was quantified using the nCounter SPRINT Analysis System (NanoString Technologies).

Initial quality control assessment was performed using the default settings in nSolver 4.0 Analysis Software (NanoString Technologies). Samples flagged for quality control were not included in normalization. Raw gene counts were processed and normalized using the R Bioconductor package NanoStringQCPro.[23] Digital counts were standardized using positive-control probes and normalized using the geometric mean of 12 internal reference genes. Log2 normalized counts were used for gene analyses. Relative log expression plots[24] of all the gene expression profiles in all biopsies were performed, after normalization, to ensure that there was no unwanted variation between samples or batches (Supplementary Figure S1).

Statistical analysis

Statistical analysis was performed using R version 4.1.1[25] and GraphPad Prism 9.0 (GraphPad Software).

Differential gene expression. We compared gene expression in AMR with No-AMR biopsies using linear regression (lm function in the stats package[25]). To account for multiple testing, *P* values were adjusted with the Benjamini-Hochberg method and statistical significance was defined as a false discovery rate threshold of <0.05.

Differential pathway expression. We used the gene set variation analysis method (GSVA R package)[26] to identify pathways that were differentially up- or downregulated in AMR and No-AMR biopsies. Gene sets were defined using Hallmark pathways (http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). We used the same modeling approach and significance thresholds as at the gene level.

Supervised learning. For development and validation of an AMR molecular classifier, the cohort was randomized 2:1 into a discovery and a validation group. The expression of 770 genes in the B-HOT Panel, in addition to age at biopsy, time from transplant to biopsy, and sex as covariates, were used to train both a molecular LASSO (Least Absolute Shrinkage and Selection Operator) and a Random Forest classifier to discriminate AMR biopsies from No-AMR biopsies using the glmnet[27] and randomForest[28] packages, respectively, within the framework of the caret package.[29] After stratified splitting of the discovery cohort into training and testing sets (73.5%:26.5%), we created 200 random data partitions and trained the LASSO and Random Forest models using leave-one-group-out cross-validation. For the Random Forest model, given that as the "mtry" parameter increases, the model performance metrics become better but the risk of model overfitting increases, we kept "mtry" constant using the "rule of thumb" mtry = sqrt(number of predictors − 1). The LASSO model was tuned using the F1 metric to address the class imbalance within our data, as there were fewer cases with AMR than with No-AMR. To avoid overfitting, the most parsimonious model with lambda within 1 SE of the optimal model was selected. The molecular classifier was then validated in the validation cohort (Supplementary Figure S2).

RESULTS

Cohort selection and RNA extraction

A total of 611 indication transplant biopsies performed at our center were identified (Figure 1), of which 585 biopsies from 445 patients met the inclusion criteria. For patients with multiple biopsies, only the first biopsy with histological features of rejection (AMR, TCMR, or BLTCMR) or the first biopsy chronologically (if there was no rejection) was included.

Sufficient RNA was isolated from 378 of 428 available FFPE blocks. The mean RNA concentration was 48.64 ng/μl (range, 7.2–225 ng/μl), and the mean 260/280 absorbance ratio was 1.96 (range, 1.6–2.4). The NanoString B-HOT Panel

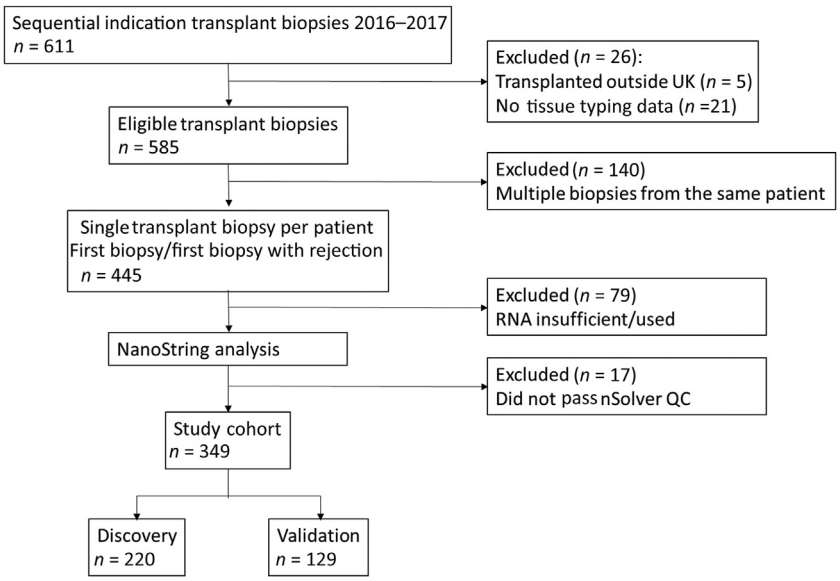


Figure 1 | Flow diagram of cohort selection. This figure shows the steps used in the selection of the biopsies used in this study. A single biopsy per patient was selected. All biopsies were randomized into either the discovery or the validation cohort.

Standard, which has synthetic targets for each of the test probes, was run 5 times to test technical reproducibility; this showed a good correlation ($r = 0.999$). Four biological replicates were included, which showed correlations of $r = 0.977$ to 0.986.

After quality control and normalization, gene expression analysis was available in 349 biopsies, of which 220 were randomized to the discovery cohort and 129 to the validation cohort.

Baseline characteristics

The relevant clinical characteristics of the entire cohort are presented in Table 1. There were no statistically significant differences between the discovery and validation cohorts.

Among the biopsies randomized to the discovery cohort, 168 (76%) had No-AMR, 32 (14%) were Susp-AMR, and 21 (9.5%) met the full criteria for active AMR. In the validation cohort, 101 (78%) biopsies had No-AMR, 18 (14%) were Susp-AMR, and 10 biopsies (7.8%) had AMR. AMR biopsies in the validation cohort were taken earlier post-transplantation than those in the discovery cohort (median 542 days vs. 2039 days; $P = 0.1$), but this did not meet statistical significance.

There were no significant differences in the overall histological features between the discovery and validation cohorts (Supplementary Table S2).

Histological characteristics

All biopsies were divided into 1 of 3 groups: AMR, No-AMR, and Susp-AMR, the histological features of which are summarized in Table 2.

The AMR group had significantly higher scores than did the others for C4d deposition and g, ptc, v, and cg as well as higher scores for interstitial inflammation and tubulitis. One biopsy with TMA alone was included in the AMR group; this case had C4d (fulfilling criteria 2 and 3) but was from an ABO-incompatible recipient. Susp-AMR biopsies had higher scores for g, ptc, and cg than did No-AMR biopsies. Scores for chronic lesions other than cg (interstitial fibrosis [ci], tubular atrophy [ct], and vascular fibrous intimal thickening [cv]) were not different between the groups, nor was the degree of interstitial fibrosis and tubular atrophy.

Of the 50 Susp-AMR cases fulfilling criterion 1, 40 were included because of a microvascular inflammation (MVI = g + ptc) score of >0; 8 because of TMA alone, and 2 because of a v score of >1 alone. Four patients had detectable DSA at biopsy, but none of these met criterion 2. Most cases, therefore, were DSA-negative.

Among No-AMR cases, 25 had BLTCMR and 26 had TCMR (13 of these also had pyelonephritis or BK virus nephropathy). Two biopsies had chronic (inactive) AMR, without meeting criterion 2 or 3. Eighteen biopsies had C4d deposition, of which 13 came from ABO-compatible donors. Ten of 13 showed C4d deposition without evidence of rejection, 2 had TCMR only (with no DSA), and 1 biopsy was DSA-positive, had MVI ≥2, but had concomitant TCMR and glomerulonephritis, so neither the ptc nor the g score was counted toward a histological diagnosis of AMR.

Clinical features at the time of biopsy

Clinical features at the time of biopsy are summarized in Table 3. Biopsies in the AMR group were taken at a later stage post-transplantation (1097 days) than those in the No AMR and Susp-AMR groups (641 and 621 days), but this was not statistically significant. Most biopsies were performed to investigate rising creatinine, but there were significantly more biopsies performed for an elevated protein-creatinine ratio in the AMR (16%) and Susp-AMR (22%) groups than in the No-AMR (8.6%) group. In addition, more biopsies in the AMR group were performed in the context of *de novo* DSA (13%).

The median serum creatinine and estimated glomerular filtration rate (as calculated using the Modification of Diet in Renal Disease formula) were not significantly different between groups; however, there was heavier proteinuria in the AMR and Susp-AMR groups than in the No-AMR group.

Most patients were maintained on tacrolimus monotherapy. The proportion of patients on either maintenance antiproliferative agents or steroids before biopsy was higher in the AMR group. Forty-eight percent of patients in the AMR group had received prior augmented immunosuppression compared with 19% in the No-AMR group and 26% in the Susp-AMR group.

HLA-DSAs were detected in 77% patients in the AMR group at the time of biopsy compared with 12% in the No-AMR group and 8% in the Susp-AMR group; their characteristics are provided in Table 4.

Differential gene expression analysis comparing AMR and No-AMR

We performed differential gene expression analysis between biopsies with AMR (AMR group) and those in the No-AMR group in both the discovery and validation cohorts (Supplementary Figure S3A and B); 391 genes were differentially expressed in the discovery cohort (369 upregulated and 22 downregulated) and 28 genes in the validation cohort.

The effect estimates in differential gene expression correlated well between the discovery and validation cohorts ($r = 0.622$; $P = 0.0001$), even though fewer genes in the validation cohort met the significance threshold (5% false discovery rate), suggesting that the smaller sample size in the latter affected the statistical power to detect differential expression (Supplementary Figure S4).

We used gene set variation analysis to explore differential pathway expression between AMR and No-AMR biopsies in the combined discovery and validation cohorts. In biopsies with AMR, we identified expression of pathways associated with allograft rejection, interferon-γ and -α signaling, complement activation, and inflammation (Supplementary Figure S5).

Selecting predictors for AMR

We next sought to develop a classifier to predict the diagnosis of AMR on the basis of biopsy gene expression. Many genes associated with AMR in the B-HOT Panel share common pathways with correlated expression. Thus, the

Table 1 | Baseline characteristics of transplant recipients in the discovery and validation cohorts

Characteristic	Overall (*N* = 349)	Discovery (*n* = 220)	Validation (*n* = 129)	*P*[a]
AMR group				0.8
AMR	31 (8.9)	21 (9.5)	10 (7.8)	
No AMR	268 (77)	167 (76)	101 (78)	
Susp-AMR	50 (14)	32 (15)	18 (14)	
Age at transplant, yr	50.00 (37.00–57.00)	48.00 (36.75–55.25)	52.00 (38.00–57.00)	0.070
Sex				0.6
Female	139 (40)	90 (41)	49 (38)	
Male	210 (60)	130 (59)	80 (62)	
Ethnicity				0.9
African/Caribbean	62 (18)	38 (17)	24 (19)	
Caucasian	140 (40)	92 (42)	48 (37)	
Indian Asian	103 (30)	63 (29)	40 (31)	
Other	44 (13)	27 (12)	17 (13)	
Cause of kidney failure				0.3
APKD	34 (9.7)	17 (7.7)	17 (13)	
Diabetes	71 (20)	44 (20)	27 (21)	
GN	98 (28)	70 (32)	28 (22)	
HTN	15 (4.3)	8 (3.6)	7 (5.4)	
Other	27 (7.7)	15 (6.8)	12 (9.3)	
Urological	23 (6.6)	13 (5.9)	10 (7.8)	
Unknown	81 (23)	53 (24)	28 (22)	
Transplant type				0.3
Deceased donor	204 (58)	123 (56)	81 (63)	
Living donor	121 (35)	81 (37)	40 (31)	
ABO-incompatible	10 (2.9)	5 (2.3)	5 (3.9)	
Simultaneous pancreas-kidney	14 (4.0)	11 (5.0)	3 (2.3)	
Graft function				0.5
Delayed	68 (22)	42 (21)	26 (24)	
Immediate	234 (77)	153 (77)	81 (76)	
Primary dysfunction	3 (1.0)	3 (1.5)	0 (0)	
Not recorded	44	22	22	
Induction immunosuppression				0.3
Basiliximab	17 (5.2)	13 (6.4)	4 (3.3)	
Campath	292 (90)	183 (90)	109 (91)	
Campath/rituximab	3 (0.9)	3 (1.5)	0 (0)	
Daclizumab	8 (2.5)	4 (2.0)	4 (3.3)	
Daclizumab/rituximab	3 (0.9)	1 (0.5)	2 (1.7)	
Methylprednisolone	1 (0.3)	0 (0)	1 (0.8)	
Not recorded	25	16	9	
HLA-A MM				0.8
0	67 (19)	44 (20)	23 (18)	
1	157 (46)	96 (44)	61 (48)	
2	121 (35)	78 (36)	43 (34)	
Unknown	4	2	2	
HLA-B MM				0.6
0	49 (14)	34 (16)	15 (12)	
1	211 (61)	131 (60)	80 (63)	
2	85 (25)	53 (24)	32 (25)	
Unknown	4	2	2	
HLA-DR MM				0.7
0	120 (35)	79 (36)	41 (32)	
1	141 (41)	89 (41)	52 (41)	
2	84 (24)	50 (23)	34 (27)	
Unknown	4	2	2	
Total mismatch number	3 (2–4)	3 (2–4)	3 (2–4)	0.3
0	25 (7.2)	18 (8.3)	7 (5.5)	
1	15 (4.3)	13 (6.0)	2 (1.6)	
2	68 (20)	43 (20)	25 (20)	
3	100 (29)	61 (28)	39 (31)	
4	80 (23)	44 (20)	36 (28)	
5	30 (8.7)	20 (9.2)	10 (7.9)	
6	27 (7.8)	19 (8.7)	8 (6.3)	
Unknown	4	2	2	
Previous transplant (yes)	49 (14)	31 (14)	18 (14)	>0.9

AMR, antibody-mediated rejection; APKD, autosomal dominant polycystic kidney disease; GN, glomerulonephritis; HLA, human leukocyte antigen; HTN, hypertension; MM, HLA mismatch; Susp-AMR, suspicious for antibody-mediated rejection.
[a]Pearson χ^2 test, Wilcoxon rank-sum test, and Fisher exact test.
Data are expressed as median (interquartile range) or n (%).

Table 2 | Histological characteristics of biopsies in the AMR, No AMR, and Susp-AMR groups

		AMR group				
Variable	Overall (N = 349)	AMR (n = 31)	No AMR (n = 268)	Susp-AMR (n = 50)	P (all groups)[a]	P (AMR vs. Susp)[a]
DSA at biopsy					<0.001	<0.001
Yes	58 (17)	24 (77)	30 (11)	4 (8.0)		
No	282 (81)	7 (23)	229 (86)	46 (92)		
Unknown	9 (2)	0 (0)	9 (3.0)	0 (0)		
g score		1 (0–2)	0 (0–0)	1 (0–1)	<0.001	0.061
0	294 (84)	8 (26)	263 (98)	23 (46)		
1	37 (11)	12 (39)	4 (1.5)	21 (42)		
2	14 (4.0)	8 (26)	1 (0.4)	5 (10)		
3	4 (1.1)	3 (9.7)	0 (0)	1 (2.0)		
ptc score		1 (1–2)	0 (0–0)	0 (0–2)	<0.001	0.020
0	286 (82)	6 (19)	254 (95)	26 (52)		
1	21 (6.0)	10 (32)	1 (0.4)	10 (20)		
2	37 (11)	13 (42)	11 (4.1)	13 (26)		
3	5 (1.4)	2 (6.5)	2 (0.7)	1 (2.0)		
v score		0 (0–0)	0 (0–0)	0 (0–0)	<0.001	0.74
0	330 (98)	26 (87)	260 (100)	44 (94)		
1	3 (0.9)	2 (6.7)	0 (0)	1 (2.1)		
2	2 (0.6)	1 (3.3)	0 (0)	1 (2.1)		
3	2 (0.6)	1 (3.3)	0 (0)	1 (2.1)		
C4d score		2 (0–3)	0 (0–0)	0 (0–0)	<0.001	<0.001
0	288 (85)	12 (39)	230 (88)	46 (94)		
1	16 (4.7)	2 (6.5)	12 (4.6)	3 (6)		
2	15 (4.4)	6 (19)	8 (3.1)	0 (0)		
3	21 (6.2)	11 (35)	10 (3.8)	0 (0)		
cg score		0 (0–2)	0 (0–0)	0 (0–0.25)		0.11
0	305 (88)	14 (48)	254 (95)	37 (74)		
1a	10 (2.9)	1 (3.4)	8 (3.0)	1 (2.0)		
1b	14 (4.0)	4 (14)	5 (1.9)	5 (10)		
2	2 (0.6)	2 (6.9)	0 (0)	0 (0)		
3	12 (3.5)	6 (21)	1 (0.4)	5 (10)		
ci score		1 (1–2)	1 (0–2)	1 (0.75–2)	0.32	0.54
0	111 (32)	6 (20)	93 (35)	12 (24)		
1	122 (35)	15 (50)	90 (33)	17 (34)		
2	80 (23)	6 (20)	58 (22)	16 (32)		
3	36 (10)	3 (10	28 (10)	5 (10)		
ct score		1 (1–2)	1 (1–2)	1 (1–2)	0.69	0.63
0	79 (23)	5 (17)	65 (24)	9 (18)		
1	154 (44)	16 (53)	118 (44)	20 (40)		
2	80 (23)	6 (20)	58 (22)	16 (32)		
3	36 (10)	3 (10)	28 (10)	5 (10)		
cv score		1 (0–2)	1 (1–2)	1 (1–2)	0.40	0.80
0	73 (22)	9 (30)	53 (20)	11 (23)		
1	122 (36)	10 (33)	92 (35)	20 (43)		
2	100 (30)	9 (30)	77 (30)	14 (30)		
3	42 (12)	2 (6.7)	38 (15)	2 (4.3)		
i score		1 (0–2)	0 (0–0)	0 (0–1)	<0.001	0.12
0	260 (75)	14 (47)	212 (79)	34 (68)		
1	48 (14)	5 (17)	34 (13)	9 (18)		
2	17 (4.9)	5 (17)	8 (3.0)	4 (8.0)		
3	23 (6.6)	6 (20)	14 (5.2)	3 (6.0)		
t score		1 (0–2)	0 (0–1)	0 (0–1.25)	0.007	0.40
0	236 (68)	13 (43)	194 (72)	29 (58)		
1	51 (15)	9 (30)	33 (12)	9 (18)		
2	39 (11)	5 (17)	24 (9.0)	10 (20)		
3	22 (6.3)	3 (10)	17 (6.3)	2 (4.0)		
ti score		1 (1–3)	1 (0–1)	1 (0–2)		0.087
0	152 (44)	3 (10)	134 (50)	15 (30)		
1	116 (33)	13 (43)	82 (30)	21 (42)		
2	45 (13)	6 (20)	30 (11)	9 (18)		
3	36 (10)	8 (27)	23 (8.6)	5 (10)		
ah score		0 (0–2)	0 (0–2)	0.5 (0–2)	0.47	0.84
0	181 (52)	16 (53)	140 (52)	25 (50)		
1	63 (18)	3 (10)	51 (19)	9 (18)		
2	59 (17)	6 (20)	46 (17)	7 (14)		

(*Continued on following page*)

Table 2 | (Continued) **Histological characteristics of biopsies in the AMR, No AMR, and Susp-AMR groups**

Variable	Overall (*N* = 349)	AMR group			*P* (all groups)[a]	*P* (AMR vs. Susp)[a]
		AMR (*n* = 31)	No AMR (*n* = 268)	Susp-AMR (*n* = 50)		
3	43 (12)	5 (17)	30 (11)	8 (16)		
8	1 (0.3)	0 (0)	0 (0)	1 (2.0)		
TMA	13 (0.37)	1 (0.03)[b]	0 (0)	12 (24)	<0.0001	0.0136
PTCML	11 (7.2)	3 (14)	5 (5.2)	3 (8.6)	0.24	0.66
IFTA (%)	15 (5–30)	15 (10–30)	15 (5–30)	20 (10–30)	0.21	0.38
Number of glomeruli	13 (8–18)	12 (8–18)	13 (8–17)	14 (8–18)	0.96	0.88
Number of obsolete glomeruli	1.00 (0.00–3.00)	1.00 (0.00–3.00)	1.00 (0.00–3.00)	1.00 (0.00–3.75)	0.76	0.91
Primary rejection diagnosis[c]						
Active AMR	18 (5.2)	18 (58)	0 (0)	0 (0)		
BLTCMR	33 (9.5)	0 (0)	26 (9.7)	7 (14)		
C4d WER	10 (2.9)	0 (0)	10 (3.7)	0 (0)		
Chronic (active) AMR	9 (2.6)	9 (29)	0 (0)	0 (0)		
Chronic (inactive) AMR	6 (1.7)	0 (0)	3 (1.1)	3 (6.0)		
Mixed	4 (1.1)	4 (13)	0 (0	0 (0)		
No rejection	237 (68)	0 (0)	202 (75)	35 (70)		
TCMR	32 (9.2)	0 (0)	27 (10)	5 (10)		

ah, arteriolar hyalinosis; AMR, antibody-mediated rejection; BLTCMR, borderline or T cell–mediated rejection; C4d WER, C4d deposition without evidence of rejection; cg, glomerular capillary double contours; ci, interstitial fibrosis; ct, tubular atrophy; cv, vascular fibrous intimal thickening; DSA, donor-specific antibody; g, glomerulitis; i, interstitial inflammation; IFTA, interstitial fibrosis and tubular atrophy; ptc, peritubular capillaritis; PTCML, peritubular capillary basement membrane multilayering; Susp-AMR, suspicious for antibody-mediated rejection; t, tubulitis; TCMR, T cell–mediated rejection; ti, total inflammation; TMA, thrombotic microangiopathy; v, intimal arteritis.
[a]Fisher exact test and Kruskal-Wallis rank-sum test.
[b]One biopsy with TMA alone was included in the validation cohort. It met criteria 2 and 3 of the Banff classification, but came from an ABO-incompatible donor.
[c]Morphological classification of rejection as defined by the most recent Banff classification, without the use of a molecular classifier as a surrogate for criterion 2 or 3 for the diagnosis of AMR. Biopsies that fulfill features of both TCMR and AMR are labeled as mixed rejection. Biopsies that do not meet the Banff classification of AMR and have no other morphological features that meet the diagnostic threshold for TCMR or BLTCMR, are defined as "No rejection."
Data are expressed as median (interquartile range) or n (%).

naive approach of selecting the most significant differentially expressed genes will not necessarily provide the best predictive performance. We therefore performed supervised learning with 2 different approaches: LASSO regression and Random Forest. Comparison of the 2 approaches revealed that LASSO outperformed Random Forest in terms of accuracy (Supplementary Table S3). Moreover, LASSO has additional advantages in relation to developing a classifier

Table 3 | Clinical features at the time of biopsy in the AMR, No AMR, and Susp-AMR groups

Characteristic	Overall (*N* = 349)	AMR (*n* = 31)	No AMR (*n* = 268)	Susp-AMR (*n* = 50)	*P*[a]
Time of biopsy, d	653 (127–2318)	1097 (301.5–3198.5)	638 (89–2215)	621 (238–1923.25)	0.2
Indication for biopsy					
Delayed graft function	23 (6.6)	0 (0)	22 (8.2)	1 (2.0)	0.11
Creatinine rise	307 (88)	28 (90)	236 (88)	43 (86)	0.8
Proteinuria	39 (11)	5 (16)	23 (8.6)	11 (22)	0.016
De novo DSA	15 (4.3)	4 (13)	8 (3.0)	3 (6.0)	0.031
Immunosuppression switch	4 (1.1)	0 (0)	3 (1.1)	1 (2.0)	0.7
Features at biopsy					
Serum creatinine, μmol/l	186 (146–240)	195.00 (149.5–222.5)	190 (148.75, 242.25)	164 (123–243.5)	0.2
MDRD eGFR, ml/min per 1.73 m^2	33 (23–43)	33.00 (26.00–44.00)	31 (23–41)	37 (23.25–51.5)	0.12
Protein/creatinine ratio, mg/mmol	35 (<20–108.00)	100.00 (<20–215.00)	28.00 (<20–80.50)	65.00 (<20–180.00)	0.004
Maintenance immunosuppression					
Maintenance CNI	348 (100)	30 (97)	268 (100)	50 (100)	0.089
Maintenance antiproliferative	127 (36)	19 (61)	92 (34)	16 (32)	0.010
Maintenance steroid	83 (24)	15 (48)	56 (21)	12 (24)	0.003
Treatment of rejection before biopsy	78 (22)	15 (48)	50 (19)	13 (26)	<0.001
Postbiopsy treatment					<0.001
No change	225 (64)	7 (23)	202 (75)	16 (32)	
Introduction of an antiproliferative agent		9 (29)	27 (10)	16 (32)	<0.001
Introduction of steroid		4 (13)	12 (4)	8 (16)	0.009
Initiation of plasma exchange		8 (26)	8 (3)	3 (6)	<0.001
Other postbiopsy treatment[b]		4 (13)	7 (3)	3 (6)	0.016

AMR, antibody-mediated rejection; CNI, calcineurin inhibitor; DSA, donor-specific antibody; eGFR, estimated glomerular filtration rate; MDRD, Modification of Diet in Renal Disease; Susp-AMR, suspicious for antibody-mediated rejection.
[a]Kruskal-Wallis rank-sum test, Fisher exact test, and Pearson $\chi^{2\ \text{test}}$.
[b]Other recorded treatments at biopsy include rituximab (1 patient in the AMR group, 5 patients in the No AMR group, 2 patients in Susp-AMR group), anti-thymocyte globulin (3 patients in the AMR group, all with mixed rejection, and 1 patient each in the No AMR and Susp-AMR groups, respectively), and alemtuzumab (1 patient in the No AMR group).
Data are expressed as median (interquartile range) or n (%).

Table 4 | DSA status at biopsy

Characteristic	AMR (n = 31)	No AMR (n = 268)	Susp-AMR (n = 50)	P[a]
DSA subtype				>0.9
De novo DSA	22 (81)	35 (80)	6 (86)	
Preformed DSA	5 (19)	9 (20)	1 (14)	
No DSA	4	224	43	
DSA history				
DSA at the time of biopsy	24 (77)	30 (12)	4 (8.0)	<0.001
Historical DSA only[b]	3 (9.7)	13 (4.9)	3 (6.0)	0.4
DSA class				0.8
I	8 (30)	16 (37)	2 (29)	
II	14 (52)	22 (51)	5 (71)	
I/II	5 (19)	5 (12)	0 (0)	
Immunodominant DSA class				0.8
I	9 (35)	17 (40)	2 (29)	
II	17 (65)	26 (60)	5 (71)	
Immunodominant DSA specificity				0.9
HLA-A	5 (21)	4 (13)	1 (25)	
HLA-B	2 (8.3)	3 (10)	1 (25)	
HLA-C	0 (0)	3 (10)	0 (0)	
HLA-DP	2 (8.3)	3 (10)	0 (0)	
HLA-DQ	13 (54)	15 (50)	2 (50)	
HLA-DR	2 (8.3)	2 (6.7)	0 (0)	
Immunodominant DSA mean fluorescence intensity	9150 (4825–12,825)	2700 (1350–5250)	1400 (1250–3350)	<0.001

AMR, antibody-mediated rejection; DSA, donor-specific antibody; HLA, human leukocyte antigen; Susp-AMR, suspicious for antibody-mediated rejection.
[a]Fisher exact test, Pearson χ^2 test, and Kruskal-Wallis rank-sum test.
[b]Previously detected DSA no longer detectable at the time of biopsy.
Data are expressed as median (interquartile range) or n (%).

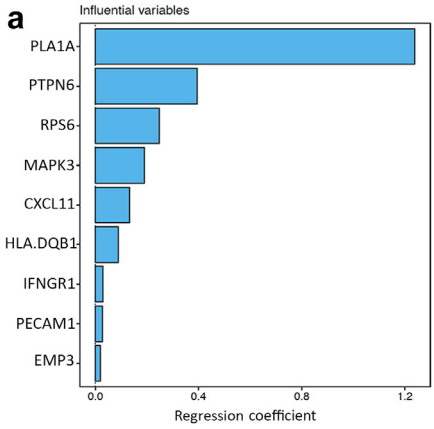
that might ultimately be translated into a clinical tool. LASSO enforces model sparsity, leading to a relatively small number of predictors and, in the case of correlated predictors, will only select 1, thus avoiding redundancy.

In our training set, age at biopsy, recipient sex, and time from transplant to biopsy were included as covariates, but none of these were selected by the LASSO algorithm for inclusion in the final predictive model.

LASSO regression in the discovery cohort selected 9 genes as predictors of diagnosis of AMR (*PLA1A*, *PTPN6*, *RPS6*, *MAPK3*, *CXCL11*, *HLA-DQB1*, *IFNGR1*, *PECAM1*, and *EMP3*; Figure 2a). All genes but *RPS6* were significantly (5% false discovery rate) upregulated in the differential gene expression analysis of AMR versus No-AMR biopsies (Supplementary Figure S6). We chose a stringent AMR cutoff to minimize the misclassification of No-AMR biopsies and then tested the model in our independent validation cohort. The model (Figure 2b) had an accuracy of 92% (range, 0.85–0.96), a specificity of 0.98, and a sensitivity of 0.3 in the validation cohort.

Histological associations of transcriptomic predictors

The correlation between individual genes and histological features on biopsy is described in Supplementary Table S4. Across the entire cohort, the model-derived score correlated with histological features of AMR, including ptc (r = 0.59; P < 0.0001), g (r = 0.59; P < 0.0001), cg (r = 0.35; P <



b

Confusion matrix –

	No	Yes
No	99	7
Yes	2	3

Accuracy (95% CI)	0.9189	(0.8517, 0.9623)
Kappa	0.36175	
Sensitivity	0.9802	
Specificity	0.3	

Figure 2 | Generation and performance of the predictive model. (**a**) Genes identified as predictors of antibody-mediated rejection from Least Absolute Shrinkage and Selection Operator regression, organized by regression coefficient (x axis) in descending order. (**b**) Confusion matrix of the predictions made by the proposed model when tested in the validation cohort. References classes were classified according to the Banff Classification of Allograft Pathology.

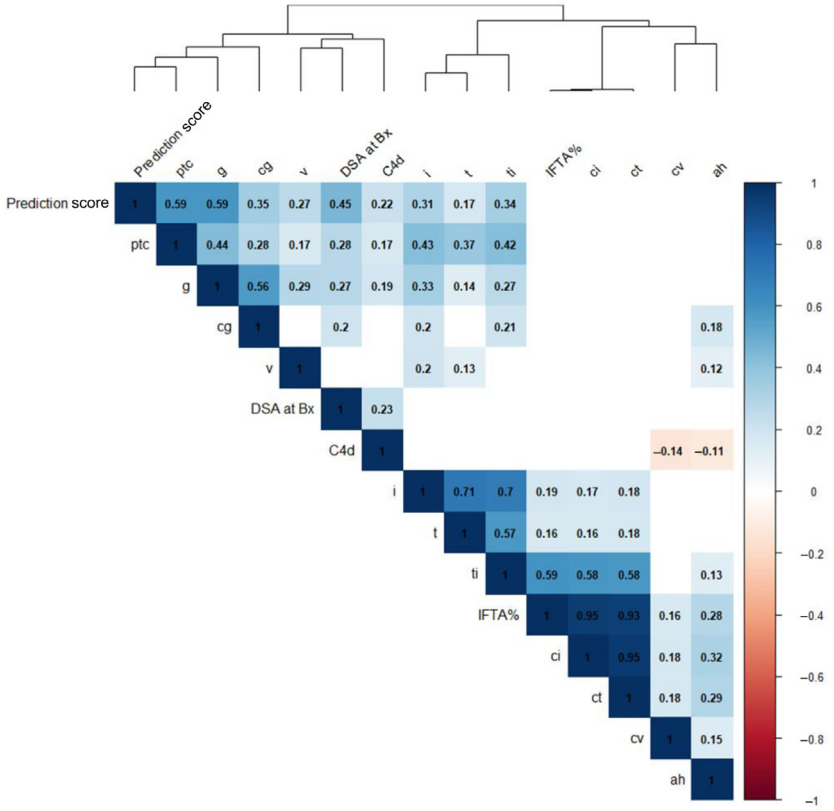


Figure 3 | Correlation matrix (Pearson *r*) between the antibody-mediated rejection molecular classifier score and the grade of histological lesions on biopsy. Results are ordered by *k*-means clustering. Numbers represent Pearson *r* coefficients. Squares are colored by strength and direction of the correlation. Where the correlation does not meet significance, squares are left blank. ah, arteriolar hyalinosis; Bx, kidney allograft biopsy; cg, glomerular capillary double contours; ci, interstitial fibrosis; ct, tubular atrophy; cv, vascular fibrous intimal thickening; DSA, donor-specific antibody; g, glomerulitis; i, interstitial inflammation; IFTA%, degree of interstitial fibrosis and tubular atrophy; ptc, peritubular capillaritis; t, tubulitis; ti, total inflammation; v, intimal arteritis.

0.0001), and C4d ($r = 0.22$; $P < 0.0001$; Figure 3). There was a correlation with other features of inflammation—tubulitis ($r = 0.18$; $P = 0.003$) and interstitial inflammation ($r = 0.31$; $P < 0.0001$)—but clustering demonstrated that the gene score associated more closely with AMR lesions than with TCMR lesions or inflammation. The AMR gene score was higher in biopsies with histological features of AMR than in those without ($P < 0.00001$; Figure 4); gene scores increased with increasing MVI. The AMR score was significantly different in biopsies with different phenotypes of rejection (Figure 5), compared with AMR.

AMR prediction for suspicious cases

Fifty biopsies in the Susp-AMR group were assessed using a model consisting of the 9 predictor genes identified by LASSO to extract the probability of AMR. Using this approach, 6 biopsies were reclassified as "AMR" and 44 biopsies were reclassified as "No-AMR."

The differences between the biopsies are outlined in Table 5. There were no significant differences between the biopsies in terms of sensitization, previously treated episodes of rejection, or maintenance immunosuppression. Those biopsies with a high AMR probability had a lower median estimated glomerular filtration rate (15.5 ml/min per 1.73 m^2) than did low-probability biopsies (38.5 ml/min per 1.73 m^2; $P = 0.045$), but there were no differences in proteinuria or the presence of detectable HLA-DSA.

High-probability biopsies had significantly higher ptc scores ($P = 0.002$), contributing to a higher median MVI score ($P = 0.01$), and higher cg scores ($P = 0.04$) than did low-probability biopsies. There were no significant differences in other histological features of AMR. Similarly, no differences were identified in i and t scores, in chronic histological features (ct, ci, and cv scores), or in the overall degree of interstitial fibrosis and tubular atrophy ($P > 0.9$). Most cases predicted to be at a low risk of AMR (32 of 44) had MVI scores of 0 or 1. Ten low-probability cases had isolated TMA or isolated v, 22 had MVI scores of 1 (1 had both TMA and g), and 12 had MVI scores of ≥2. In the group of 6 with a high AMR probability, there were 3 biopsies with MVI scores of 1, 1 had ptc alone, and the other 2 had concurrent glomerulonephritis or BLTCMR, so 1 component of the MVI score was not counted. The other 3 had MVI scores of ≥2; 1 of these also had TMA, and 1 also had v. There were no statistical differences in AMR score in patients

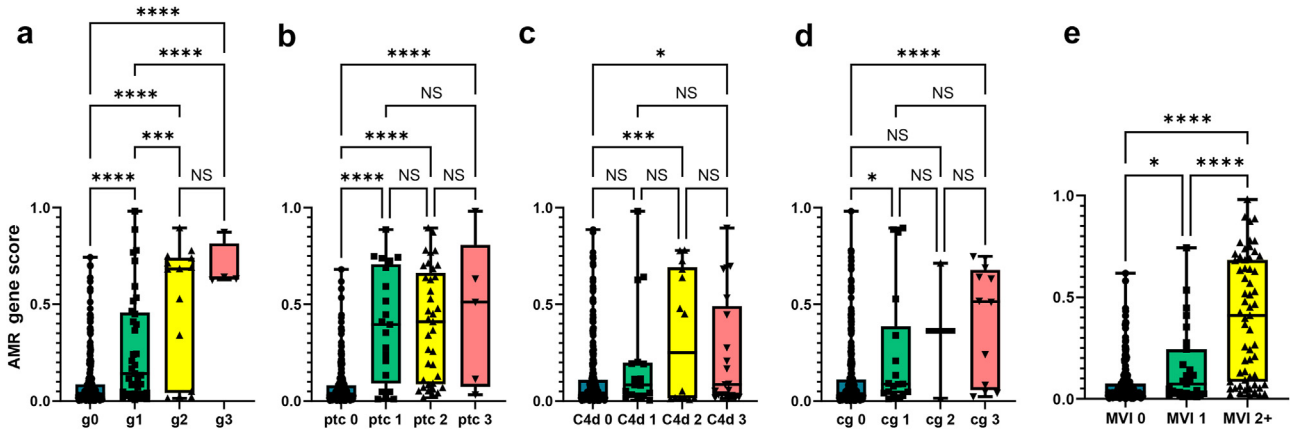


Figure 4 | Antibody-mediated rejection (AMR) prediction score in biopsies with histological features of AMR. (**a**) AMR score in biopsies with glomerulitis (g) graded from 0 to 3. (**b**) AMR score in biopsies with (ptc score >0) and without (ptc score 0) peritubular capillaritis (ptc). (**c**) AMR score in biopsies with C4d staining graded from 0 to 3. (**d**) AMR gene score with chronic glomerulopathy (cg) graded from 0 to 3. Biopsies with cg noted on electron microscopy alone (cg 1a) were included with cg 0 biopsies. There were only 2 biopsies with cg 2 in this cohort. (**e**) AMR score in biopsies with microvascular inflammation (MVI = g + ptc) scores of 1 ($P = 0.0248$) or ≥2 ($P < 0.0001$) compared with those without MVI (MVI score 0). AMR scores between groups were compared using the *t* test. NS, not significant. *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$.

with and without detectable HLA-DSA at the time of biopsy nor was MVI score alone responsible for an elevated AMR score (Figure 6a and b).

Those biopsies with a high AMR probability had significantly worse allograft outcomes (Figure 7). All patients with high-probability biopsies had allograft failure within 5 years of biopsy. The AMR molecular prediction score correlated strongly with the risk of graft loss ($\chi^2 = 5.7242$; $P = 0.0167$) overall and was independently associated with allograft loss in multivariable analysis (Supplementary Table S6).

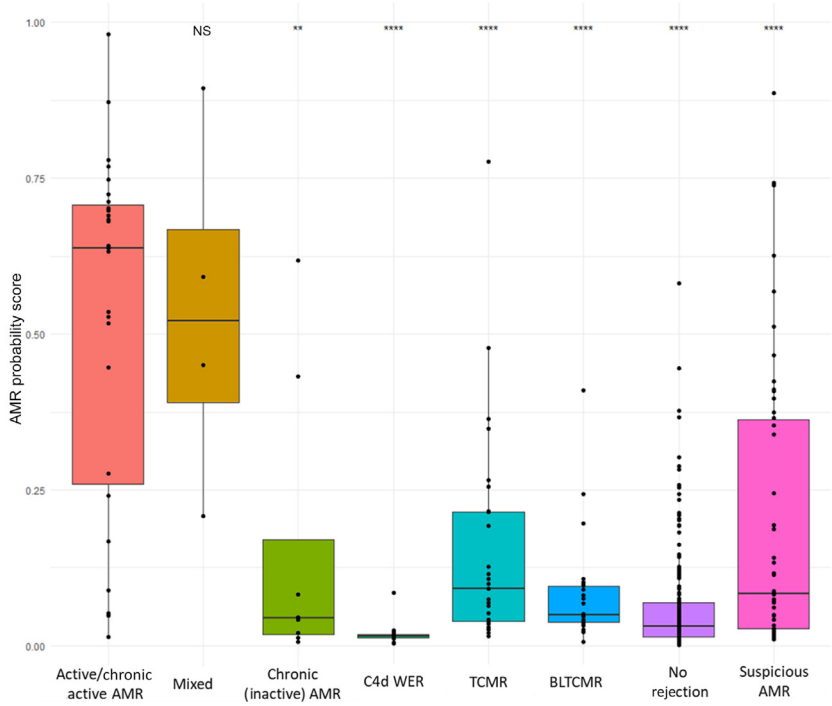


Figure 5 | Box plots comparing antibody-mediated rejection (AMR) probability scores across Banff diagnostic groups. The adjusted *P* values for the scores in each group compared to biopsies with active AMR are shown: AMR versus mixed rejection (not significant [NS], $P = 0.977$); AMR versus chronic (inactive) AMR (**$P = 0.0013$); AMR versus C4d deposition without evidence of rejection (C4d WER; ***$P < 0.0001$); T cell–mediated rejection (TCMR; ****$P < 0.0001$), borderline or T cell–mediated rejection (BLTCMR; ****$P < 0.0001$), and no rejection (****$P < 0.0001$). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers in the Tukey method.

Table 5 | Comparison of Susp-AMR biopsies divided according to the AMR molecular classifier

Characteristic	AMR not predicted (n = 44)	AMR predicted (n = 6)	P[a]
Allograft loss	17 (39)	6 (100)	0.006
Age at biopsy, yr	53 (44–61)	57.5 (53.75–62)	0.2
Sex (male)	18 (41)	3 (50)	0.7
Time from transplant to biopsy, d	703 (248.50–2095.75)	328.5 (157–807.5)	0.12
Transplant type			0.6
Deceased donor	25 (57)	5 (83)	
Live donor	16 (36)	1 (17)	
SPK	3 (6.8)	0 (0)	
Sensitization			
Total mismatch number (HLA-A, -B, -DR)	2.9	2	0.4
Previous transplant (yes)	3 (6.8)	1 (17)	0.4
Graft function at biopsy			
Creatinine, μmol/l	314.5 (75–818)	145.5 (76–677)	
eGFR, ml/min per 1.73 m^2	38.5 (25.5–49)	15.5 (12.75–22)	0.045
Proteinuria (uPCR)	47 (20–185)	101.50 (82.25–116.25)	0.5
Treatment			
Prior treatment of rejection (yes)	11 (25)	2 (33)	0.6
Calcineurin inhibitor	44 (100)	6 (100)	
Antiproliferative	16 (36)	1 (17)	0.6
Steroid	11 (25)	1 (17)	>0.9
Change in treatment after biopsy? (yes)	32 (73)	5 (83)	>0.9
Histology			
DSA (yes)	3 (6.8)	1 (16.7)	0.4
C4d score	0 (0–0)	0 (0–0.25)	0.3
g score	1 (0–1)	0.5 (0–2.25)	0.10
ptc score	0 (0–1.75)	1.5 (1–2.25)	0.002
MVI score	1 (1–2)	1.5 (1–3.5)	0.01
v score	0 (0–0)	0 (0–1.5)	0.3
i score	0 (0–1)	0 (0–1.5)	0.5
t score	0 (0–2)	0 (0–1)	0.5
ci score	1 (0.25–2)	1 (0.75–2.25)	0.6
ct score	1 (1–2)	1 (1–2.25)	0.3
cv score	1 (1–2)	1 (0–2.5)	0.2
cg score	0 (0–0.5)	0.5 (0–3)	0.04
ah score	0.5 (0–2)	1.5 (0–3)	0.2
ti score	1 (0–1.75)	1.5 (0.75–3)	0.3
IFTA (%)	20 (8.75–30)	20 (10–25)	>0.9
AMR histological features[b]			
MVI ≥ 2	12 (27)	3 (50)	0.346
MVI = 1	22 (50)	3 (50)	>0.9
MVI = 0	10 (23)	0 (0)	0.327
Thrombotic microangiopathy	9 (1 with g = 1)	1 (with ptc = 2)	>0.9
v lesion	2 (with mvi = 0)	1 (with g = 2, ptc = 3)	0.378

ah, arteriolar hyalinosis; AMR, antibody-mediated rejection; cg, glomerular capillary double contours; ci, interstitial fibrosis; ct, tubular atrophy; cv, vascular fibrous intimal thickening; DSA, donor-specific antibody; eGFR, estimated glomerular filtration rate; g, glomerulitis; HLA, human leukocyte antigen; i, interstitial inflammation; IFTA, interstitial fibrosis and tubular atrophy; MVI, microvascular inflammation; ptc, peritubular capillaritis; SPK, simultaneous pancreas-kidney transplant; Susp-AMR, suspicious for antibody-mediated rejection; t, tubulitis, ti, total inflammation; uPCR, urine protein-creatinine ratio; v, intimal arteritis.

[a]Fisher exact test and Wilcoxon rank-sum test.

[b]MVI is defined as the sum of g and ptc scores (in the absence of glomerulonephritis for g and of borderline or T cell–mediated rejection (BLTCMR or TCMR) or infection for peritubular capillaritis). Where ptc exists in the presence of BLTCMR or TCMR or infection or g exists in the presence of glomerulonephritis or BL/TCMR, these scores are not counted toward either histological evidence of AMR (criterion 1) or included in the sum g + ptc = MVI.

Data are expressed as median (interquartile range) or n (%).

AMR score is associated with graft loss when using the 2013 Banff definition of suspicious

Under the 2013 Banff classification, "Suspicious for AMR" existed as a separate category,[30] requiring the presence of 2 of 3 Banff criteria. We assessed whether the AMR gene scores in Susp-AMR biopsies fulfilling 2 of 3 Banff criteria had significantly different AMR scores than did those that fulfilled only criterion 1. There was no significant difference in score between biopsies meeting the historical (2013) definition of "suspicious" and those fulfilling criterion 1 alone (Figure 6c). There was no significant difference between HLA-DSA–positive and –negative biopsies in the Susp-AMR group. We subsequently reclassified all biopsies in the cohort according to the 2013 Banff classification. Thirty biopsies met the historical definition of "suspicious." In this subset, 7 of 30 had a high AMR probability, and the outcome in these biopsies was significantly ($P = 0.003$) worse than those in equivalent biopsies with a lower predicted probability of AMR (Supplementary Figure S7).

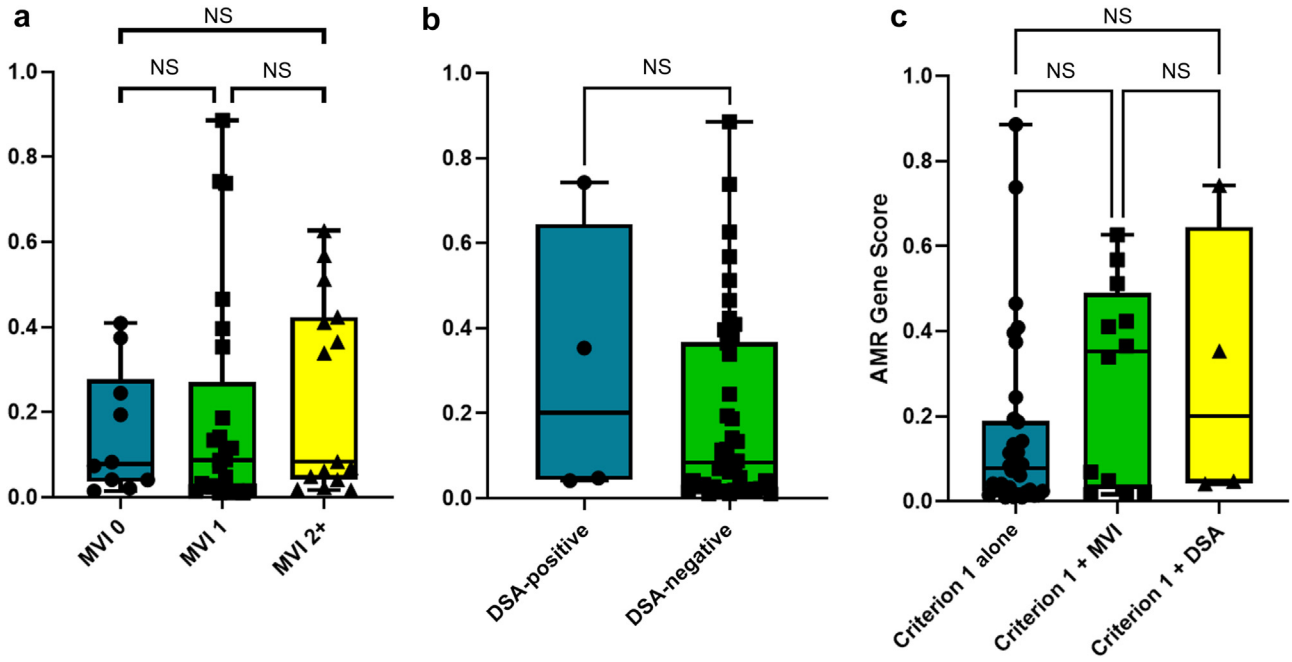


Figure 6 | In biopsies with incomplete histological features of antibody-mediated rejection (AMR) (suspicious for AMR [Susp-AMR]), gene score is not significantly different between biopsies with and without the addition of Banff criteria 2 and 3. (**a**) There was no significant difference in AMR score in Susp-AMR biopsies with varying degrees of microvascular inflammation. (**b**) There was no significant ($P = 0.1026$) difference in AMR gene score between anti–human leukocyte antigen–donor-specific antibody (HLA-DSA)-positive and -negative biopsies within the Susp-AMR group. (**c**) Gene score in Susp-AMR biopsies is not significantly different between biopsies that have histological features (criterion 1) of AMR alone or criterion 1 with either a microvascular inflammation (MVI score of ≥2 or positive DSA at the time of biopsy (the historical 2013 Banff definition of "suspicious for AMR"). NS, not significant.

AMR prediction correlates with allograft loss

To simulate the clinical use of histological features of AMR (H), the detection of circulating DSA or C4d (D), and the addition of the molecular classifier score (G) on allograft loss across all biopsies in the retrospective cohort, we generated logistic regression models including combinations of these variables (Figure 8), as described previously.[16,31] The full D + H + G model had a larger area under the receiver operating characteristic curve (0.628; range, 0.563–0.695) than the combination of DSA and histology (0.611; range, 0.544–0.668) or histology alone (0.586; range, 0.534–0.638). Net reclassification indices demonstrated improved reclassification of risk in the full model as compared with histology alone (net reclassification index 0.368 [range, 0.153–0.584]; $P = 0.0008$) and the combination of DSA and histology (net reclassification index 0.286; $P = 0.01$).

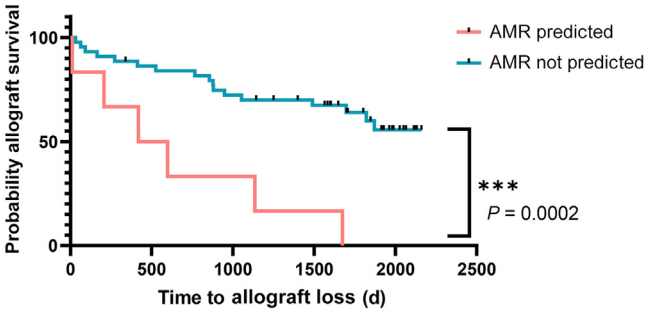


Figure 7 | Kaplan-Meier survival curve of time for biopsy to death-censored allograft failure in "suspicious for antibody-mediated rejection (AMR)" biopsies. Allograft loss was seen in 6 of 6 biopsies (100%) with a high probability of AMR (pink line) and in 17 of 44 biopsies (39%) with a low probability of AMR (blue line).

DISCUSSION

This is the first study to develop and apply an AMR-associated gene score using the B-HOT Panel on FFPE samples. Using an unselected retrospective cohort of transplant biopsies to minimize the sample bias, we developed a parsimonious 9-gene panel and demonstrated that it effectively identifies cases with worse outcomes in biopsies with features suspicious of AMR. This work therefore confirms the proposed application of molecular diagnosis as defined in the Banff classification. Our classifier is highly specific at ruling out AMR in biopsies with incomplete phenotypes. It performed less well at predicting AMR in our smaller validation cohort, which comprised only 10 cases of AMR, including 1 from an ABO-incompatible donor with TMA alone. However, the accuracy of our molecular classifier in the validation cohort (0.918) compares favorably with the accuracy seen in a microarray AMR classifier (0.85)[13] and a 3-gene AMR classifier identified in nonhuman primates (0.881).[32]

AMR has a heterogeneous histological appearance, including variable activity and chronicity relating to multiple

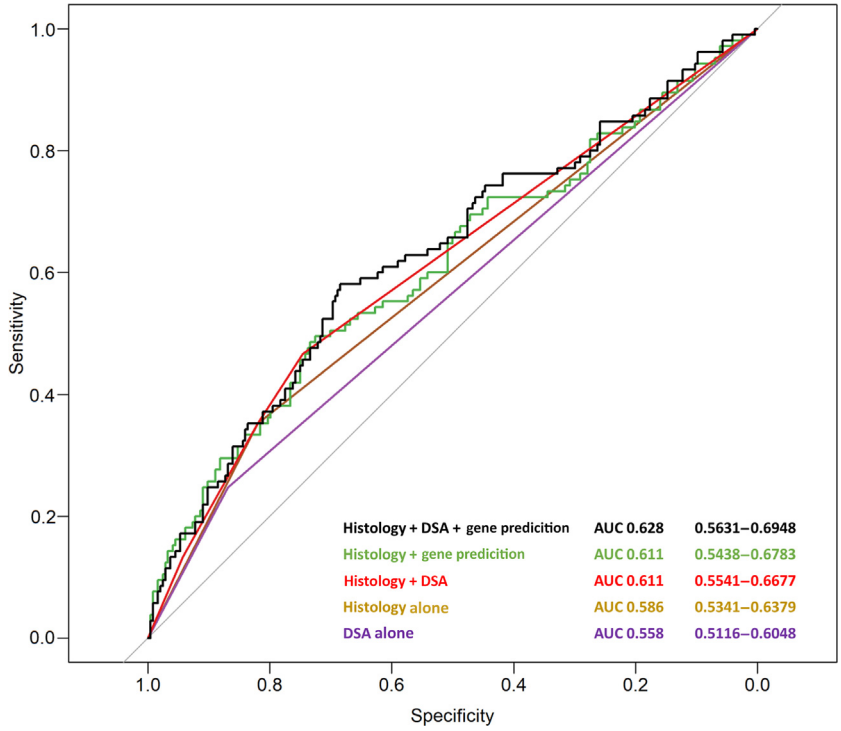


Figure 8 | Prediction of death-censored graft failure ($n = 349$). Receiver operating characteristic curve analysis of the prediction of death-censored allograft failure revealed an area under the curve (AUC) of 0.558 for the presence of donor-specific antibody (DSA) or C4d alone, 0.586 for the presence of histological features of antibody-mediated rejection (AMR) alone, 0.608 for the combination of histological and serological features of AMR, 0.611 for the addition of the AMR molecular classifier to histology, and 0.632 for the combination of serological, histological, and molecular features in predicting allograft loss.

biopsy features.[3,4,33] Within the most recent versions of the Banff classification, a subset of biopsies that were previously referred to as Susp-AMR, including some with MVI but without C4d or detectable DSA, were reclassified as No-AMR.[33] Biopsies with an incomplete AMR phenotype represent a diagnostic quandary. In series that include protocol biopsies, biopsies with an incomplete phenotype for AMR have inferior outcomes to patients without rejection but with a more indolent disease course than do histologically similar patients with detectable circulating DSA.[15,34] Other studies of indication biopsies have shown similarly poor outcomes in biopsies with MVI from patients with and without DSA.[35,36] Studies with the molecular microscope (MMDx, Kashi Clinical Laboratories) showed no difference in gene expression profiles of biopsies with MVI, regardless of the presence of DSA.[36] Our research complements these studies by showing that in a group of biopsies with incomplete features of AMR, some have a higher molecular probability of AMR, whether they had only histological features (criterion 1) alone or met the historical 2013 Banff definition of "suspicious" (with 2 of 3 Banff criteria for AMR). Our classifier correlated strongly with histological features of AMR (ptc, g, v, cg, and C4d), which reflects findings by other groups using microarray[8,9,13] and NanoString in transplant biopsies,[18,37] and we see significant differences in gene score in biopsies with varying levels of MVI. In combination with the findings of other groups, our results confirm that AMR gene scores correlate more with MVI scores than with DSA.

However, MVI is not the sole determinant of elevation of our AMR gene score. It is impossible to tell at this stage why some biopsies with MVI show increased expression of our AMR gene score and inferior outcome and others not; there may be a variety of causal and treatment-related factors that influence this.

Our results show that the group we defined as Susp-AMR contains a majority of cases that do not align with HLA-DSA–positive AMR (44 of 50) and a minority that do. Only 8% of the Susp-AMR cases had HLA-DSA; the etiology of the other cases may be diverse: related to non–HLA-DSA, to missing self,[38,39] or other causes of TMA or v lesions. One of the 6 reclassified was HLA-DSA–positive and likely represents "subdiagnostic threshold" AMR. Five of the 6 reclassified were HLA-DSA–negative and may be related to anti–HLA-DSA undetectable with our current methods. Our Susp-AMR group included both cases with MVI and other histological features of AMR, namely, TMA and v. The majority (35 of 50) of Susp-AMR cases had MVI scores of 0 or 1, and these were more likely to be classified as low-probability AMR. In the Susp-AMR group, there was no difference in scores between DSA-

positive and DSA-negative biopsies, regardless of the histological lesions.

Although this is the first study using NanoString in the context of use defined by the current Banff classification, we previously used a 10-gene score using quantitative polymerase chain reaction to classify biopsies with features suspicious of AMR into high- and low-risk groups for allograft loss.[16] The genes in this former 10-gene score were selected from the literature as the most reliably increased transcripts described by others in AMR. When "top transcripts" are selected, these may belong to the same biological pathways and be strongly coexpressed, offering no additive predictive value to a model. In the present study, we used LASSO regression to generate a sparse predictive model, a more robust approach for generating a predictive score. LASSO selects 1 feature from multiple correlated features to provide a list of genes that convey orthogonal information. This explains why the gene sets derived in our 2 studies only partially overlap (for *PLA1A*, *CXCL11*, and *PECAM1*). NanoString analysis on FFPE samples has the added advantage of enabling direct comparisons between histological features and transcript expression in the same tissue.[18,37]

Although developed for classification of biopsies, our score correlates strongly with the risk of allograft loss in biopsies with incomplete AMR. The addition of our molecular classifier to the standard histological and serological diagnosis of AMR improved the prediction of allograft loss in all biopsies in our cohort.

This is the largest cohort of transplant biopsies in which B-HOT NanoString analysis has been reported to date. Nevertheless, AMR represents a relatively rare diagnosis (<10% of biopsies in our cohort). Relatively small numbers affect the power to detect significant differences in genes between groups, while class imbalance may mean models are more accurate at predicting the dominant class—biopsies without AMR. We compared LASSO regression with Random Forest and found that within our training sets, LASSO outperformed Random Forest. We did not prefilter our gene expression data before LASSO regression, as we did not want to exclude genes that, even though not significantly differentially expressed, could retain strong discriminatory potential. The genes identified by LASSO have been described in other studies of transplant biopsies. *PLA1A* and *CXCL11*, interferon-inducible endothelial transcripts, have been well described as being specific for AMR.[12,13,16–18,40] *PTPN6* is a regulator of *STAT4* and has been associated with allograft loss after transplant rejection.[41] *PECAM1*, a cellular adhesion molecule, is upregulated in AMR.[11] *EMP3* and *HLA-DQB1* are both markers in the cell-extracellular matrix pathways that are upregulated in chronic AMR.[40] *RPS6*, a mammalian target of rapamycin pathway–associated protein kinase gene, was the only gene among our predictors that was not significantly differentially expressed between AMR and No-AMR. Studies of chronic AMR in primates have shown only a variable association between *RPS6* and chronic AMR stages.[42,43] The role of *RPS6* in the mammalian target of rapamycin pathway has been extensively described in heart and lung allografts with AMR[44,45] and kidney transplants where it was shown to be strongly associated with the presence of anti–HLA-DSA and poor allograft survival.[46] Our model suggests that *RPS6* may help discriminate some biopsies with AMR within this cohort, providing further evidence that the mammalian target of rapamycin pathway may be important in a subset of AMR.[44–46]

Although the unselected cohort represents a "real-world" approach to biopsies, there are limitations. This is a single-center study, and although it is the largest published cohort of NanoString analysis of transplant biopsies to date, there may be center-specific differences in patient demographics or immunosuppression that limit its generalizability to other transplant centers. Patients in our cohort received a wide range of treatments both before and after their index biopsies, which may have influenced both their gene expression profiles at the time of biopsy and subsequent allograft outcomes. We selected allograft failure as a clinically relevant "hard" outcome measure, but this can be influenced by factors other than AMR; we acknowledge that an improvement in prognostic capability is not the same as diagnostic accuracy and prospective evaluation of the utility of a molecular score compared with the current standard of care alone will be necessary to assess diagnostic accuracy. To investigate the reproducibility of our model and generalizability to patients in other centers, further external validation will be needed.

In summary, we have demonstrated that a 9-gene AMR classifier applied to biopsies with incomplete AMR phenotypes enables the identification of those with worse outcomes. To validate an AMR gene score for clinical practice, the next steps are as follows: external validation with other groups using NanoString (work with the International Consortium for Diagnostics & Outcomes in Transplantation; http://icdot.org/) or other platforms and prospective investigation in clinical trials of the added value in patient care.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

JB was supported by the Auchi Renal Research Fund through the Imperial Health Charity. CR was supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. CR's research activity is made possible with generous support from Sidharth and Indira Burman.

Human samples used in this research project were obtained from the Imperial College Healthcare Tissue and Biobank (ICHTB). ICHTB is supported by the NIHR Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB is approved by Wales REC 3 to release human material for research (22/WA/2836).

SUPPLEMENTARY FILE (PDF)

Supplementary File (PDF)

Supplementary Table S1. NanoString Banff Human Organ Transplant (B-HOT) Panel genes.

Supplementary Table S2. Differences between the discovery and validation cohorts.

Supplementary Table S3. Confusion matrices of Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest models in the testing, training, and validation sets.

Supplementary Table S4. Pearson correlation between Least Absolute Shrinkage and Selection Operator (LASSO)–generated predictors and histological lesions on biopsy.

Supplementary Table S5. Multivariable analysis of survival in suspicious for antibody-mediated rejection (Susp-AMR) biopsies by histological and clinical parameters at the time of biopsy.

Supplementary Table S6. Banff Human Organ Transplant Panel gene annotations and KEGG pathway labels.

Supplementary Figure S1. Relative log expression (RLE) plots of normalized data. RLE plots illustrating any unwanted variation or batch effects. (**A**) Colored by cohort: discovery (D) and validation (V). (**B**) Colored by diagnostic group: antibody-mediated rejection (AMR; pink), No AMR (blue), or suspicious for AMR (Susp-AMR; yellow). (**C**) Colored by Banff Human Organ Transplant (Banff-HOT) Panel codeset batch (1–4).

Supplementary Figure S2. Model for the development of the antibody-mediated rejection (AMR) gene score. Biopsies with complete features of AMR ("AMR") and without histological features of AMR ("No-AMR") were used to develop a molecular model predictive of AMR. Step 1: A stratified split (73.5:26.5) of the discovery cohort was performed to partition the data into training and testing sets. Two models were fit on the training set: Least Absolute Shrinkage and Selection Operator (LASSO) regression and Random Forest (RF). Two hundred models in the training set were performed using leave-one-group-out cross-validation to tune the optimal lambda for LASSO. RF was performed on 200 models using leave-one-group-out cross-validation with "mtry" kept constant [mtry = sqrt(number of predictors − 1)]. The LASSO and RF models were then tested in the testing set, and a confusion matrix was generated for both. These parameters were then used on the entire discovery cohort, and the coefficients of the model were extracted to validate in the validation cohort. Step 2: LASSO and RF, tested and tuned in the discovery cohort, were validated in "AMR" and "No-AMR" biopsies in the validation cohort. Step 3: The AMR prediction model was tested on suspicious for AMR ("Susp-AMR") biopsies, and AMR prediction scores were extracted for all biopsies with incomplete AMR phenotypes. AUC, area under the curve.

Supplementary Figure S3. Volcano plots showing the differential gene expression between antibody-mediated rejection (AMR) and No AMR biopsies, adjusted for 5% false discovery rate (FDR). (**A**) Discovery cohort (AMR, $n = 21$; No AMR, $n = 167$): 391 genes were significantly differentially expressed. (**B**) Validation cohort (AMR, $n = 10$; No AMR, $n = 101$): 28 genes were significantly differentially expressed (5% FDR).

Supplementary Figure S4. Correlation in differential gene expression profiles in the discovery and validation cohorts. (**A**) Scatter plot of differential gene expression estimates in the discovery and validation cohorts, colored by 5% false discovery rate (FDR); 17 genes were significantly differentially expressed by FDR in both cohorts. (**B**) Scatter plot of differential gene expression estimates in the discovery and validation cohorts, colored by *P* value. The differential expression profiles of genes in antibody-mediated rejection (AMR) versus No-AMR biopsies in both the discovery and validation cohorts correlate significantly (rho = 0.6215; $P < 0.0001$). Genes in pink were significantly differentially expressed in AMR versus No AMR biopsies in both the discovery and validation cohorts; genes in green were significantly different in the discovery cohort only; genes in purple were significantly different in the validation cohort only; and genes in blue were not significantly different in AMR versus No-AMR biopsies in either cohort.

Supplementary Figure S5. Gene set variation analysis (GSVA) method has been used to explore differential pathway expression in antibody-mediated rejection (AMR) and No AMR biopsies in the combined discovery and validation cohorts. (**A**) Volcano plot illustrating the differential expression of pathways in AMR versus No AMR. Upregulated pathways in AMR, such as interferon gamma response and allograft rejection, are red and seen on the right. (**B**) Pathways differentially expressed in AMR versus No AMR biopsies, ranked by $-\log_{10}$(*P*-value significance scores).

Supplementary Figure S6. Expression of the selected predictors in antibody-mediated rejection (AMR) versus No AMR biopsies in the combined discovery and validation cohorts. *PLA1A*, *PTPN6*, *MAPK3*, *CXCL11*, *HLA-DQB1*, *IFNGR1*, *PECAM1*, and *EMP3* were all significantly differentially expressed. The expression of *RPS6* between AMR and No AMR biopsies did not reach statistical significance.

Supplementary Figure S7. Kaplan-Meier survival curve of time for biopsy to death-censored allograft failure in biopsies reclassified as "Suspicious for antibody-mediated rejection (AMR)" under the 2013 Banff classification. Allograft loss was seen in 6 of 7 (85%) biopsies with a high probability of AMR (pink line) and in 6 of 23 (26%) biopsies with a low probability of AMR (blue line).

REFERENCES

1. Sellarés J, de Freitas DG, Mengel M, et al. Understanding the causes of kidney transplant failure: the dominant role of antibody-mediated rejection and nonadherence. *Am J Transplant*. 2012;12:388–399.
2. Solez K, Axelsen RA, Benediktsson H, et al. International standardization of criteria for the histologic diagnosis of renal allograft rejection: the Banff working classification of kidney transplant pathology. *Kidney Int*. 1993;44:411–422.
3. Roufosse C, Simmonds N, Clahsen-van Groningen M, et al. 2018 Reference Guide to the Banff Classification of Renal Allograft Pathology. *Transplantation*. 2018;102:1795–1814.
4. Loupy A, Haas M, Roufosse C, et al. The Banff 2019 Kidney Meeting Report (I): updates on and clarification of criteria for T cell- and antibody-mediated rejection. *Am J Transplant*. 2020;20:2318–2331.
5. Mengel M, Sis B, Halloran PF. SWOT analysis of Banff: strengths, weaknesses, opportunities and threats of the international Banff consensus process and classification system for renal allograft pathology. *Am J Transplant*. 2007;7:2221–2226.
6. Furness PN, Taub N, Assmann KJ, et al. International variation in histologic grading is large, and persistent feedback does not improve reproducibility. *Am J Surg Pathol*. 2003;27:805–810.
7. Halloran PF, de Freitas DG, Einecke G, et al. The molecular phenotype of kidney transplants. *Am J Transplant*. 2010;10:2215–2222.
8. Sarwal M, Chua MS, Kambham N, et al. Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. *N Engl J Med*. 2003;349:125–138.
9. 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;7:2712–2722.
10. Hidalgo LG, Sis B, Sellares J, et al. NK cell transcripts and NK cells in kidney biopsies from patients with donor-specific antibodies: evidence for NK cell involvement in antibody-mediated rejection. *Am J Transplant*. 2010;10:1812–1822.
11. Sis B, Jhangri GS, Bunnag S, et al. Endothelial gene expression in kidney transplants with alloantibody indicates antibody-mediated damage despite lack of C4d staining. *Am J Transplant*. 2009;9:2312–2323.
12. Venner JM, Hidalgo LG, Famulski KS, et al. The molecular landscape of antibody-mediated kidney transplant rejection: evidence for NK

involvement through CD16a Fc receptors. *Am J Transplant*. 2015;15:1336–1348.
13. Sellarés J, Reeve J, Loupy A, et al. Molecular diagnosis of antibody-mediated rejection in human kidney transplants. *Am J Transplant*. 2013;13:971–983.
14. Loupy A, Lefaucheur C, Vernerey D, et al. Molecular microscope strategy to improve risk stratification in early antibody-mediated kidney allograft rejection. *J Am Soc Nephrol*. 2014;25:2267–2277.
15. Callemeyn J, Lerut E, de Loor H, et al. Transcriptional changes in kidney allografts with histology of antibody-mediated rejection without anti-HLA donor-specific antibodies. *J Am Soc Nephrol*. 2020;31:2168–2183.
16. Toulza F, Dominy K, Willicombe M, et al. Diagnostic application of transcripts associated with antibody-mediated rejection in kidney transplant biopsies. *Nephrol Dial Transplant*. 2022;37:1576–1584.
17. Dominy KM, Roufosse C, de Kort H, et al. Use of quantitative real time polymerase chain reaction to assess gene transcripts associated with antibody-mediated rejection of kidney transplants. *Transplantation*. 2015;99:1981–1988.
18. Adam B, Afzali B, Dominy KM, et al. Multiplexed color-coded probe-based gene expression assessment for clinical molecular diagnostics in formalin-fixed paraffin-embedded human renal allograft tissue. *Clin Transplant*. 2016;30:295–305.
19. Dominy KM, Willicombe M, Al Johani T, et al. Molecular assessment of C4d-positive renal transplant biopsies without evidence of rejection. *Kidney Int Rep*. 2019;4:148–158.
20. Sigdel TK, Nguyen M, Dobi D, et al. Targeted transcriptional profiling of kidney transplant biopsies. *Kidney Int Rep*. 2018;3:722–731.
21. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26:317–325.
22. Mengel M, Loupy A, Haas M, et al. Banff 2019 Meeting Report: molecular diagnostics in solid organ transplantation—Consensus for the Banff Human Organ Transplant (B-HOT) gene panel and open source multicenter validation. *Am J Transplant*. 2020;20:2305–2317.
23. NanoStringQCPro: Quality metrics and data processing methods for NanoString mRNA gene expression data. R package version 1.28.0. 2023. https://doi.org/10.18129/B9.bioc.NanoStringQCPro
24. Gandolfo L, Speed T. RLE plots: visualizing unwanted variation in high dimensional data. *PLoS One*. 2018;13:e0191629.
25. R. A language and environment for statistical computing. R Foundation for Statistical Computing; 2021. https://www.R-project.org/
26. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
28. Liaw AaWM. Classification and regression by randomForest. *R News*. 2002;2:11–18.
29. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28:1–26.
30. Haas M, Sis B, Racusen LC, et al. Banff 2013 Meeting Report: inclusion of c4d-negative antibody-mediated rejection and antibody-associated arterial lesions. *Am J Transplant*. 2014;14:272–283.
31. Roufosse C, Drachenberg C, Renaudin K, et al. Molecular assessment of antibody-mediated rejection in human pancreas allograft biopsies. *Clin Transplant*. 2020;34:e14065.
32. Adam BA, Smith RN, Rosales IA, et al. Chronic antibody-mediated rejection in nonhuman primate renal allografts: validation of human histological and molecular phenotypes. *Am J Transplant*. 2017;17:2841–2850.
33. Callemeyn J, Ameye H, Lerut E, et al. Revisiting the changes in the Banff classification for antibody-mediated rejection after kidney transplantation. *Am J Transplant*. 2021;21:2413–2423.
34. Senev A, Coemans M, Lerut E, et al. Histological picture of antibody-mediated rejection without donor-specific anti-HLA antibodies: clinical presentation and implications for outcome. *Am J Transplant*. 2019;19:763–780.
35. Parajuli S, Redfield RR, Garg N, et al. Clinical significance of microvascular inflammation in the absence of anti-HLA DSA in kidney transplantation. *Transplantation*. 2019;103:1468–1476.
36. Halloran PF, Madill-Thomsen KS, Pon S, et al. Molecular diagnosis of ABMR with or without donor-specific antibody in kidney transplant biopsies: differences in timing and intensity but similar mechanisms and outcomes. *Am J Transplant*. 2022;22:1976–1991.
37. Rosales I, Mahowald G, Tomaszewski K, et al. Banff human organ transplant transcripts correlate with renal allograft pathology and outcome: importance of capillaritis and subpathologic rejection. *J Am Soc Nephrol*. 2022;33:2306–2319.
38. Callemeyn J, Senev A, Coemans M, et al. Missing self-induced microvascular rejection of kidney allografts: a population-based study. *J Am Soc Nephrol*. 2021;32:2070–2082.
39. Koenig A, Chen CC, Marçais A, et al. Missing self triggers NK cell-mediated chronic vascular rejection of solid organ transplants. *Nat Commun*. 2019;10:5350.
40. Gambella A, Barreca A, Osella-Abate S, et al. Caveolin-1 in kidney chronic antibody-mediated rejection: an integrated immunohistochemical and transcriptomic analysis based on the Banff Human Organ Transplant (B-HOT) gene panel. *Biomedicines*. 2021;9:1318.
41. Chen Y, Zhang B, Liu T, et al. T cells with activated *STAT4* drive the high-risk rejection state to renal allograft failure after kidney transplantation. *Front Immunol*. 2022;13:895762.
42. Smith RN, Matsunami M, Adam BA, et al. RNA expression profiling of nonhuman primate renal allograft rejection identifies tolerance. *Am J Transplant*. 2018;18:1328–1339.
43. Legaz I, Bernardo MV, Alfaro R, et al. PCR array technology in biopsy samples identifies up-regulated mTOR pathway genes as potential rejection biomarkers after kidney transplantation. *Front Med (Lausanne)*. 2021;8:547849.
44. Li F, Wei J, Valenzuela NM, et al. Phosphorylated S6 kinase and S6 ribosomal protein are diagnostic markers of antibody-mediated rejection in heart allografts. *J Heart Lung Transplant*. 2015;34:580–587.
45. Lepin EJ, Zhang Q, Zhang X, et al. Phosphorylated S6 ribosomal protein: a novel biomarker of antibody-mediated rejection in heart allografts. *Am J Transplant*. 2006;6:1560–1571.
46. Raïch-Regué D, Gimeno J, Llinàs-Mallol L, et al. Phosphorylation of S6RP in peritubular capillaries of kidney grafts and circulating HLA donor-specific antibodies. *Front Med (Lausanne)*. 2022;9:988080.