

**An integrated molecular diagnostic report for heart transplant biopsies
using an ensemble of diagnostic algorithms**

Michael D Parkes¹, Arezu Z Aliabadi², Martin Cadeiras³, Marisa G Crespo-Leiro⁴, Mario Deng³, Eugene C Depasquale³, Johannes Goekler², Daniel H Kim⁵, Jon Kobashigawa⁶, Alexandre Loupy⁷, Peter Macdonald⁸, Luciano Potena⁹, Andreas Zuckermann², and Philip F Halloran^{1,5}

¹Alberta Transplant Applied Genomics Centre, Edmonton, AB, Canada; ²Medical University of Vienna, Vienna, Austria; ³Ronald Reagan UCLA Medical Center, Los Angeles, CA, USA; ⁴Complejo Hospitalario Universitario A Coruña, A Coruña, Spain; ⁵Department of Medicine, University of Alberta, Edmonton, AB, Canada; ⁶Cedars-Sinai Medical Center, Beverly Hills, CA, USA; ⁷Hôpital Necker, Paris, France; ⁸The Victor Chang Cardiac Research Institute, Sydney, Australia; ⁹Cardiovascular Department, University of Bologna, Bologna, Italy.

Corresponding Author & Reprint Requests

Philip F. Halloran MD PhD
Alberta Transplant Applied Genomics Centre
#250 Heritage Medical Research Centre, University of Alberta
Edmonton, AB T6G 2S2, Canada
Phone: 780-492-6160; Fax: 780-407-7450
Email: phallora@ualberta.ca

Keywords: antibody-mediated rejection, T cell-mediated rejection, injury, heart transplant

Short title: Molecular diagnosis of heart injury & rejection

Abstract

Background. We previously reported a microarray-based diagnostic system for heart transplant endomyocardial biopsies (EMBs), using either 3-archetype (3AA) or 4-archetype (4AA) unsupervised algorithms to estimate rejection. The present study aimed to examine the stability of machine-learning algorithms in new biopsies, compare 3AA vs. 4AA algorithms, assess supervised binary classifiers trained on histologic or molecular diagnoses, create a report combining many scores into an ensemble of estimates, and examine possible automated sign-outs.

Methods. We studied 889 EMBs from 454 transplant recipients at eight centers: the initial cohort (N=331) and a new cohort (N=558). Published 3AA algorithms derived in cohort 331 were tested in cohort 558; the 3AA and 4AA models were compared; and supervised binary classifiers were created.

Results. Algorithms derived in cohort 331 performed similarly in new biopsies despite differences in case mix. In the combined cohort, the 4AA model, including a parenchymal injury score, retained correlations with histologic rejection and DSA similar to the 3AA model. Supervised molecular classifiers predicted molecular rejection (AUCs>0.87) better than histologic rejection (AUCs<0.78), even when trained on histology diagnoses. A report incorporating many AA and binary classifier scores interpreted by one expert showed highly significant agreement with histology ($p<0.001$), but with many discrepancies as expected from the known noise in histology. An automated random forest score closely predicted expert diagnoses, confirming potential for automated sign-outs.

Conclusions. Molecular algorithms are stable in new populations and can be assembled into an ensemble that combines many supervised and unsupervised estimates of the molecular disease states. (ClinicalTrials.gov NCT02670408).

Introduction

Molecular examination of endomyocardial biopsies (EMBs) presents an opportunity to reclassify disease states and improve the precision and accuracy of diagnoses. The current standard for diagnosing heart transplant rejection is histology, interpreted with limited reproducibility¹ using the International Society for Heart and Lung Transplantation (ISHLT)²⁻⁶ guidelines for T cell-mediated rejection (TCMR) and antibody-mediated rejection (ABMR). Molecular diagnosis offers the possibility of improved quantitation, objectivity, and mechanistic insights. We first reported assessments using gene sets⁷ and later assessed EMBs using kidney-derived rejection-associated transcript (RAT) expression. The genes associated with ABMR in hearts were highly similar to those in kidney⁸. We reported a first-generation centralized microarray-based Molecular Microscope[®] Diagnostic System (MMDx) for kidney⁹ and heart transplants^{7,10-12}. For EMBs, rejection was assessed using an unsupervised archetypal analysis (AA) that initially assigned three archetype (3AA) scores: $S1_{\text{normal}}$, $S2_{\text{TCMR}}$, and $S3_{\text{ABMR}}$ ¹². We explored a modification of this approach by a four-archetype (4AA) model that added a parenchymal injury score ($S4_{\text{injury}}$)¹³, taking advantage of the fact that molecular features of adaptive immunity (rejection) are shared by innate immunity (the inflammatory response to injury)¹⁴⁻¹⁶.

In kidney transplant biopsies, we showed the value of supervised classifiers trained on histology diagnoses^{17,18} as well as unsupervised archetypal scores⁹. This raised the possibility of incorporating the multiple supervised and unsupervised machine learning estimates into one comprehensive report based on the principle that multiple estimates improve accuracy¹⁹⁻²¹. With this in mind, the present study had four aims: to test whether diagnostic algorithms derived in one biopsy set performed similarly in new biopsies despite variations in case mix; to study whether incorporation of injury estimates affected the estimates of rejection; to develop new supervised classifier equations trained on rejection diagnoses; and to integrate many estimates into an ensemble that reported the molecular and injury phenotype, potentially using an automated sign-out. Our strategy is outlined in Supplementary Figure 1.

Methods

Population. This prospective study was approved by the ethics review board of each center (Supplementary Table 1) and is registered at ClinicalTrials.gov (NCT02670408). As previously described¹², biopsies were collected prospectively for clinical indications or by protocol, processed for histology and HLA antibody testing as per local standard of care. Histology interpretation followed ISHLT guidelines^{3;22}.

Microarray analysis. As detailed elsewhere¹², purified total RNA from EMBs including RNA available from 97 Edmonton samples used in a previous study⁷ was labeled with the 3' IVT Plus kit (Affymetrix, Santa Clara) and hybridized to PrimeView microarrays (Affymetrix) according to manufacturer protocols (www.affymetrix.com).

Rejection-Associated Transcripts (RATs). As previously described, RATs were derived as union (after eliminating overlaps) of the top 200 probe sets associated with each of three kidney transplant biopsy histology comparisons: all rejection vs. everything else, ABMR vs. everything else, and TCMR vs. everything else^{9;23}.

Pathogenesis-based transcript sets (PBTs). PBTs were previously annotated to reflect biological mechanisms in rejection and injury, defined in human cell lines, mouse experimental models, and human transplant biopsies (detailed at <https://www.ualberta.ca/medicine/institutes-centres-groups/atagc/research/gene-lists>) PBTs are expressed as their mean fold difference in biopsies compared to a control group. We focused on PBTs for injury-and-repair transcripts (IRRATs)^{24;25}, quantitative constitutive macrophage transcripts (QCMATs)²⁶, quantitative cytotoxic T cell transcripts (QCATs)²⁷, T cell burden transcripts (TCBs)²⁸, NK cell burden transcripts (NKBs)²⁸, IFNG-inducible transcripts (GRITs)²⁹, DSA-selective transcripts (DSASTs)³⁰, and the endothelium-expressed subset of DSASTs (eDSASTs).

Archetypal analysis (AA). The methods for AA are published³¹, including model 1 for three archetypes (3AA)¹² and model 2 for four archetypes (4AA)¹³. Both were based on RAT expression and assigned each EMB a set of scores representing no rejection ($S1_{normal}$), TCMR ($S2_{TCMR}$), and ABMR ($S3_{ABMR}$). The 4AA model assigned an additional score for injury ($S4_{injury}$).

Rejection diagnoses. To facilitate comparison between MMDx and histology, histologic and molecular scores were translated into diagnoses of ABMR, possible ABMR (pABMR), TCMR, possible TCMR (pTCMR), mixed ABMR plus TCMR, and no rejection¹³. Histologic ABMR 1H+ and 1I+ were pABMR; grades 2 and 3 were ABMR. Histologic TCMR 1R was pTCMR; grades 2R and 3R were TCMR. Using 4AA scores, $0.3 \leq S3_{ABMR} < 0.5$ constituted pABMR, $S3_{ABMR} \geq 0.5$ constituted ABMR, $0.2 \leq S2_{TCMR} < 0.3$ constituted pTCMR, and $S2_{TCMR} \geq 0.3$ constituted TCMR. $S4_{injury} \geq 0.2$ constituted possible injury. Cut-offs were derived arbitrarily from examination of the data.

Binary rejection classifiers. The binary classifiers comparing a positive class to a negative class were derived to predict the probability of histologic or molecular diagnoses of ABMR, TCMR, or any rejection (using the above nomenclature). For features they used the top 20 transcripts (selected using Bayesian t-tests of all possible transcripts) associated with the relevant diagnosis, either histologic or molecular, compared to all other biopsies. They were constructed using linear discriminant analysis using 10-fold cross-validation. The top transcripts and the classifiers were derived from scratch in each fold of cross-validation, yielding test set predictions for all biopsies and preventing overfitting.

Results

Biopsy population. Clinically indicated, protocol and follow-up EMBs from 454 patients were collected in two cohorts: cohort 331 (N=331) from the earlier study¹² and a new cohort of 558 biopsies (Table 1 and Supplementary Table 2). Cardiomyopathy and coronary artery disease were the common primary diseases.

Histology grades and molecular scores were translated to a common nomenclature¹³ (see Methods): TCMR, pTCMR, ABMR, and pABMR. Frequencies of the histologic classes in cohort 331 and 558 are in Table 2 and Supplementary Table 3. As expected, in cohort 558 ABMR was less frequent than in cohort 331, (3% vs. 10%) particularly in the late biopsies, and no rejection was more frequent (44% vs. 27%) than in cohort 331. Case mix differed between cohorts because cohort 331 was enriched in high risk heart transplants performed in Paris, whereas cohort 558 was intended to be more representative of the prevalent international heart transplant population.

At their latest biopsy, 35% percent of patients in cohort 331 and 37% of patients in cohort 558 were DSA positive (Table 2).

Performance of machine learning algorithms in a new cohort. We studied whether the previously published 3AA model for classification of rejection in cohort 331¹² would detect rejection in a new cohort 558. Relationships between $S1_{\text{normal}}$, $S2_{\text{TCMR}}$, and $S3_{\text{ABMR}}$ scores and the corresponding histologic classes were similar in both cohorts: $S1_{\text{normal}}$ scores were highest in no rejection, $S2_{\text{TCMR}}$ scores were highest in TCMR increasing through no rejection and pTCMR, and $S3_{\text{ABMR}}$ scores were highest in ABMR increasing through no rejection and pABMR (Figure 1, Supplementary Table 4). Thus, unsupervised machine learning algorithms derived in one population performed as expected in new biopsies despite differences in case mix and participating centers.

Does incorporating injury estimates affect the assessment of rejection? We compared relationships between histologic diagnoses and molecular scores in the 3AA and 4AA models (Figure 2, Supplementary Table 5). Because the 4AA model was developed in the combined population¹³, we re-derived the 3AA model in the 889 population to permit comparison of the 3AA and 4AA scores.

Relationships between histologic diagnoses and $S1_{\text{normal}}$, $S2_{\text{TCMR}}$, and $S3_{\text{ABMR}}$ scores (Figure 2 A-D) were similar in both models: highest $S2_{\text{TCMR}}$ scores in TCMR and mixed, highest $S3_{\text{ABMR}}$ scores in ABMR, and a gradient in scores from normal to definite TCMR and ABMR. Adding the $S4_{\text{injury}}$ score separated no rejection biopsies into injured and normal. The $S4_{\text{injury}}$ scores were elevated in histologic TCMR and mixed compared to no rejection (Figure 2 E-F), more so than in ABMR. Moving from a 3AA to a 4AA model preserved the relationship of the molecular scores to histologic rejection.

Histologic classes were studied in RAT-based principal component analysis (PCA) (Supplementary Table 6). TCMR and ABMR had higher PC1 scores than no rejection. ABMR had high PC2 scores and TCMR had low PC2 scores. There were no significant differences among rejection classes in median PC3 scores, as expected since PC3 reflects abnormalities unrelated to rejection, collectively termed injury¹³. However, among rejection classes the highest PC3 was in TCMR/mixed biopsies, reflecting parenchymal injury due to TCMR. This presumably reflects the fact that TCMR is an interstitial inflammatory process adjacent to the parenchyma whereas ABMR is primarily an intercapillary process and usually has little effect on the parenchyma until late.

The fold differences in 3AA and 4AA scores between DSA positive and DSA negative patients were calculated. $S3_{\text{ABMR}}$ scores were higher and $S1_{\text{normal}}$ scores were lower in biopsies from DSA positive patients (Supplementary Table 7).

Developing binary molecular rejection classifiers. Supervised binary molecular classifiers comparing a positive histology class to a negative class have proven useful in kidney transplant biopsies^{17;18}. To see if this approach would be useful for heart biopsies, we developed binary molecular classifiers trained using either histologic or molecular EMB diagnoses (see Methods).

Table 3 shows the areas under the receiver-operating characteristic curves (AUCs) for predicting histologic or molecular diagnoses of rejection with the new binary molecular classifiers, compared to the 4AA $S1_{\text{normal}}$, $S2_{\text{TCMR}}$, and $S3_{\text{ABMR}}$ scores. The AUCs for the 4AA scores' ability to predict histology diagnoses of any rejection, ABMR, or TCMR were 0.72, 0.69, and 0.67, similar to our previous report using cohort 331¹².

The molecular classifiers trained on molecular diagnoses also had low AUCs for predicting histology diagnoses (0.69, 0.75, and 0.70) and high AUCs for predicting molecular diagnoses (0.98, 0.98, and 1.00). Their AUCs for predicting molecular diagnoses made according to AA score cut-offs are expected because this is what these classifiers were trained to predict using many of the same features (transcripts) that defined the training labels. Of interest, the molecular classifiers trained on histology had low scores for histology diagnoses (0.73, 0.77, and 0.74) but higher scores for molecular diagnoses (0.88, 0.89, and 0.92), albeit not as high as classifiers trained on molecular diagnoses.

Thus, whether unsupervised, trained on molecular diagnoses, or trained on histologic diagnoses, molecular estimates of rejection gave relatively low scores for predicting histology diagnoses, compatible with the limitations imposed by the inter-observer disagreement within histology diagnoses¹.

Developing a molecular report using an ensemble of algorithms. Because diagnoses are more reliable when they include multiple independent estimates, we prepared a new report incorporating 3AA (model 1) scores, 4AA (model 2) scores, binary classifiers trained on molecular diagnosis labels from model 1 or model 2, and PCA scores (Figure 3). The report visualizes the new biopsy (based on its RAT expression) projected into PCA of the 889 biopsy reference set. Page 2 lists scores for some relevant PBTs and single genes.

One expert (PFH) assigned molecular diagnoses to all 889 biopsies taking into account model 1 and model 2 scores, the binary classifier scores, and PBT scores available on the MMDx-Heart report. Table 4 compares agreement of 885 available histologic diagnoses with the corresponding molecular expert diagnoses, broken down by reason for biopsy, with statistical analyses in Supplementary Table 8. The overall relationship between MMDx and histology was significant (Chi-square test on 7x7 matrix for all biopsies is $p=1 \times 10^{-49}$). Most histologic pTCMR cases (61%) were molecular no rejection, and only 10% were molecular TCMR-related (pTCMR 4%, TCMR 5%, or mixed 1%), indicating that histology grade 1R seldom reflects true TCMR. Conversely, 49% of histologic pABMR cases were signed out as molecular ABMR-related (pABMR 14%, ABMR 32%, or mixed 3%), indicating that pABMR is often ABMR-related. Treating pTCMR as no rejection and pABMR as ABMR-related, MMDx and pathology were 4.1 times more likely to agree than disagree on a diagnosis of rejection vs. no rejection (Fisher's exact $p=9 \times 10^{-21}$),

5.4 times more likely to agree on ABMR vs. no ABMR ($p=1 \times 10^{-22}$), and 8.1 times more likely to agree on TCMR vs. no TCMR ($p=1 \times 10^{-11}$) (Supplementary Table 8). The MMDx-histology relationships were preserved when all available protocol and indication/follow-up biopsies were analyzed separately (Table 4) despite the lower frequency of rejection in protocol biopsies. The relationship of MMDx to histology was also similar across post-transplant intervals: first year; Year 2-5 and beyond 5 years (Supplementary Table 9).

We also examined the frequencies of rejection in indication/follow-up vs protocol biopsies when one biopsy per patient was selected randomly. Using one random biopsy per patient from 162 patients biopsied for cause or follow-up, the estimated frequencies were molecular ABMR 20% (5% by histology), TCMR 11% (12% by histology), and no rejection 52% (35% by histology). This is similar to our cohort of kidney transplant indication biopsies where the frequencies of molecular ABMR and TCMR were approximately 20-29% and 7-9% respectively^{9;32}. In 306 patients who underwent protocol EMBs, the frequencies were molecular ABMR 9% (5% by histology), TCMR 4% (8% by histology), and no rejection 68% (40% by histology).

Automated sign-out that predicts expert sign-out. Because it would be useful to make automated sign-outs, we used all archetype, classifier, and PBT scores available on the report in random forest analyses that ranked the ability of each score to predict the expert's molecular diagnoses by bootstrapping thousands of decision trees (Figure 4). The random forest models predicted the expert's diagnoses of ABMR 98% of the time and TCMR 99% of the time. Model 2 S3_{ABMR} and the model 2 ABMR binary classifier score were most predictive of MMDx expert ABMR diagnoses. Model 2 S2_{TCMR}, T cell burden transcripts (TCB) score, and the model 2 binary TCMR classifier score were most predictive of MMDx expert TCMR diagnoses.

PBT scores on the report were not the main predictors of the MMDx expert's sign-out in random forests when all scores on the report were used as inputs, but they add confidence to molecular diagnoses nonetheless. In random forest analyses using only PBTs as variables, DSASTs and NKBs were most important for predicting molecular ABMR, and TCBs and QCATs were most important for predicting TCMR (Supplementary Figure 2). The PBT random forests predicted the expert's diagnoses of ABMR 95% of the time and TCMR 98% of the time.

Discussion

Having found value in molecular EMB assessment (especially after incorporating injury estimates)¹³, the present study aimed to develop an ensemble of estimates into a biopsy report and examine automated sign-out. We established that machine-learning algorithms trained in one cohort perform similarly in future biopsies despite differences in case mix. We showed that model 2 incorporating injury demonstrated similar relationships to histologic rejection as model 1 and correlated with DSA status. We then developed supervised binary classifiers to complement the unsupervised model 1 and 2 scores. We found that binary molecular classifiers trained on histology diagnoses predicted molecular diagnoses better than histology diagnoses. As expected, the classifiers trained on molecular diagnoses gave better predictions of molecular diagnoses than those trained on histology diagnoses. Indeed, no molecular score predicted histology diagnoses with an AUC>0.78, which is expected given the limits imposed by inter-observer disagreements in histology diagnoses reported in CARGO II¹. We assembled a molecular report signed out by an expert, and found significant agreement with histology but also extensive discrepancies. An automated random forest estimate accurately predicted MMDx expert sign-out of molecular rejection. We conclude that an ensemble of supervised and unsupervised molecular estimates of rejection can form a stable reporting system that avoids undue reliance on single methods, and that combining all scores into an ensemble using random forest assessment makes automated sign-outs feasible for central reporting of EMBs.

The results present a picture of the distribution of rejection phenotypes in the prevalent heart transplant population. In both indication and protocol biopsies, MMDx diagnosed ABMR and no rejection more frequently than histology. Histology diagnosed pTCMR much more frequently than MMDx, but these histology TCMR biopsies often bore little resemblance to TCMR in molecular terms. The molecular results suggest that histologic pTCMR (i.e. grade 1R) in heart transplant is seldom true TCMR, as pathologists and clinicians have long suspected. MMDx shows some positive TCMR and ABMR in biopsies where histology is negative, and vice versa. Such discrepancies are the subject of ongoing and detailed analyses in the extension of the INTERHEART study.

While no diagnostic system is perfect, we consider molecular assessment a more accurate reflection of the true disease state than histology. Molecular assessment utilizes precise measurements

of features expressed as continuous numbers, not semi-quantitative or binary scores like histology, and has high technical and biological reproducibility³³. MMDx machine learning algorithms are purely data-driven rather than derived through a consensus of opinions. Molecules outperform histology when predicting phenotypes with accurate gold standards e.g. survival^{18;34}, and simulations show that molecular classifiers can outperform flawed gold standards over a wide range of sample labeling errors. Thus, classifiers trained on molecules associated with histology are probably more accurate than the histologic diagnosis because machine learning can overcome noise in histology. MMDx analysis has provided evidence for updating the Banff guidelines for kidney transplants with C4d-negative ABMR^{15;16}. The biological relevance of the gene expression measured by microarrays can often be established²³.

The observation that binary molecular classifiers, whether trained on histology diagnoses or molecular diagnoses predict molecular diagnosis but could not predict histologic diagnoses with AUCs >0.78 probably reflects the limitations posed by the known noise in histology, including some misclassification caused by injury-induced inflammation masquerading as rejection¹³. This is lower than the agreement between MMDx and histology in kidney, as expected given that CARGO II reported lower agreement between pathologists assessing EMBs (Cohen's kappa 0.28)¹ than is usually seen with kidney transplant biopsies¹⁷, probably because EMBs are more challenging to assess histologically than kidney core biopsies. Nevertheless, all machine learning molecular estimates, including unsupervised archetype scores and supervised binary classifiers trained on histology or MMDx diagnoses, were significantly associated with histology diagnoses, confirming that there is substantial truth in histology.

Caution is warranted in benchmarking the performance of a new molecular system against a standard with known problems such as histology. Seeking maximum agreement between molecular tests and a flawed conventional test would generate a molecular system that retains the errors of the conventional test. The MMDx phenotype is real and reproducible, and analysis of discrepancies with histology will improve understanding. We propose that the MMDx system and the histology system can develop reciprocally, each offering a platform to help develop the other. With this in mind, a detailed analysis of the MMDx-histology discrepancies is in progress, with the goal of eventually improving both systems. One limitation of this study is that the clinical implications need to be further explored, and will be addressed in the ongoing extension of the INTERHEART study.

Acknowledgments and Sources of Funding

This research has been supported by funding and/or resources from Genome Canada, Canada Foundation for Innovation, the University of Alberta Hospital Foundation, the Alberta Ministry of Advanced Education and Technology, the Mendez National Institute of Transplantation Foundation, Industrial Research Assistance Program and the Roche Organ Transplant Research Foundation. Partial support was also provided by funding from a licensing agreement with the One Lambda division of Thermo Fisher. Dr. Halloran held a Canada Research Chair in Transplant Immunology until 2008 and currently holds the Muttart Chair in Clinical Immunology.

Authors thank Jeff Reeve for his critical revision of manuscript and Jessica Chang for preparing and editing the final version of the manuscript.

Disclosure

P F Halloran holds shares in Transcriptome Sciences Inc., a University of Alberta research company with an interest in molecular diagnostics and has been a symposium speaker for One Lambda. The other authors have declared no conflict of interest exists.

Reference List

1. Crespo-Leiro MG, Zuckermann A, Bara C, Mohacsi P, Schulz U, Boyle A, Ross HJ, Parameshwar J, Zakliczynski M, Fiocchi R, Stypmann J, Hofer D, Lehmkuhl H, Deng MC, LePrince P, Berry G, Marboe CC, Stewart S, Tazelaar HD, Baron HM, Coleman IC, Vanhaecke J: Concordance Among Pathologists in the Second Cardiac Allograft Rejection Gene Expression Observational Study (CARGO II). *Transplant* 94:1172-1177, 2012
2. Stewart S, Winters GL, Fishbein MC, Tazelaar HD, Kobashigawa J, Abrams J, Andersen CB, Angelini A, Berry GJ, Burke MM, Demetris AJ, Hammond E, Itescu S, Marboe CC, McManus B, Reed EF, Reinsmoen NL, Rodriguez ER, Rose AG, Rose M, Suci-Focia N, Zeevi A, Billingham ME: Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant* 24:1710-1720, 2005
3. Billingham M, Kobashigawa JA: The revised ISHLT heart biopsy grading scale. *J Heart Lung Transplant* 24:1709, 2005
4. Mehra MR, Canter CE, Hannan MM, Semigran MJ, Uber PA, Baran DA, nziger-Isakov L, Kirklin JK, Kirk R, Kushwaha SS, Lund LH, Potena L, Ross HJ, Taylor DO, Verschuuren EA, Zuckermann A: The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: A 10-year update. *J Heart Lung Transplant* 35:1-23, 2016
5. Leone O, Veinot JP, Angelini A, Baandrup UT, Basso C, Berry G, Bruneval P, Burke M, Butany J, Calabrese F, d'Amati G, Edwards WD, Fallon JT, Fishbein MC, Gallagher PJ, Halushka MK, McManus B, Pucci A, Rodriguez ER, Saffitz JE, Sheppard MN, Steenbergen C, Stone JR, Tan C, Thiene G, van der Wal AC, Winters GL: 2011 consensus statement on endomyocardial biopsy from the Association for European Cardiovascular Pathology and the Society for Cardiovascular Pathology. *Cardiovasc Pathol* 21:245-274, 2012
6. Bruneval P, Angelini A, Miller D, Potena L, Loupy A, Zeevi A, Reed EF, Dragun D, Reinsmoen N, Smith RN, West L, Tebutt S, Thum T, Haas M, Mengel M, Revelo P, Fedrigo M, Duong Van Huyen JP, Berry GJ: The XIIIth Banff Conference on Allograft Pathology: The Banff 2015 Heart Meeting Report: Improving Antibody-Mediated Rejection Diagnostics: Strengths, Unmet Needs, and Future Directions. *Am J Transplant* 17:42-53, 2017
7. Mengel M, Sis B, Kim D, Chang J, Famulski K, Hidalgo L, Einecke G, De Freitas D, Tymchak W, Burton J, Halloran PF: The molecular phenotype of heart transplant biopsies: relationship to histopathological and clinical variables. *Am J Transplant* 10:2105-2115, 2010
8. Loupy A, Duong Van Huyen JP, Hidalgo LG, Reeve J, Racape M, Venner J, Famulski K, Bories MC, Beuscart T, Guillemain R, Francois A, Pattier S, Toquet C, Gay A, Rouvier P, Varnous S, LePrince P, Empana JP, Lefaucheur C, Bruneval P, Jouven X, Halloran PF: Gene Expression Profiling for the Identification and Classification of Antibody-Mediated Heart Rejection. *Circulation* 135:917-935, 2017
9. Reeve J, Bohmig GA, Eskandary F, Einecke G, Lefaucheur C, Loupy A, Halloran PF, MMDx-Kidney study group: Assessing rejection-related disease in kidney transplant biopsies based on archetypal analysis of molecular phenotypes. *JCI Insight* 2:e94197, 2017
10. Halloran PF, Famulski K, Reeve J: The molecular phenotypes of rejection in kidney transplant biopsies. *Current Opinion In Organ Transplantation* 20:359-367, 2015
11. Halloran PF, Reeve J, Pereira AB, Hidalgo LG, Famulski KS: Antibody-mediated rejection, T cell-mediated rejection, and the injury-repair response: new insights from the Genome Canada studies of kidney transplant biopsies. *Kidney Int* 258-264, 2014

12. Halloran PF, Potena L, Duong Van Huyen JP, Bruneval P, Leone O, Kim DH, Jouven X, Reeve J, Loupy A: Building a tissue-based molecular diagnostic system in heart transplant rejection: the heart molecular microscope MMDx. *J Heart Lung Transplant* 36:1192-1200, 2017
13. Halloran PF, Aliabadi AZ, Cadeiras M, Crespo-Leiro MG, Deng M, Depasquale EC, Goekler J, Jouven X, Kim DH, Kobashigawa J, Loupy A, Macdonald P, Potena L, Zuckermann A, Parkes MD: Exploring the cardiac response-to-injury in heart transplant biopsies. *JCI Insight* 3:e123674, 2018
14. Sis B, Jhangri GS, Riopel J, Chang J, de Freitas DG, Hidalgo L, Mengel M, Matas A, Halloran PF: A new diagnostic algorithm for antibody-mediated microcirculation inflammation in kidney transplants. *Am J Transplant* 12:1168-1179, 2012
15. Sis B, Jhangri G, Bunnag S, Allanach K, Kaplan B, Halloran PF: Endothelial gene expression in kidney transplants with alloantibody indicates antibody-mediated damage despite lack of C4d staining. *Am J Transplant* 9:2312-2323, 2009
16. Einecke G, Sis B, Reeve J, Mengel M, Campbell PM, Hidalgo LG, Kaplan B, Halloran PF: Antibody-Mediated Microcirculation Injury Is the Major Cause of Late Kidney Transplant Failure. *Am J Transplant* 9:2520-2531, 2009
17. Reeve J, Sellares J, Mengel M, Sis B, Skene A, Hidalgo L, De Freitas D, Famulski K, Halloran PF: Molecular diagnosis of T cell-mediated rejection in human kidney transplant biopsies. *Am J Transplant* 13:645-655, 2013
18. Sellares J, Reeve J, Loupy A, Mengel M, Sis B, Skene A, De Freitas D, Kreepala C, Hidalgo L, Famulski K, Halloran PF: Molecular diagnosis of antibody-mediated rejection in human kidney transplants. *Am J Transplant* 13:971-983, 2013
19. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second ed, Springer, 2009
20. Oza NC, Turner K: Classifier ensembles: Select real-world applications. *Information Fusion* 9:4-20, 2008
21. Rokach L: Ensemble Methods in Supervised Learning. In: *Data Mining and Knowledge Discovery Handbook*, edited by Maimon O, Rokach L, Boston, MA, Springer US, 2010, pp 959-979
22. Berry GJ, Burke MM, Andersen C, Bruneval P, Fedrigo M, Fishbein MC, Goddard M, Hammond EH, Leone O, Marboe C, Miller D, Neil D, Rassl D, Revelo MP, Rice A, Rene RE, Stewart S, Tan CD, Winters GL, West L, Mehra MR, Angelini A: The 2013 International Society for Heart and Lung Transplantation Working Formulation for the standardization of nomenclature in the pathologic diagnosis of antibody-mediated rejection in heart transplantation. *J Heart Lung Transplant* 32:1147-1162, 2013
23. Halloran PF, Venner JM, Famulski KS: Comprehensive analysis of transcript changes associated with allograft rejection: Combining universal and selective features. *Am J Transplant* 17:1754-1769, 2017
24. Famulski KS, de Freitas DG, Kreepala C, Chang J, Sellares J, Sis B, Mengel M, Reeve J, Halloran PF: Molecular phenotypes of acute kidney injury in human kidney transplants. *JASN* 23:948-958, 2012
25. Venner JM, Famulski KS, Reeve J, Chang J, Halloran PF: Relationships among injury, fibrosis, and time in human kidney transplants. *Journal of Clinical Investigation Insight* 1:e85323-
doi:10.1172/jci.insight.85323., 2016

26. Famulski KS, Einecke G, Sis B, Mengel M, Hidalgo LG, Kaplan B, Halloran PF: Defining the Canonical Form of T-Cell-Mediated Rejection in Human Kidney Transplants. *Am J Transplant* 10:810-820, 2010
27. Hidalgo LG, Einecke G, Allanach K, Mengel M, Sis B, Mueller TF, Halloran PF: The transcriptome of human cytotoxic T cells: measuring the burden of CTL-associated transcripts in human kidney transplants. *Am J Transplant* 8:637-646, 2008
28. Hidalgo LG, Sellares J, Sis B, Mengel M, Chang J, Halloran PF: Interpreting NK cell transcripts versus T cell transcripts in renal transplant biopsies. *Am J Transplant* 12:1180-1191, 2012
29. Famulski KS, Einecke G, Reeve J, Ramassar V, Allanach K, Mueller T, Hidalgo LG, Zhu L-F, Halloran PF: Changes in the transcriptome in allograft rejection: IFN- γ induced transcripts in mouse kidney allografts. *Am J Transplant* 6:1342-1354, 2006
30. Hidalgo LG, Sis B, Sellares J, Campbell PM, Mengel M, Einecke G, Chang J, Halloran PF: 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* 10:1812-1822, 2010
31. Cutler A, Breiman L: Archetypal Analysis. *Technometrics* 36:338-347, 1994
32. Halloran PF, Akalin E, Aubert O, Bohmig G, Brennan D, Bromberg J, Einecke G, Eskandary F, Gosset C, Duong Van Huyen JP, Gupta G, Lefaucheur C, Malone A, Seron D, Sellares J, Weir M, Loupy A: Real time central assessment of kidney transplant indication biopsies by microarrays: the INTERCOMEX Study. *Am J Transplant* 17:2851-2862, 2017
33. Madill-Thomsen KS, Wiggins RC, Eskandary F, Bohmig GA, Halloran PF: The effect of cortex/medulla proportions on molecular diagnoses in kidney transplant biopsies: rejection and injury can be assessed in medulla. *Am J Transplant* 17:2117-2128, 2017
34. Einecke G, Reeve J, Sis B, Mengel M, Hidalgo L, Famulski K, Matas A, Kasiske B, Kaplan B, Halloran PF: A molecular classifier for predicting future graft loss in late kidney transplant biopsies. *The Journal of Clinical investigation* 120:1862-1872, 2010

Figure Legends

Figure 1. Comparison of molecular scores between histologic classes in cohort 331 and cohort 558 using the 3-archetype model scores derived by supervised archetype analysis in cohort 331.

The plots on the left-hand side (A, C, E) compare relationships between histology classes and 3AA $S1_{normal}$ (A), $S2_{TCMR}$ (C), and $S3_{ABMR}$ scores (E) in cohort 331. The plots on the right-hand side (B, D, and F) compare these same relationships in cohort 558. The $S1_{normal}$, $S2_{TCMR}$, and $S3_{ABMR}$ scores in both cohorts were taken from 3AA models trained on cohort 331. Mixed rejection was included in the TCMR and ABMR histology groups; possible mixed rejection was included in the pTCMR and pABMR groups. Asterisks below the plot labels indicate the significance of the Mann-Whitney U test between the pairs: ns – not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$.

Figure 2. Comparison of molecular scores between histologic classes using the 3-archetype (3AA) and 4-archetype (4AA) model scores derived from algorithms trained in cohort 889. Results are shown for cohort 889. Panels A and B compare TCMR histology classes with $S2_{TCMR}$ scores in 3AA (A) and 4AA (B). Panels C and D compare ABMR histology classes with $S3_{ABMR}$ scores in 3AA (C) and 4AA (D). 4AA $S4_{injury}$ scores are compared with TCMR histology classes in panel E and with ABMR classes in panel F. Mixed rejection was included in the TCMR and ABMR histology groups; possible mixed rejection was included in the pTCMR and pABMR groups. Asterisks below the x-axis labels indicate the significance of the Mann-Whitney U test between the indicated pairs of histology groups: ns – not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$.

Figure 3. Molecular Microscope[®] Report for heart transplant biopsies (MMDx-Heart). The new biopsy (yellow triangle) is compared to the reference set of 889 endomyocardial biopsies and given a series of molecular scores and assigned a molecular interpretation. Patient information has been redacted. Archetype scores $S1_{normal}$, $S2_{TCMR}$, $S3_{ABMR}$, and $S4_{injury}$ from the 3-archetype model (3AA/Model 1) and the 4-archetype model (4AA/Model 2) are given for the new biopsy, in addition to corresponding binary classifier scores predicting the probability of molecular non-rejection, TCMR, and ABMR, trained

on the molecular assessments. The report visualizes of the new biopsy projected into the rejection-associated transcript-based principal component analysis of the 889 reference set biopsies. Biopsies in the reference set are colored according to their highest 4AA archetype scores: grey for $S1_{normal}$; red for $S2_{TCMR}$; blue for $S3_{ABMR}$; and cyan for $S4_{Injury}$. Page 2 of the report (right side of the figure) provides additional molecular data including pathogenesis-based transcript (PBT) set scores and transcript expression scores relating to all rejection, ABMR, TCMR, and injury. Score are represented as the log fold change in the new biopsy vs. normal biopsies (i.e. reference set biopsies with $S1_{normal} > 0.7$). For each score a normal limit is given, defined as the 95th percentile score in the normal biopsies. Scores in the 95th-99th percentile are labeled “slightly abnormal” and scores in the 99th percentile are labeled “abnormal.”

Figure 4. Relative predictive value of molecular scores used as variables in random forest predictions of molecular diagnoses assigned by an expert. Random forest analyses used molecular scores available on the report and time of biopsy post-transplant as features. Random forests were built using 2000 trees. Error rates with 2000 trees were stable. Model 1 corresponds to the RAT-based 3-archetype model trained in 889 hearts, and model 2 corresponds to the 4-archetype model. The classifiers refer to the binary classifiers trained on biopsy diagnoses assigned using model 1 or model 2 based on score cut-offs of $S3_{ABMR} \geq 0.5$ for ABMR and $S2_{TCMR} \geq 0.3$ for TCMR. (A) Relative importance of variables in predicting expert-assigned molecular ABMR diagnoses. (B) Relative importance of variables in predicting expert-assigned molecular TCMR diagnoses.

Supplementary Figure 1. Outline of the strategy in present study.

Supplementary Figure 2. Relative predictive value of PBT scores used as variables in random forest predictions of molecular diagnoses assigned by an expert. Random forest analyses used PBT scores available on the report. Random forests were built using 2000 trees. Error rates with 2000 trees were stable. (A) Relative importance of variables in predicting expert-assigned ABMR diagnoses. (B) Relative importance of variables in predicting expert-assigned TCMR diagnoses.

Figure 1

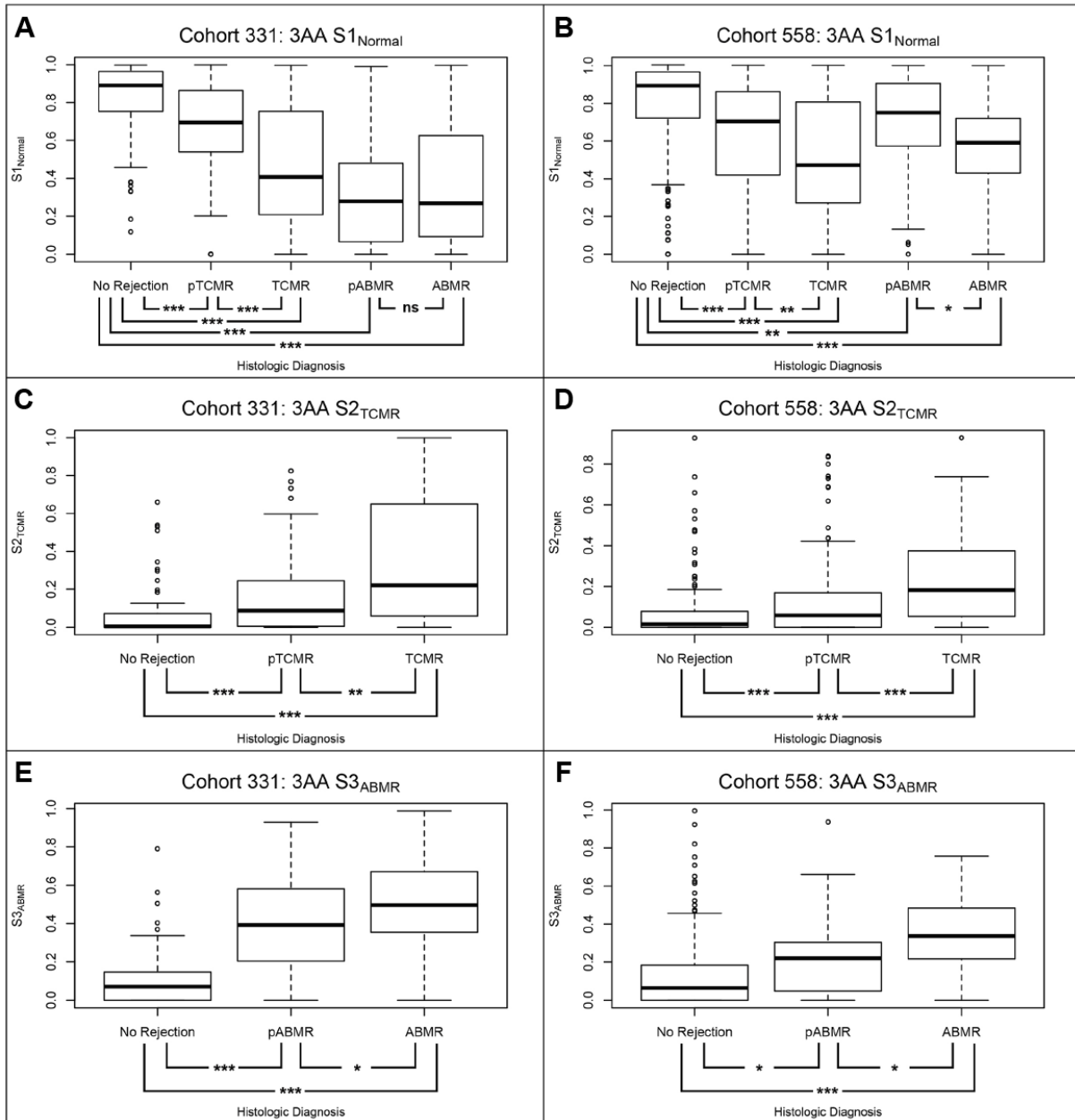
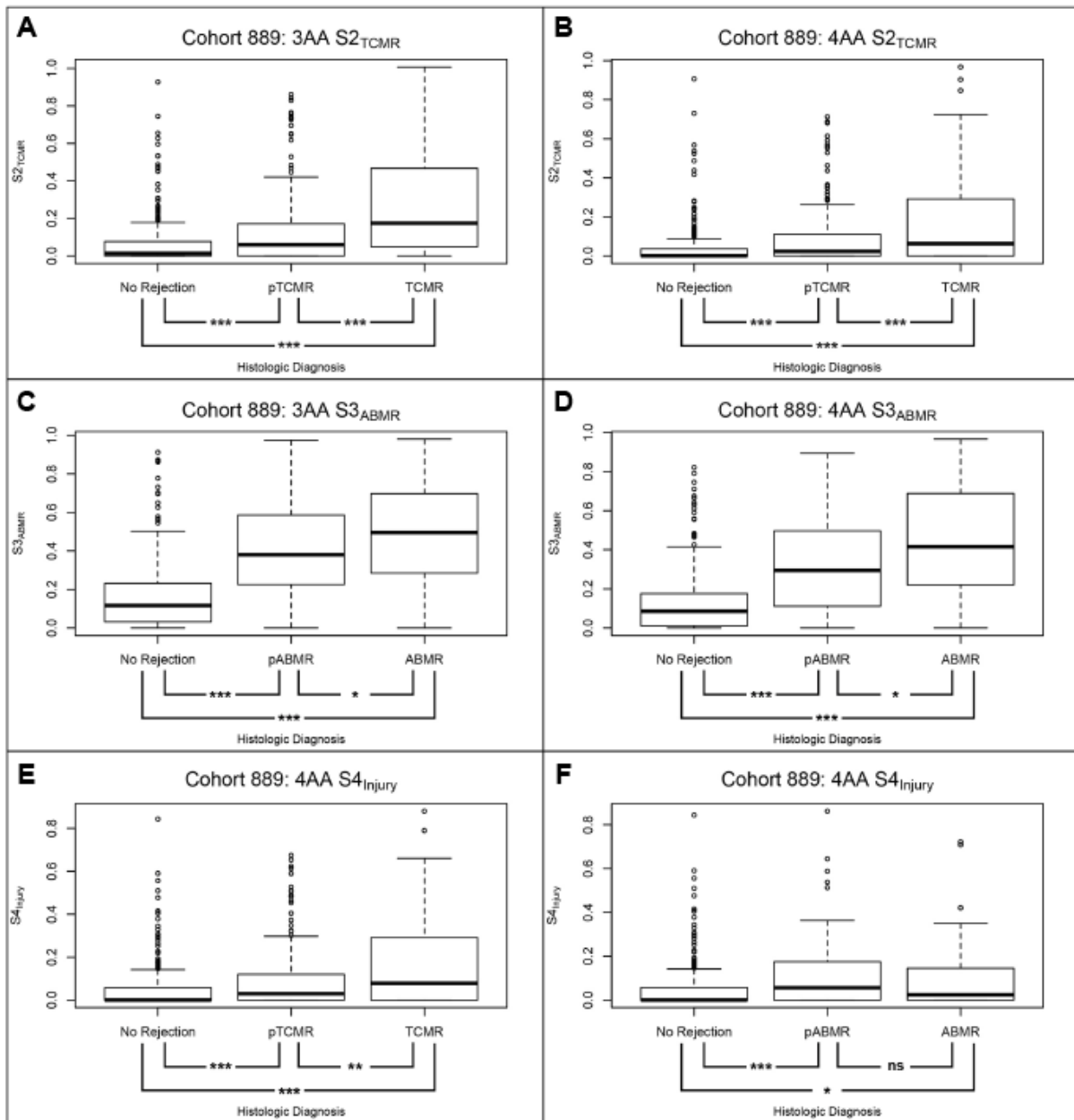


Figure 2



AC

IPT

Figure 3

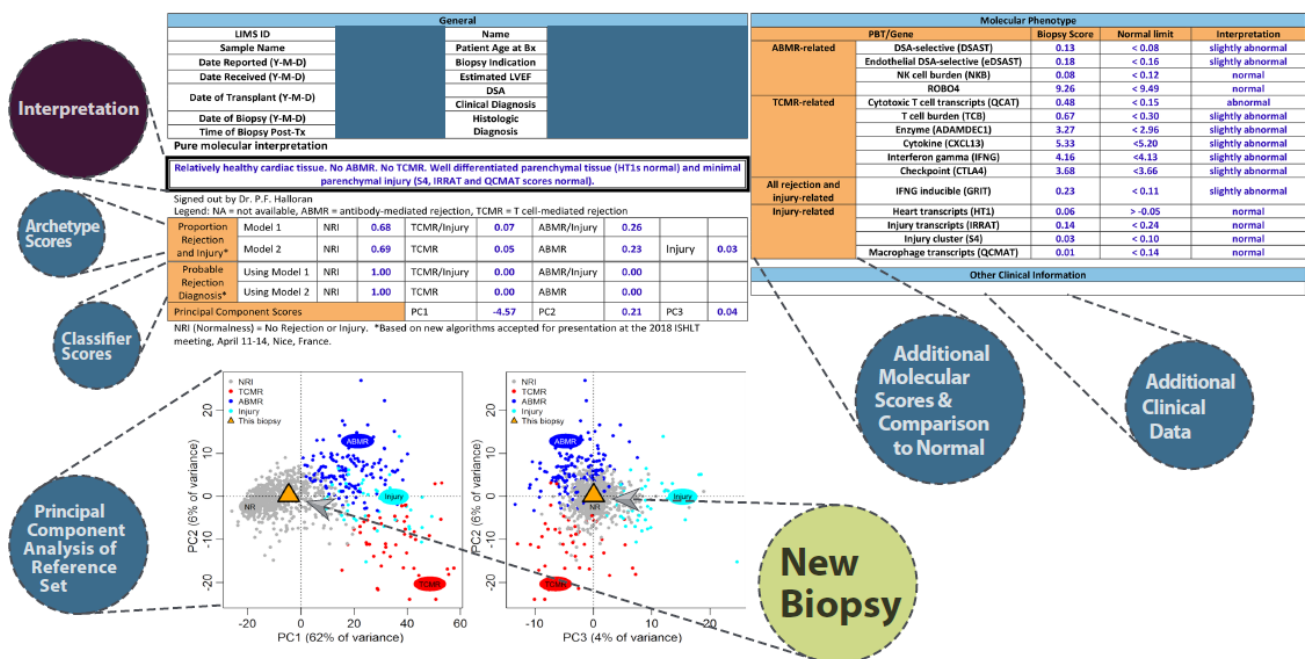
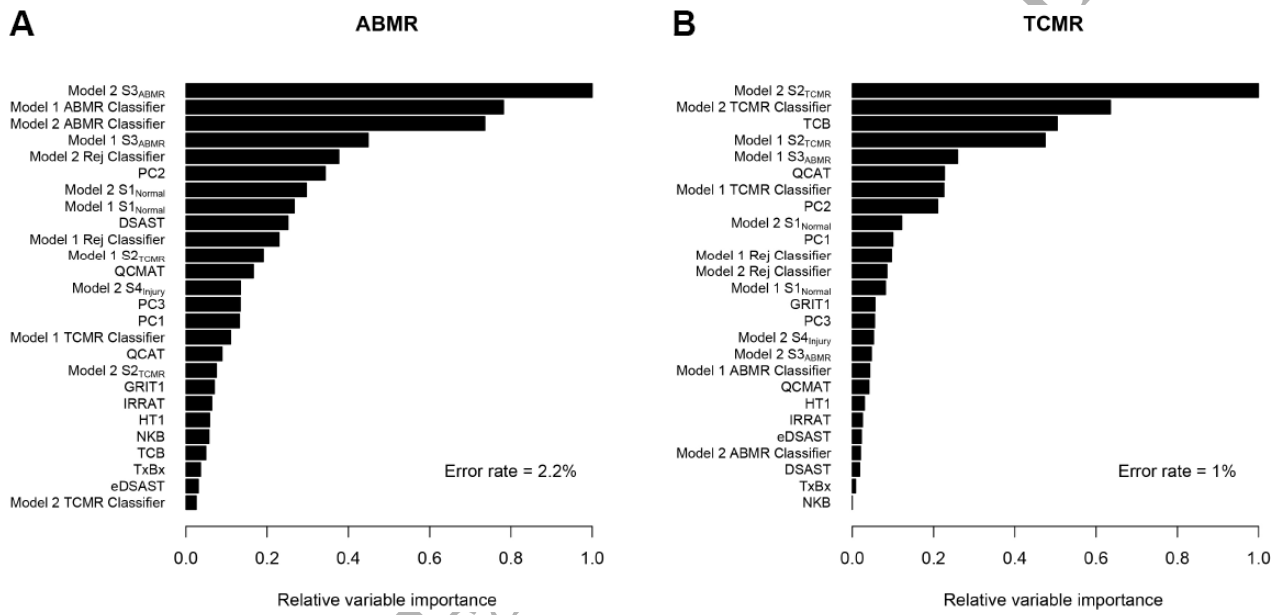


Figure 4



Biopsy Characteristics	Cohort 331 (331 Biopsies)	Cohort 558 (558 Biopsies)
Days to biopsy post-transplant (TxBx)		
Mean	984	750
Median (range)	337 (8 – 10150)	197 (6 – 9745)
Days to most recent follow-up after biopsy		
Mean	1522	358
Median (range)	1282 (13 – 3854)	299 (1 – 3713)
Indication for biopsy		
Clinical (% of known)	62 (19%)	92 (17%)
Follow-up (% of known)	19 (6%)	89 (16%)
Protocol biopsy (% of known)	238 (75%)	375 (67%)
Not recorded (% of total)	12 (4%)	2 (<1%)
Patient Characteristics*	Cohort 331* (210 Patients)	Cohort 558* (273 Patients)
Mean patient age at first biopsy (range)	49 (2 – 78)	53 (16 – 76)
Mean donor age (range)	40 (6 – 66)	41 (15 – 68)
Patient sex		
Male (% of known)	147 (70%)	192 (73%)
Female (% of known)	63 (30 %)	71 (27%)
Not available (% of total)	0	10 (4%)
Donor sex		
Male (% of known)	138 (68%)	155 (66%)
Female (% of known)	65 (32%)	81 (34%)
Not available (% of total)	7 (3%)	37 (14%)
Patient had a previous failed heart transplant	5 (2%)	8 (3%)
Transplant status		
Alive at last follow up (% of known)	173 (82%)	247 (90%)
Deceased (% of known)	36 (18%)	22 (9%)
Failed and retransplanted (% of known)	1 (<1%)	3 (1%)
Not available (% of total)	0	1 (<1%)
Primary disease†		
Dilated cardiomyopathy (% of known)	96 (46%)	134 (55%)
Hypertrophic cardiomyopathy (% of known)	11 (5%)	21 (9%)
Restrictive cardiomyopathy (% of known)	3 (1%)	5 (2%)
Other cardiomyopathy (% of known)	21 (10%)	18 (7%)
Congenital heart defect (% of known)	17 (8%)	14 (6%)
Coronary artery disease (% of known)	50 (24%)	35 (14%)
Other (% of known)	28 (13%)	18 (7%)
Not recorded (% of total)	1 (<1%)	28 (10%)

* Some patients were part of both cohorts because biopsies were taken from them during both legs of the INTERHEART study. The total number of unique patients between both cohorts is 454.

† Some patients fell under multiple categories

Table 2. Histology and DSA

Histology diagnoses* (% of known diagnoses)		Cohort 331 (331 biopsies)	Cohort 558 (558 biopsies)
TCMR Related	TCMR	36 (11%)	48 (9%)
	pTCMR	90 (27%)	183 (33%)
ABMR Related	ABMR	33 (10%)	18 (3%)§
	pABMR	44 (13%)	19 (3%)§
Other	ABMR/TCMR (Mixed)	5 (2%)	4 (1%)
	pABMR/pTCMR	34 (10%)	37 (7%)
No Rejection		89 (27%)	245 (44%)§
Missing		0 (0%)	4 (1%)
DSA Status		Cohort 331 [†] (210 patients)	Cohort 558 [†] (273 patients)
DSA status at most recent biopsy[‡]			
<i>Positive</i>		69 (35%)	94 (37%)
<i>Negative</i>		130 (65%)	161 (63%)
<i>Unknown</i>		11 (5%)	18 (7%)
* Biopsies in the 889 cohort were labeled as follows:			
pAMR0..... No ABMR;			
pAMR1, pAMR1+, pAMR1H+..... Possible ABMR (pABMR);			
pAMR2, pAMR3..... ABMR;			
TCMR0R..... No TCMR;			
TCMR1R..... Possible TCMR (pTCMR);			
TCMR2R, TCMR3R..... TCMR			
† Some patients were in both cohorts because biopsies were taken from them during both legs of the INTERHEART study. The total number of unique patients between both cohorts is 454.			
‡ The most recent DSA status at time of most recent biopsy was used, if known. DSA statuses dated more than 14 days after the biopsy were not considered. If the most recent DSA status at time of biopsy was not known, but the patient was most recently PRA negative, the DSA status was presumed negative.			
§ p-value<0.001 by Chi-square test of rejection vs no rejection between cohort 331 and cohort 558			

Algorithms	Prediction tested	Areas under the receiver-operator characteristic curve (AUCs) for classifiers predicting the diagnosis of:		
		All Rejection ^{*,†} (ABMR, TCMR, Mixed)	ABMR [*] (ABMR, Mixed)	TCMR [*] (TCMR, Mixed)
Control: unsupervised 4AA Scores <i>S1_{normal}, S2_{TCMR}, and S3_{ABMR}[‡]</i>	<i>Histologic diagnoses</i>	0.72	0.69	0.67
Supervised classifiers trained on:	Prediction tested	All Rejection^{*,†} (ABMR, TCMR, Mixed)	ABMR[*] (ABMR, Mixed)	TCMR[*] (TCMR, Mixed)
<i>Histologic diagnoses^{*,‡}</i>	<i>Histologic diagnoses</i>	0.73	0.77	0.74
	<i>Molecular diagnoses</i>	0.89	0.88	0.92
<i>Molecular diagnoses^{*,§}</i>	<i>Histologic diagnoses</i>	0.69	0.75	0.70
	<i>Molecular diagnoses</i>	0.98	0.98	1.00

* All rejection = ABMR, TCMR, Mixed (ABMR/TCMR), ABMR/pTCMR, and pABMR/TCMR vs all other biopsies. ABMR = ABMR, Mixed, and ABMR/pTCMR vs all other biopsies. TCMR = TCMR, Mixed, and pABMR/TCMR vs all other biopsies.

[†] Molecular scores derived from the RAT-based four-archetype model of rejection. For classification purposes, we used cut-offs of S2_{TCMR}≥0.3 for TCMR, S3_{ABMR}≥0.5 for ABMR, and either cut-off for all rejections.

[‡] Genes used: top 20 transcripts associated with histologic diagnoses. TCMR included biopsies with histologic TCMR grades > 1R, ABMR included biopsies with histologic ABMR grades > 1, and either cut-off for all rejections.

[§] Genes used: top 20 transcripts associated with molecular diagnoses based on the four-archetype model of rejection using cut-offs of S2_{TCMR}≥0.3 for TCMR, S3_{ABMR}≥0.5 for ABMR, and either cut-off for all rejections.

Table 4. Comparing expert-assigned MMDx diagnoses with histology diagnoses in 885/889 endomyocardial biopsies, 259 for cause/follow-up endomyocardial biopsies, and 612 protocol endomyocardial biopsies with available histology assessments

885/889 endomyocardial biopsies with available histology assessments							
Expert diagnoses (% of row)*							
Histology diagnoses†	TCMR (N=51)	pTCMR (N=33)	ABMR (N=106)	pABMR (N=120)	ABMR/TCMR (Mixed) (N=11)	pABMR/pTCMR (N=11)	No Rejection (N=553)
TCMR (N=84)	18 (21%)	9 (11%)	8 (10%)	4 (5%)	1 (1%)	0 (0%)	44 (52%)
pTCMR (N=273)	14 (5%)	10 (4%)	27 (10%)	47 (17%)	2 (1%)	6 (2%)	167 (61%)
ABMR (N=51)	0 (0%)	1 (2%)	23 (45%)	15 (29%)	0 (0%)	0 (0%)	12 (24%)
pABMR (N=63)	5 (8%)	4 (6%)	20 (32%)	9 (14%)	2 (3%)	2 (3%)	21 (33%)
ABMR/TCMR (Mixed) (N=9)	6 (67%)	0 (0%)	1 (11%)	1 (11%)	1 (11%)	0 (0%)	0 (0%)
pABMR/pTCMR (N=71)	3 (4%)	1 (1%)	13 (18%)	21(30%)	2 (3%)	1 (1%)	30 (42%)
No rejection (N=334)	5 (1%)	8 (2%)	14 (4%)	23 (7%)	3 (1%)	2 (1%)	279 (84%)
259 for cause/follow-up endomyocardial biopsies with available histology assessments							
Expert diagnoses (% of row)*							
Histology diagnoses	TCMR (N=26)	pTCMR (N=11)	ABMR (N=51)	pABMR (N=30)	ABMR/TCMR (Mixed) (N=6)	pABMR/pTCMR (N=5)	No Rejection (N=130)
TCMR (N=26)	8 (31%)	2 (8%)	3 (12%)	1 (4%)	0 (0%)	0 (0%)	12 (46%)
pTCMR (N=74)	7 (9%)	3 (4%)	12 (16%)	15 (20%)	0 (0%)	4 (5%)	33 (45%)
ABMR (N=16)	0 (0%)	1 (6%)	7 (44%)	4 (25%)	0 (0%)	0 (0%)	4 (25%)
pABMR (N=27)	2 (7%)	3 (11%)	13 (48%)	1 (4%)	0 (0%)	1 (4%)	7 (26%)
ABMR/TCMR (Mixed) (N=7)	4 (57%)	0 (0%)	1 (14%)	1 (14%)	1 (14%)	0 (0%)	0 (0%)
pABMR/pTCMR (N=18)	2 (11%)	0 (0%)	7 (39%)	5 (28%)	2 (11%)	0 (0%)	2 (11%)
No rejection (N=91)	3 (3%)	2 (2%)	8 (9%)	3 (3%)	3 (3%)	0 (0%)	72 (79%)
612 protocol endomyocardial biopsies with available histology assessments							
Expert diagnoses (% of row)*							
Histology diagnoses	TCMR (N=25)	pTCMR (N=22)	ABMR (N=50)	pABMR (N=88)	ABMR/TCMR (Mixed) (N=5)	pABMR/pTCMR (N=6)	No Rejection (N=416)
TCMR (N=58)	10 (17%)	7 (12%)	5 (9%)	3 (5%)	1 (2%)	0 (0%)	32 (55%)
pTCMR (N=198)	7 (4%)	7 (4%)	15 (8%)	31 (16%)	2 (1%)	2 (1%)	134 (68%)
ABMR (N=30)	0 (0%)	0 (0%)	11 (37%)	11 (37%)	0 (0%)	0 (0%)	8 (27%)
pABMR (N=35)	3 (9%)	1 (3%)	7 (20%)	7 (20%)	2 (6%)	1 (3%)	14 (40%)
ABMR/TCMR (Mixed) (N=2)	2 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
pABMR/pTCMR (N=53)	1 (2%)	1 (2%)	6 (11%)	16 (30%)	0 (0%)	1 (2%)	28 (53%)
No rejection (N=236)	2 (1%)	6 (3%)	6 (3%)	20 (8%)	0 (0%)	2 (1%)	200 (85%)

* Diagnoses assigned by expert following individual review of molecular assessments. Diagnoses were translated from the "pure molecular interpretation" text on the report based on the following phrases:

"low probability of" possible rejection (pABMR and/or pTCMR);
 "possible" possible rejection (pABMR and/or pTCMR);
 "cannot exclude" possible rejection (pABMR and/or pTCMR);
 "[ABMR/TCMR]-like changes" possible rejection (pABMR and/or pTCMR);
 "probable" rejection (ABMR and/or TCMR);
 "mild/moderate/extensive/severe" rejection (ABMR and/or TCMR)

† Chi-square test on 7x7 matrix is $p=1.3 \times 10^{-49}$