Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Continuous Transcriptomic Indices for Assessing the Spectrum of Kidney Transplant Rejection: International Derivation and Validation Study

Maarten Naesens

maarten.naesens@uzleuven.be

Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium
https://orcid.org/0000-0002-5625-0792

Karolien Wellekens

KU Leuven

Thibaut Vaulet

KU Leuven https://orcid.org/0000-0002-1226-3153

Hao zhang

CareDX

Angelica pagliazzi

KU Leuven

Jetty de Loor

KU Leuven

Priyanka Koshy

UZ Leuven https://orcid.org/0000-0002-2313-5122

Dirk Kuypers

Department of Microbiology, Immunology and Transplantation, KU Leuven

Thomas Vanhoutte

KU Leuven

Arthur Vranken

Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

Maarten Coemans

KU Leuven https://orcid.org/0000-0001-8442-3673

Chris larsen

Arkana Laboratories

Clarissa Cassol

Arkana Laboratories https://orcid.org/0000-0001-7536-6806

Robert woodward

Article

Keywords:

Posted Date: March 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9012525/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Kidney allograft rejection is currently defined by categorical Banff histologic thresholds, despite reflecting a continuous spectrum of immune injury. Here we developed four continuous molecular indices capturing rejection severity and injury stage: antibody-mediated microvascular injury (AMR/MVI), T cell–mediated tubulointerstitial inflammation (TCMR/TI), inflammatory activity and chronic injury. Using transcriptomic profiling of formalin-fixed biopsies with the Banff Human Organ Transplant (B-HOT) panel, indices were trained against continuous histologic scores in 415 biopsies and validated in independent European and US cohorts (total N=2,076). Molecular indices accurately discriminated Banff phenotypes while revealing substantial within-category heterogeneity and clinically meaningful reclassification of intermediate cases. Elevated molecular scores in histologically rejection-free biopsies anticipated subsequent rejection. In multivariable Cox models, all four indices independently predicted five-year graft failure and improved prognostic discrimination beyond histology alone. These continuous molecular indices provide a quantitative framework that complements biopsy interpretation and refines risk stratification across the rejection spectrum.

Introduction

Despite advances in non-invasive monitoring of kidney transplant injury, kidney allograft biopsies remain central to diagnosing rejection and guiding immunosuppression. The Banff Classification provides an internationally standardized framework by categorizing biopsies into antibody-mediated rejection and microvascular inflammation (AMR/MVI) and T cell–mediated rejection (TCMR) based on semiquantitative lesion scores and diagnostic thresholds.[1,2] However, rejection biology is neither dichotomous nor static. Alloimmune mechanisms overlap with non-alloimmune injury pathways and discrete Banff categories can obscure clinically relevant heterogeneity; small differences in lesion scoring may shift biopsies across diagnostic categories, while biologically distinct cases may share the same diagnostic label,[3] limiting precision for risk stratification and individualized management.[3]

To better align biopsy interpretation with continuous disease biology, continuous histologic indices have recently been proposed (AMR/MVI, TCMR/TI, activity and chronicity indices), representing rejection phenotypes and injury stage along a spectrum rather than as discrete categories.[4-7] These indices, derived from Banff lesion scores, provide granularity beyond categorical diagnoses and independently associate with graft failure risk, even within Banff-defined rejection groups. Nevertheless, histological assessment remains constrained by sampling error, interobserver variability and limited sensitivity.

Molecular profiling of kidney transplant biopsies offers an opportunity to complement histology. In 2019, the Banff Molecular Diagnostics Working Group introduced the Banff Human Organ Transplant (B-HOT) panel, a curated transcript set capturing key alloimmune pathways and enabling standardized gene expression analysis from formalin-fixed paraffin-embedded tissue.[8-10] Molecular approaches such as the Molecular Microscope Diagnostic System (MMDx) and related transcriptomic classifiers have largely been developed to discriminate Banff diagnostic categories.[11-17] However, categorical modelling may

mask meaningful variation in molecular signal intensity, particularly near diagnostic cutoffs, in intermediate phenotypes, or in biopsies with mixed rejection features.

We therefore hypothesized that modelling intragraft gene expression as continuous molecular indices, rather than aligning transcriptomic signatures to categorical disease categories, would better capture the full spectrum and severity of rejection. To test this, we developed four a priori biopsy-based molecular indices reflecting antibody-mediated microvascular injury (AMR/MVI), T cell–mediated tubulointerstitial inflammation (TCMR/TI), inflammatory activity and chronic injury.

We show that these continuous molecular indices (i) align with individual histologic lesion scores and previously defined composite histologic indices; (ii) accurately discriminate the major Banff diagnostic categories, including mixed rejection, with good calibration in independent European and US cohorts; (iii) provide clinically meaningful reclassification of biopsies with intermediate phenotypes; and (iv) are interpretable across their full range, improving prediction of clinically relevant outcomes, including graft failure, beyond histology alone. Together, these four molecular indices capture rejection severity along a continuum and offer a quantitative complement to conventional biopsy assessment.

Results

Cohort characteristics

The derivation and validation cohorts included 415 retrospectively selected biopsies for model training, and 471 biopsies from 338 recipients prospectively collected for independent validation, respectively (**Supplementary Fig. S1**). The US validation cohort included 1,190 biopsies. The clinical and histological characteristics are summarized in **Table 1**. In the derivation cohort, 90.1% of transplants were from deceased donors, 18.8% were repeat transplants, and 12.2% of recipients had pre-transplant HLA-DSA; the corresponding values in the prospective validation cohort were 86.7%, 12.1%, and 7.1%. Clinical metadata were not available for the US cohort.

Banff 2022-defined diagnoses were made in 45.5% of the biopsies in the derivation cohort, with representation of incomplete (borderline changes, isolated v lesions, probable AMR) and complete phenotypes (AMR, MVI, TCMR and mixed AMR-TCMR). Rejection was less frequent in the European validation cohort, consistent with its prospective, unselected design and inclusion of 77.9% protocol biopsies (**Table 1**). The US cohort was enriched for TCMR (38.0%) and AMR (28.2%) and contained fewer intermediate phenotypes. Individual Banff lesion score distributions of the three cohorts are shown in **Supplementary Table S1**.

Across Banff 2022 phenotypes in the derivation cohort, the histological AMR/MVI indices[7] were highest in mixed rejection and progressively lower in AMR, DSAnegC4dneg MVI, probable AMR, and no rejection. The histological TCMR/TI indices[7] were highest in TCMR, intermediate in borderline TCMR and isolated v-lesions, and lowest in no rejection; PyVAN showed intermediate histological TCMR/TI indices comparable to isolated v-lesions (**Supplementary Table S2**). The histological activity index[7] was highest

in mixed rejection, whereas the histological chronicity indices correlated with time post-transplantation (r=0.49, p<0.001).

Development of the continuous molecular indices

To develop four continuous molecular indices reflecting antibody-mediated microvascular injury (AMR/MVI), T cell–mediated tubulointerstitial inflammation (TCMR/TI), inflammatory activity and chronicity, we applied a supervised modeling framework where the continuous histologic indices, rather than categorical Banff diagnoses, were used as ground truth outcomes. Candidate transcripts were identified using an ensemble feature-selection strategy in the derivation cohort, and parsimonious linear regression models were trained to generate continuous molecular scores. Model complexity was reduced through cross-validation–guided backward selection, and final models were chosen using the Bayesian Information Criterion to favor sparse yet informative transcript sets.

Distribution of the continuous molecular indices

In the derivation cohort, the four continuous histological indices showed consistently right-skewed distributions, with most biopsies having low scores and fewer biopsies showing higher values (**Fig. 1a**). Unlike histological lesion scores, which are semi-continuous and influenced by diagnostic thresholds, the molecular indices spanned a broad and continuous range, consistent with increasing injury severity.

Joint plots of AMR/MVI and TCMR/TI further illustrated differences between histological and molecular representations of rejection (**Fig. 1b**). Histological indices clustered along the axes, reflecting the constraints of categorical scoring. In contrast, molecular indices were distributed more evenly across the joint space, with many biopsies showing elevation of both AMR/MVI and TCMR/TI signals. Biopsies with high molecular AMR/MVI scores and minimal TCMR/TI scores, or the reverse, were uncommon, indicating that molecular injury signals often occur together rather than as isolated processes.

Across the derivation cohort, molecular indices showed moderate to strong correlations with their corresponding continuous histological indices[7], with Spearman correlation coefficients of r=0.59 for AMR/MVI, r=0.68 for TCMR/TI, r=0.63 for inflammatory activity and r=0.74 for chronicity (all p<0.001) (**Fig. 1c**). Similar correlations were observed in the US validation cohort, while correlations were weaker in the prospective European validation cohort, likely due to its narrower range of phenotypes and predominance of protocol biopsies (**Supplementary Figs. S2–S3**).

To evaluate whether the molecular indices capture continuous inflammatory biology, we examined their relationships with the histological spectrum of kidney transplant biopsies. In the derivation cohort (**Supplementary Figs. S4–S7**), the molecular AMR/MVI index increased stepwise with microvascular inflammation severity, including glomerulitis (g) and peritubular capillaritis (ptc), with weaker associations for etiology-linked lesions such as C4d deposition. The molecular TCMR/TI index rose with increasing interstitial inflammation (i), tubulitis (t) and total inflammation (ti), while showing limited association with chronic injury. The molecular activity index reflected acute inflammatory lesions across

both antibody- and T cell–mediated compartments, whereas the chronicity index primarily tracked chronic damage features, including interstitial fibrosis (ci), tubular atrophy (ct), transplant glomerulopathy (cg) and vascular fibrous intimal thickening (cv). Index distributions were broadly comparable between protocol and indication biopsies, although indication biopsies showed modestly higher index values (**Supplementary Fig. S8**).

Relationship between the continuous molecular indices and Banff-defined rejection

Across Banff diagnostic categories, the continuous molecular indices differed importantly, although there was substantial overlap of the indices between the histologic categories (**Fig. 2a**). Biopsies classified as AMR, DSAnegC4dneg MVI, and mixed rejection exhibited high molecular AMR/MVI index values, whereas molecular TCMR/TI indices increased across the spectrum of T cell–mediated rejection, with intermediate levels in borderline TCMR and isolated v-lesions. Intermediate Banff phenotypes, including probable AMR and borderline TCMR, showed marked molecular heterogeneity, but overall higher index values than biopsies classified as no rejection. Notably, a subset of biopsies without histological rejection displayed elevated molecular indices, indicating subclinical inflammatory activity not captured by categorical diagnosis. Exact molecular index values and pairwise comparisons between Banff phenotypes are shown in **Supplementary Table S2**.

In discrimination analyses (**Fig. 2b; Table 2**), molecular AMR/MVI and TCMR/TI indices robustly distinguished their corresponding Banff phenotypes, with AUCs of 0.94 (95% CI 0.90–0.97) and 0.90 (95% CI 0.86–0.95), respectively. Discrimination was modestly reduced for intermediate entities such as borderline TCMR, isolated v-lesions, and probable AMR, reflecting their biological overlap. The molecular activity index effectively discriminated abnormal from normal biopsies, achieving an AUC of 0.82 (95% CI 0.78–0.86), whereas the chronicity index showed limited discrimination, consistent with chronic injury reflecting cumulative rather than phenotype-specific processes. Raw and prevalence-adjusted AUPRCs are provided in **Supplementary Table S3**. Discriminative performance was similar in protocol and indication biopsies (**Supplementary Table S4**), with lower AUPRCs in protocol biopsies reflecting the lower prevalence of rejection in this subset. Calibration analyses showed good agreement between predicted and observed probabilities for the AMR/MVI and TCMR/TI indices, with low Brier scores (**Supplementary Fig. S9**).

Validation of the continuous molecular indices for diagnosing Banff-defined rejection

We next evaluated the performance of the four continuous molecular indices in two independent validation cohorts comprising a prospective European cohort (N=471) and a retrospective US cohort (N=1,190). Overall discriminative performance was consistent with that observed in the derivation cohort (**Fig. 2b; Table 2**), with the molecular AMR/MVI and TCMR/TI indices showing strong discrimination for their corresponding Banff-defined rejection phenotypes (all AUCs >0.86). In the US validation cohort, the chronicity index showed improved discrimination for abnormal versus normal biopsies and for mixed rejection compared with the derivation and prospective European cohorts, likely reflecting the higher prevalence of chronic active TCMR in this dataset (63.8% of active TCMR cases, versus 9.5% and 38.9%

in the derivation and European cohorts, respectively). Calibration of the AMR/MVI and TCMR/TI indices remained satisfactory in both validation cohorts, with good agreement between predicted and observed probabilities (**Supplementary Fig. S9**).

Reference ranges of the continuous molecular indices and relation to Banff categories

Although the continuous molecular indices were designed to capture injury across the full biological spectrum, clinically interpretable reference ranges are required to support diagnostic use and application as clinical trial endpoints. We therefore derived reference thresholds from the distribution of index values in biopsies classified as no rejection in the derivation cohort (N=226). Thresholds were defined at the 75th and 95th percentiles. For the molecular AMR/MVI and TCMR/TI indices, these corresponded to values of 0.76 and 1.43 for AMR/MVI, and 1.43 and 2.22 for TCMR/TI, respectively. Corresponding thresholds for the molecular activity and chronicity indices were 2.00 and 3.12 for activity, and 3.96 and 5.98 for chronicity. Continuous molecular indices were thus stratified as low (<P75), intermediate (P75–P95), or high (>P95).

Use of two thresholds for a continuous molecular marker align the biological continuum of allograft injury with the practical requirements of clinical decision-making. The lower threshold prioritizes sensitivity, identifying biopsies unlikely to show significant rejection, whereas the upper threshold prioritizes specificity, defining a high-confidence rule-in range (**Supplementary Table S5**). Values between these thresholds represent an intermediate zone in which molecular evidence of rejection is present but not definitive. This framework reduces binary decision-making, limits misclassification at diagnostic extremes, and enables probabilistic interpretation aligned with the continuous nature of allograft rejection.

Mapping the continuous molecular AMR/MVI and TCMR/TI indices onto Banff 2022 diagnostic categories revealed substantial within-category heterogeneity (**Fig. 3a**). Most AMR/MVI and TCMR cases were enriched in the highest corresponding molecular strata, while no-rejection biopsies mapped mainly to the lowest strata, with only a small minority showing unexpectedly elevated molecular activity (**Fig. 3b**). In contrast, diagnostically ambiguous phenotypes, such as probable AMR, borderline TCMR, and isolated v, were broadly distributed across molecular strata, indicating that molecular indices provide the greatest refinement in intermediate phenotypes. Because biopsies often exhibit overlapping rejection patterns in practice, we examined AMR/MVI and TCMR/TI jointly (**Supplementary Table S6**): most no-rejection biopsies clustered in the low–low stratum, whereas high–high biopsies were enriched for Banff-defined rejection, particularly mixed rejection.

Higher continuous molecular indices in non-rejection biopsies anticipate later rejection

All previous analyses associated the continuous molecular indices with the concomitant histology. Some cases showed increased molecular indices without meeting histologic criteria for Banff-defined disease. In initial biopsies classified as no rejection with a follow-up biopsy (N=377), higher molecular AMR/MVI indices in these initial biopsies were associated with subsequent development of any rejection

phenotype (β=0.69±0.25; p=0.005; OR 1.99 per unit), primarily future DSAnegC4dneg MVI or AMR. Molecular TCMR/TI indices in the initial biopsies showed only a borderline association with subsequent (borderline) TCMR (β=0.39±0.22; p=0.07; OR 1.48) (**Supplementary Fig. S10**).

Stratification of the molecular AMR/MVI and TCMR/TI indices yielded consistent results. High (≥P95) molecular AMR/MVI indices in the rejection-free initial biopsies were more frequent in kidneys that progressed to Banff 2022 AMR/MVI phenotypes (33.3% developed DSAnegC4dneg MVI or AMR) compared with biopsies having low (<P75) and intermediate (P75-P95) AMR/MVI indices (1.1% and 2.6%, respectively) (**Supplementary Table S7**).

In time-to-event analyses, higher molecular AMR/MVI indices were associated with shorter time to subsequent rejection (HR 1.55 per unit increase; 95% CI 1.02–2.36; p=0.04), and the high molecular AMR/MVI stratum conferred increased hazard compared to the low stratum (HR 2.39; 95% CI 1.13–5.06; p=0.02) (**Supplementary Table S8**). The continuous molecular TCMR/TI index was not significantly associated with time to subsequent rejection (HR 1.31; p=0.20), but the high TCMR/TI stratum showed an increased hazard versus the low stratum (HR 2.05; 95% CI 1.03–4.09; p=0.04). Importantly, while most biopsies classified as no rejection lacked detectable histologic injury, a minority exhibited isolated subthreshold inflammatory lesions, most commonly mild tubulitis (**Supplementary Fig. S11**).

Low molecular indices in patients with histological rejection

Although Banff-defined rejection phenotypes were generally enriched in the high molecular index strata, a substantial subset of biopsies meeting histologic criteria for AMR or TCMR mapped to low molecular index groups. To better understand this discordance, we examined the histologic features driving the Banff diagnoses in molecular-low rejection biopsies in comparison with molecular-high rejection biopsies (**Supplementary Table S9**).

Among Banff-defined AMR biopsies with low molecular AMR/MVI indices (N=87), microvascular inflammation was less pronounced than in AMR biopsies with high molecular indices (N=299). In particular, low-index AMR biopsies more frequently exhibited C4d positivity in the context of minimal microvascular inflammation (g+ptc ≤ 1). Similarly, Banff-defined TCMR biopsies with low molecular TCMR/TI indices (N=127) showed lower overall tubulointerstitial inflammatory activity compared with those in the high molecular group (N=298), and molecular-low TCMR biopsies also exhibited fewer v-lesions. Notably, the vast majority of histology-positive but molecular-low discordant cases originated from the US validation cohort (98.9% for AMR/MVI and 94.5% for TCMR/TI). Together, these findings demonstrate that histologic rejection phenotypes with low molecular indices represent biologically distinct states characterized by reduced inflammatory activity. Molecular indices therefore refine the interpretation of Banff-defined rejection by distinguishing highly active immune injury from more limited or incomplete phenotypes.

Continuous molecular indices are associated with graft failure

We next assessed associations between the continuous molecular indices and graft failure in the derivation and prospective validation cohorts (N=886 biopsies from 659 transplants; outcome data were unavailable for the US cohort). Higher molecular AMR/MVI, TCMR/TI, and activity indices were associated with lower eGFR at time of biopsy (r −0.25 to −0.33; all p<0.001), whereas the molecular chronicity index showed a weaker association (r −0.10; p=0.004) (**Supplementary Fig. S12**).

Restricted cubic spline analyses showed no evidence of non-linearity for any index (all nonlinear Wald p>0.1) (**Supplementary Fig. S13**). In multivariable Cox models using a randomly selected biopsy per transplant, all four indices were independently associated with graft failure, with adjusted hazard ratios per 1 standard deviation increase of 1.82 for AMR/MVI, 1.80 for TCMR/TI, 1.50 for activity and 1.25 for chronicity (all p<0.001; chronicity p=0.004) (**Fig. 4a; Supplementary Table S10**). Estimates were similar using first- or last-biopsy sampling strategies. Stratified analyses showed that graft failure risk among biopsies without rejection was most consistently associated with higher TCMR/TI and activity scores, whereas among biopsies with any rejection, associations were strongest for AMR/MVI, activity and chronicity; analyses were not further subdivided by individual Banff categories owing to limited event counts.

We next evaluated prognostic discrimination for 5-year graft failure (**Fig. 4b; Supplementary Table S11**). Across biopsy-selection strategies, continuous molecular indices consistently showed better discrimination than models based on Banff-defined rejection phenotypes alone. In the random-biopsy analysis, the molecular AMR/MVI index achieved an AUC of 0.67 (95% CI 0.57–0.74), compared with 0.61 (95% CI 0.55–0.67) for the Banff phenotype model, and performance was highest when molecular and histologic indices were combined (AUC 0.68; 95% CI 0.60–0.76). Similarly, the molecular TCMR/TI index outperformed the Banff TCMR phenotype model (AUC 0.64 vs 0.54) and the histologic index (AUC 0.57). Molecular activity indices also improved discrimination compared with Banff phenotype alone (AUC 0.66 vs 0.60), whereas chronicity indices showed more modest performance (AUC 0.60), comparable to Banff phenotype and slightly lower than histologic chronicity scores. Together, these findings indicate that continuous molecular indices refine prognostic risk stratification beyond categorical Banff phenotypes and provide complementary information to continuous histologic indices for predicting graft failure.

Discussion

Using transcriptomic profiling of formalin-fixed kidney allograft biopsies, we developed and externally validated four continuous molecular indices that quantify antibody-mediated microvascular injury, T cell–mediated tubulointerstitial inflammation, inflammatory activity and chronic injury across the full spectrum of allograft rejection. By training these models against continuous histologic indices rather than categorical Banff diagnoses, we preserved quantitative information on injury severity and generated molecular scores that are interpretable across their entire range. The resulting indices performed consistently across independent European and US cohorts, captured clinically relevant heterogeneity within Banff phenotypes, and provided prognostic information for subsequent rejection and graft failure.

These findings support a conceptual shift in kidney transplant pathology: rejection is best understood as a multidimensional continuum of immune-mediated injury and evolving chronic damage, rather than a binary state defined by discrete diagnostic thresholds. Although Banff categories remain essential for standardized reporting and clinical trial definitions, they inevitably compress spatially heterogeneous and dynamic biological processes into simplified labels. The continuous molecular indices presented here map rejection activity onto orthogonal dimensions aligned with compartment-specific injury patterns, while revealing substantial overlap across categories and marked within-category variation, particularly in intermediate entities such as borderline TCMR, isolated v lesions, probable AMR and mixed rejection.

Importantly, the indices were most informative in these diagnostic “gray zones,” where small histologic differences can shift categorical assignment, but underlying molecular injury varies widely. Percentile-based stratification redistributed intermediate phenotypes across low, intermediate and high molecular strata, suggesting that continuous molecular readouts may provide clinically meaningful refinement when histology alone is ambiguous. Conversely, biopsies with unequivocal normal histology or clear-cut rejection more often exhibited concordant molecular patterns, indicating that the incremental interpretive value of transcriptomics depends on the clinical and histologic context.[18-22]

A key challenge for biopsy-based molecular diagnostics is molecular–histologic discordance, particularly elevated molecular activity in biopsies classified as no rejection. Despite overall alignment, molecular and histologic indices were only moderately correlated, indicating that they capture overlapping yet distinct dimensions of intra-graft injury. While histological assessment and Banff lesion scoring describe the spatial distribution of injury, they do not capture the underlying molecular changes or the composition of infiltrating cell populations. Conversely, bulk transcriptomic analyses provide broad insight into molecular alterations but lack spatial resolution. Also, higher AMR/MVI scores in histology-negative biopsies anticipated subsequent antibody-mediated phenotypes and were associated with shorter time to rejection. These observations align with prior work[23] showing that subthreshold molecular rejection activity is common and clinically informative, supporting the view that transcriptomic profiling can detect immune activation before it crosses conventional diagnostic thresholds. Such findings suggest that molecular indices may complement histology by identifying early or evolving alloimmune injury rather than simply reproducing categorical diagnoses.

Beyond diagnosis, we demonstrated that molecular profiling provides clinically relevant prognostic information. To date, such associations with outcome were most described for MMDx-derived AMR classifiers[24], less consistent for MMDx TCMR[25], and largely absent for B-HOT–based classifiers[16,26,27] Here, all four continuous molecular indices were independently associated with impaired graft function and increased the risk of five-year graft failure, even after adjustment for clinical covariates. Moreover, molecular indices consistently improved discrimination compared with Banff phenotypes and continuous histological indices, supporting their potential role as quantitative prognostic markers and as candidate endpoints for clinical trials.

A major strength of this work is the portability and interpretability of the modeling framework. The indices rely on standardized B-HOT profiling from FFPE tissue with sparse transcript sets including housekeeping gene and an interpretable algorithm, enabling application to new biopsies without the need for cohort-specific retraining. This approach differs from transcriptome-wide categorical classifiers[15,26] and is designed to support reproducible quantitative risk estimation across centers and over time, including in mixed or overlapping rejection phenotypes commonly encountered in clinical practice.

Several limitations should be acknowledged. First, the indices were trained against histologic composite scores which, although more granular than categorical Banff diagnoses, remain imperfect surrogates of underlying biology and are subject to sampling error and interobserver variability. Second, outcome data were unavailable for the US validation cohort, limiting prognostic generalizability in that dataset. Third, treatment adaptations were not captured and may have influenced observed associations. Finally, molecular indices are not rejection-specific and may reflect shared immune activation pathways in non-alloimmune processes such as viral infection, drug toxicity and recurrent or de novo disease[28,29], underscoring the need to interpret molecular scores in conjunction with histology and clinical context.

In summary, continuous molecular indices derived from biopsy-based transcriptomics provide a quantitative, multidimensional framework for kidney allograft biopsy interpretation. By capturing rejection severity along a continuum, refining risk stratification in diagnostically ambiguous cases, identifying subclinical immune activity and improving prediction of graft failure, these indices complement the Banff Classification and advance precision transplant medicine. Prospective interventional studies are now required to determine whether integrating these molecular indices into clinical decision-making improves long-term graft outcomes.

Methods

Study population and data collection

The derivation cohort consisted of retrospectively selected adult kidney allograft biopsies performed at University Hospitals Leuven between October 2011 and December 2021 (ClinicalTrials.gov ID NCT06505200). Clinical and laboratory data were prospectively recorded in the electronic health record. Inclusion required sufficient biopsy tissue and representation of diverse histologic phenotypes; combined or sequential non-renal transplants were excluded. All included cases had negative complement-dependent cytotoxicity crossmatches. Biopsies were obtained at protocol time points (0, 3, 12, and 24 months post-transplantation) and when clinically indicated. Data collection was completed on June 11, 2022. Validation was performed in two independent cohorts. The first validation cohort comprised a prospective, unselected series of all kidney allograft biopsies performed at University Hospitals Leuven between March 2022 and March 2024 (data lock May 6, 2025). The US validation cohort consisted of the multicenter dataset reported by Zhang et al.,[16] including biopsies reviewed by Arkana Laboratories between January 2012 and November 2021, supplemented with selected cases of

rare diagnoses dating back to 2008. The study was approved by the Ethics Committee of University Hospitals Leuven (S63996). Written informed consent was obtained from all participants, and all procedures were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

Biopsy and HLA-DSA status evaluation

All biopsies from the derivation and prospective validation cohorts were prospectively documented and scored by experienced transplant pathologists using semiquantitative Banff criteria and subsequently classified according to the Banff 2022 classification. HLA donor-specific antibodies (HLA-DSA) were assessed according to local center protocols. At the University Hospitals Leuven, testing was performed pre- and post-transplantation, and at the time of biopsy using a Luminex single-antigen bead assay (HILA, Red Cross Flanders; Immucor®) targeting HLA-A, -B, -C, -DR, -DQ, and -DP loci.[30] Biopsies were classified according to the Banff 2022 classification (https://banfffoundation.org/central-repository-for-banff-classification-resources-3/): no rejection, probable antibody-mediated rejection (AMR), donor-specific antibody–negative C4d-negative microvascular inflammation (DSAnegC4dneg MVI), AMR (acute, chronic active, or chronic), isolated v-lesion (v > 0 with t0-i0 to t3-i0 or t0-i1–3),[31] borderline T cell–mediated rejection (TCMR), TCMR, or polyomavirus-associated nephropathy (PyVAN). “Mixed rejection” denoted concurrent AMR or DSAnegC4dneg MVI with TCMR. Intimal arteritis occurring within AMR/MVI phenotypes was not classified as mixed rejection unless additional criteria for (borderline) TCMR were fulfilled.[31] Histopathologic assessment was performed as part of routine clinical practice before molecular analyses, and pathologists had no access to molecular gene-expression results.

Calculation of continuous histological indices

Continuous histologic indices for AMR/microvascular inflammation (AMR/MVI) and TCMR/tubulointerstitial inflammation (TCMR/TI) were calculated from Banff lesion scores using previously established methods, as described by Vaulet et al.[4,5,7] Histological activity and chronicity indices were computed according to the Banff 2024 proposal.[32]

RNA extraction and gene expression analysis by NanoString nCounter

Three 10-µm sections of formalin-fixed paraffin-embedded (FFPE) tissue per biopsy were processed for RNA extraction using the Quick-RNA FFPE Kit (Zymo Research) and stored at 4–6 °C until analysis. RNA concentration was quantified using the Qubit RNA HS Assay Kit on a Qubit Flex Fluorometer (Invitrogen). Gene expression profiling was performed using the NanoString nCounter® FLEX system with the Banff Human Organ Transplant (B-HOT) gene panel. Hybridization was carried out using 150 ng of total RNA (8 µL input); when RNA yield was lower, the entire sample was used and vacuum-concentrated. Molecular analysts were blinded to histopathologic diagnoses and Banff lesion scores during RNA processing and gene-expression analysis.

Gene expression data processing

Samples yielding <75 ng total RNA after concentration were excluded according to predefined quality-control criteria. Batch calibration at the cartridge level was performed using panel standards following the NanoString procedure,[33] except for the Arkana dataset, for which panel standards were unavailable. Data normalization was conducted in two sequential steps: (1) positive control normalization using the four highest-concentration probes, and (2) housekeeping gene normalization ("CodeSet content normalization"), analogous to the nSolver pipeline.[34] Among the twelve B-HOT candidate housekeeping genes, the three most stable were identified using the geNorm genorm[35] in the derivation cohort, and their geometric mean was used to generate sample-specific scaling factors. This predefined, transparent normalization strategy, based on a restricted set of stable transcripts, enables reproducible application to new samples without cohort-specific refitting. Derivation and validation cohorts were processed independently to prevent data leakage. Expression values were $\log_2(X + 1)$–transformed and standardized using the mean and standard deviation of the derivation cohort. Uniform Manifold Approximation and Projection (UMAP) inspection revealed no major batch effects; therefore, no additional batch correction was applied.

Supervised feature selection and model development

Four continuous molecular indices—antibody-mediated microvascular injury (AMR/MVI), T cell–mediated tubulointerstitial inflammation (TCMR/TI), inflammatory activity and chronicity—were developed using a supervised feature-selection and modeling approach where continuous histologic indices, rather than categorical Banff diagnostic labels, were used as ground truth. Ensemble feature selection was performed in the derivation cohort to identify transcripts associated with each histologic index. Five complementary approaches were applied, including one univariate method (Spearman rank correlation) and four multivariate methods: Boruta[36] implemented with gradient-boosted decision trees, Stability Selection[37] using randomized LASSO, and Recursive Feature Elimination[38] applied to Partial Least Squares regression and Support Vector Regression. Each method initially selected the top 40 candidate transcripts, and transcripts retained by at least two approaches were carried forward for model training. For each molecular index, linear regression models were trained using the selected transcript sets, prioritizing interpretability, stability and portability. Model complexity was reduced using sequential backward feature elimination guided by mean squared error within five-fold cross-validation to mitigate overfitting. Final models were selected using the Bayesian Information Criterion (BIC), favoring sparse yet informative transcript combinations. Repeated biopsies were treated as independent observations during model development.

Outcome

The primary objective was to evaluate discrimination of the four molecular indices for Banff 2022 diagnostic categories in the prospective and US validation cohorts. Secondary objectives included: (i) comparison of index distributions across Banff phenotypes; (ii) associations with individual Banff lesion scores and continuous histologic indices; (iii) determine whether molecular activity in biopsies classified as no rejection predicted subsequent biopsy phenotypes; and (iv) assess associations with graft failure

and the incremental prognostic value of the molecular indices beyond Banff phenotype and histology-based models.

Statistical methods

The study followed STARD reporting guidelines (**Supplementary Table S12**). Given the retrospective design with external validation, no formal sample-size calculation was performed; all eligible biopsies within the study periods were included. Analyses were conducted using complete cases for variables required to assign Banff lesion scores and diagnostic phenotypes.

Diagnostic performance of each molecular index for prespecified Banff 2022 rejection categories was evaluated in the validation cohorts using the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (raw and prevalence-adjusted) to account for class imbalance. Calibration was assessed within each cohort using logistic regression with intercept recalibration to cohort-specific outcome prevalence, as described by Van Calster et al.[39] Because molecular index distributions were right-skewed with long upper tails, percentile-based molecular strata were prespecified for interpretability. Thresholds were derived from biopsies classified as no rejection in the derivation cohort (N=226) and applied unchanged across cohorts, defining low (<P75), intermediate (P75–P95) and high (>P95) strata.

Associations between molecular indices, histopathology and kidney function were assessed using Spearman rank correlation coefficients (ρ) with individual Banff lesion scores, continuous histologic indices and eGFR at time of the biopsy. Graft failure was defined as return to dialysis or retransplantation. Patients were censored at last follow-up or death. Associations with graft failure were evaluated using Cox proportional-hazards models, with hazard ratios reported per 1 standard deviation increase in each index and adjusted for time post-transplantation at biopsy, donor age, recipient age and recipient sex. Linearity of associations on the log-hazard scale was assessed using restricted cubic spline Cox models, and the nonlinear component was tested using Wald. Prognostic discrimination was quantified using 5-year time-dependent AUC. To address repeated biopsies per transplant, survival analyses were repeated using three prespecified sampling strategies (first, random, or last biopsy per transplant).

All tests were two-sided and $p<0.05$ was considered statistically significant.

Data pre-processing, feature selection and model derivation was performed in Python 3.9.10 with the following packages: genorm, boruta 0.3, sklearn 1.2.2, pandas 2.0.3, matplotlib 3.7.5, numpy 1.26.4, xgboost 2.1.3, statsmodels 0.13.5.

Analyses were performed in R (v4.5.0) using ggplot2[40], pROC[41], PRROC[42], survival[43], rms[44] and riskRegression[45] packages.

Data preprocessing, feature selection, and model development were performed in Python (v3.9.10) using scikit-learn[46], Boruta[36], and XGBoost[47]. Statistical analyses were conducted in R (v4.5.0) using the survival[43], rms[44], pROC[41], and riskRegression[45] packages. All tests were two-sided, and $p<0.05$ was considered statistically significant.

Declarations

Author contributions

KW, TV, and MN designed the study. KW, TV, HZ, AP, JL, PK, DK, TV, AV, MC, CL, CC, RW and MN were involved in data collection and curation, including histopathologic assessment of kidney allograft biopsies. NanoString gene-expression analyses were performed by HZ. TV designed and implemented the analytical pipeline for the molecular indices, including feature selection, model training, and validation. KW performed all other analyses with input from MC, AV and MN. KW and MN wrote the manuscript, which was revised by all co-authors.

Acknowledgements

We thank the clinicians, surgeons, nursing staff, and patients at University Hospitals Leuven, as well as the laboratory technicians of the Nephrology and Renal Transplantation Research Group (KU Leuven), for their essential contributions. We also acknowledge the centers participating in the Leuven Collaborative Group for Nephrology (LSGN). We are grateful to Arkana Laboratories for their collaboration and data sharing, which significantly contributed to this work.

Funding

This work was supported by grants from The Research Foundation Flanders (FWO) (project grant G038024N) and by the KU Leuven Research Council (project grant C2M/24/057), and by a research collaboration between KU Leuven and CareDx. KW and AP hold a FWO fellowship grant (11P1524N and 1S93023N respectively). MN is a senior clinical investigator of FWO (1842919N).

Disclosure

HZ and RW are salaried employees of CareDx. CL and CC are salaried employees of Arkana Laboratories. KU Leuven applied for a patent related to the content of this manuscript.

References

1. Roufosse C, Simmonds N, Clahsen-van Groningen M, et al. A 2018 Reference Guide to the Banff Classification of Renal Allograft Pathology. *Transplantation*. Nov 2018;102(11):1795-1814. doi:10.1097/tp.0000000000002366

2. Naesens M, Roufosse C, Haas M, et al. The Banff 2022 Kidney Meeting Report: Reappraisal of microvascular inflammation and the role of biopsy-based transcript diagnostics. *Am J Transplant*. Oct 28 2023;doi:10.1016/j.ajt.2023.10.016
3. Vaulet T, Callemeyn J, Lamarthée B, et al. The Clinical Relevance of the Infiltrating Immune Cell Composition in Kidney Transplant Rejection. *J Am Soc Nephrol*. Jul 1 2024;35(7):886-900. doi:10.1681/asn.0000000000000350
4. Vaulet T, Divard G, Thaunat O, et al. Data-driven Derivation and Validation of Novel Phenotypes for Acute Kidney Transplant Rejection using Semi-supervised Clustering. *J Am Soc Nephrol*. May 3 2021;32(5):1084-1096. doi:10.1681/asn.2020101418
5. Vaulet T, Divard G, Thaunat O, et al. Data-Driven Chronic Allograft Phenotypes: A Novel and Validated Complement for Histologic Assessment of Kidney Transplant Biopsies. *J Am Soc Nephrol*. Nov 2022;33(11):2026-2039. doi:10.1681/asn.2022030290
6. Sikosana MLN, Reeve J, Madill-Thomsen KS, Halloran PF. Using Regression Equations to Enhance Interpretation of Histology Lesions of Kidney Transplant Rejection. *Transplantation*. Feb 1 2024;108(2):445-454. doi:10.1097/tp.0000000000004783
7. Vaulet T, Koshy P, Wellekens K, et al. Continuous indices to assess the phenotypic spectrum of kidney transplant rejection. *Nat Commun*. Nov 26 2025;16(1):10417. doi:10.1038/s41467-025-65153-9
8. 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*. Sep 2020;20(9):2305-2317. doi:10.1111/ajt.16059
9. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. Mar 2008;26(3):317-25. doi:10.1038/nbt1385
10. 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*. Mar 2016;30(3):295-305. doi:10.1111/ctr.12689
11. Madill-Thomsen KS, Halloran PF. Precision diagnostics in transplanted organs using microarray-assessed gene expression: concepts and technical methods of the Molecular Microscope® Diagnostic System (MMDx). *Clin Sci (Lond)*. Jun 5 2024;138(11):663-685. doi:10.1042/cs20220530
12. Varol H, Ernst A, Cristoferi I, et al. Feasibility and Potential of Transcriptomic Analysis Using the NanoString nCounter Technology to Aid the Classification of Rejection in Kidney Transplant Biopsies. *Transplantation*. Apr 1 2023;107(4):903-912. doi:10.1097/tp.0000000000004372
13. Giarraputo A, Goutaudier V, Robin B, et al. Relevance of the Banff Human Organ Transplant Consensus Gene Panel for Detecting Antibody and T-Cell-Mediated Rejection of Kidney Allografts. *Kidney Int Rep*. Jul 2024;9(7):2290-2294. doi:10.1016/j.ekir.2024.04.054
14. Smith RN, Rosales IA, Tomaszewski KT, et al. Utility of Banff Human Organ Transplant Gene Panel in Human Kidney Transplant Biopsies. *Transplantation*. May 1 2023;107(5):1188-1199. doi:10.1097/tp.0000000000004389

15. van Baardwijk M, Cristoferi I, Ju J, et al. A Decentralized Kidney Transplant Biopsy Classifier for Transplant Rejection Developed Using Genes of the Banff-Human Organ Transplant Panel. *Front Immunol*. 2022;13:841519. doi:10.3389/fimmu.2022.841519
16. Zhang H, Haun RS, Collin F, et al. Development and Validation of a Multiclass Model Defining Molecular Archetypes of Kidney Transplant Rejection: A Large Cohort Study of the Banff Human Organ Transplant Gene Expression Panel. *Lab Invest*. Mar 2024;104(3):100304. doi:10.1016/j.labinv.2023.100304
17. Halloran PF, Madill-Thomsen KS, Reeve J. The Molecular Phenotype of Kidney Transplants: Insights From the MMDx Project. *Transplantation*. Jan 1 2024;108(1):45-71. doi:10.1097/tp.0000000000004624
18. Weidmann L, Harmacek D, Castrezana Lopez K, et al. Limitations of biopsy-based transcript diagnostics to detect T-cell-mediated allograft rejection. *Nephrol Dial Transplant*. Feb 4 2025;40(2):294-307. doi:10.1093/ndt/gfae147
19. van Baardwijk M, Wagenmakers A, van den Bosch TPP, et al. Borderline (suspicious) for T-cell-mediated rejection, the Banff classification's Achilles' heel. *Nephrol Dial Transplant*. Feb 4 2025;40(2):224-226. doi:10.1093/ndt/gfae192
20. Salazar ID, Merino López M, Chang J, Halloran PF. Reassessing the Significance of Intimal Arteritis in Kidney Transplant Biopsy Specimens. *J Am Soc Nephrol*. Dec 2015;26(12):3190-8. doi:10.1681/asn.2014111064
21. Reeve J, Sellarés J, Mengel M, et al. Molecular diagnosis of T cell-mediated rejection in human kidney transplant biopsies. *Am J Transplant*. Mar 2013;13(3):645-55. doi:10.1111/ajt.12079
22. Buxeda A, Crespo M, Chamoun B, et al. Clinical and molecular spectrum of v-lesion. *Am J Transplant*. Jul 29 2024;doi:10.1016/j.ajt.2024.07.025
23. Halloran PF, Madill-Thomsen KS, Böhmig G, et al. Subthreshold rejection activity in many kidney transplants currently classified as having no rejection. *Am J Transplant*. Jan 2025;25(1):72-87. doi:10.1016/j.ajt.2024.07.034
24. Halloran PF, Pereira AB, Chang J, et al. Microarray diagnosis of antibody-mediated rejection in kidney transplant biopsies: an international prospective study (INTERCOM). *Am J Transplant*. Nov 2013;13(11):2865-74. doi:10.1111/ajt.12465
25. Halloran PF, Pereira AB, Chang J, et al. Potential impact of microarray diagnosis of T cell-mediated rejection in kidney transplants: The INTERCOM study. *Am J Transplant*. Sep 2013;13(9):2352-63. doi:10.1111/ajt.12387
26. Zielinski D, Goutaudier V, Sablik M, et al. Molecular diagnosis of kidney allograft rejection based on the Banff Human Organ Transplant gene panel: A multicenter international study. *Am J Transplant*. Aug 2025;25(8):1631-1642. doi:10.1016/j.ajt.2025.04.025
27. Beadle J, Papadaki A, Toulza F, et al. Application of the Banff Human Organ Transplant Panel to kidney transplant biopsies with features suspicious for antibody-mediated rejection. *Kidney International*. 2023;104(3):526-541. doi:10.1016/j.kint.2023.04.015

28. Roufosse C, Naesens M, Haas M, et al. The Banff 2022 Kidney Meeting Work Plan: Data-driven refinement of the Banff Classification for renal allografts. *Am J Transplant*. Mar 2024;24(3):350-361. doi:10.1016/j.ajt.2023.10.031
29. Rho E, Weidmann L, Korach R, et al. Additional Diagnoses Other Than Rejection in the Kidney Allograft Biopsy: Pitfalls for Biopsy-based Transcript Diagnostics. *Transplant Direct*. Mar 2025;11(3):e1759. doi:10.1097/txd.0000000000001759
30. Senev A, Lerut E, Van Sandt V, et al. Specificity, strength, and evolution of pretransplant donor-specific HLA antibodies determine outcome after kidney transplantation. *Am J Transplant*. Nov 2019;19(11):3100-3113. doi:10.1111/ajt.15414
31. Wellekens K, Coemans M, Koshy P, et al. The impact of the Banff v-lesion on rejection classification and outcomes: Insights from a multicenter study. *Am J Transplant*. May 5 2025;doi:10.1016/j.ajt.2025.04.023
32. Naesens M RC, Cornell L, et al. The Banff 2024 Kidney Meeting Report: Rejection as a Spectrum of Phenotypes and Focus on Non-Rejection Glomerular Disease. *SSRN*. 2025;
33. Technologies N. TECH NOTE: Panel Standard and Calibration Sample Usage. https://nanostring.com/wp-content/uploads/TN_MK3415_Panel-Standard_r9.pdf
34. Technologies N. nSolver™ 4.0: Analysis Software User Manual. https://nanostring.com/wp-content/uploads/MAN-C0019-08_nSolver_4.0_Analysis_Software_User_Manual.pdf
35. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. Jun 18 2002;3:RESEARCH0034.
36. Kursa M, Rudnicki W. Feature Selection with the Boruta Package. *J Stat Softw*. 2010;36:1-13.
37. Meinshausen N, Bühlmann P. Stability selection: Stability Selection. *Stat Soc Ser B Stat Methodol* 2010;72:417–473
38. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach Learn* 2002;46:389–422.
39. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. Dec 16 2019;17(1):230. doi:10.1186/s12916-019-1466-7
40. Wickham H CW, Henry L, Pedersen T, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 2021.
41. Robin X, Turck N, Hainard A, et al. pROC' package: Display and Analyze ROC Curves. 2023.
42. Grau J, Keilwagen J. PRROC package: Precision-Recall and ROC Curves for Weighted and Unweighted Dat. 2025.
43. Therneau T. A package for survival analysis in R. 2024.
44. F H. rms: Regression Modeling Strategies. 2025.
45. Gerds T OJ, Blanche P, Mortensen R, et al. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks. 2025.

46. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
47. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785–794.

Tables

Table 1 | Descriptive summary of patient and biopsy characteristics.

Cohort characteristics

Recipients	**Derivation cohort**	**European validation**
	N=351 transplants	N=338 transplants
Age (years), mean ± SD	53.2 ± 12.9	55.7 ± 12.3
Missing, n (%)	1 (0.3%)	0 (0.0%)
Male, n (%)	230 (65.3%)	227 (67.2%)
Female, n (%)	121 (34.4%)	111 (32.8%)
Repeat transplantations, N (%)	66 (18.8%)	41 (12.1%)
Ethnicity		
White, n (%)	333 (94.6%)	320 (94.7%)
Asian, n (%)	4 (1.1%)	8 (2.4%)
Black, n (%)	13 (3.7%)	9 (2.7%)
Other, n (%)	1 (0.3%)	1 (0.3%)
Pre-transplant HLA-DSA, n (%)	43 (12.2%)	24 (7.1%)
De novo HLA-DSA at any time post-transplantation, n (%)	13 (3.7%)	15 (4.4%)
Immunosuppressive regimen		
TAC-MMF-CS, n (%)	336 (95.7%)	314 (92.9%)
Other, n (%)	15 (4.3%)	24 (7.1%)
Induction therapy, n (%)	186 (53.0%)	188 (55.6%)
Basiliximab, n (%)	184 (52.4%)	182 (53.8%)
Anti-thymocyte globulins, n (%)	2 (0.6%)	6 (1.8%)
Missing, n (%)	2 (0.6%)	2 (0.6%)
Follow-up time post-transplantation (years), median [IQR]	5.2 [2.6-7.6]	2.1 [1.4-3.1]
Delayed graft function, n (%)	63 (17.9%)	68 (20.1%)
Missing, n (%)	1 (0.3%)	2 (0.6%)
Donors		
Age (years), mean ± SD	49.2 ± 13.8	54.0 ± 12.8
Missing, n (%)	0 (0.0%)	1 (0.3%)

Male, n (%)	183 (52.0%)		177 (52.4%)
Female, n (%)	168 (47.7%)		160 (47.3%)
Missing, n (%)	0 (0.0%)		1 (0.3%)
Donor type			
Donation after brain death, n (%)	239 (68.1%)		173 (51.2%)
Donation after cardiac death, n (%)	77 (21.9%)		120 (35.5%)
Living donation, n (%)	34 (9.7%)		44 (13.0%)
Missing, n (%)	1 (0.3%)		1 (0.3%)
Cold ischemia time (h), mean ± SD	12.3 ± 5.8		9.4 ± 6.1
Missing, n (%)	1 (0.3%)		3 (0.9%)
Biopsy characteristics	**Derivation cohort**	**European validation cohort**	**US validation cohort**
	N=415 biopsies	N=471 biopsies	N=1190 biopsies
Days since transplantation, median [IQR]	97 [76-483]	365 [92-743]	
Missing, n (%)	0 (0.0%)	0 (0.0%)	1190 (100.0%)
Protocol biopsy, n (%)	261 (62.9%)	367 (77.9%)	
Indication biopsy, n (%)	154 (37.1%)	104 (22.1%)	
Missing, n (%)	0 (0.0%)	0 (0.0%)	1190 (100.0%)
HLA-DSA positivity at time of the biopsy, n (%)	53 (12.8%)	22 (4.7%)	168 (14.1%)
Banff phenotypes*			
No rejection, n (%)	226 (54.5%)	388 (82.4%)	232 (19.5%)
PyVAN, n (%)	14 (3.4%)	17 (3.6%)	61 (5.1%)
Isolated v, n (%)	26 (6.2%)	4 (0.8%)	1 (0.1%)
Borderline TCMR, n (%)	65 (15.7%)	26 (5.5%)	4 (0.3%)
TCMR, n (%)	21 (5.1%)	18 (3.8%)	453 (38.1%)
Probable AMR, n (%)	6 (1.4%)	0 (0.0%)	0 (0.0%)
DSAneg C4dneg MVI, n (%)	17 (4.1%)	6 (1.3%)	1 (0.1%)
AMR, n (%)	21 (5.1%)	8 (1.7%)	336 (28.2%)
Mixed rejection, n (%)	19 (4.6%)	4 (0.8%)	102 (8.6%)

(1) Student t-test (2) Chi-square test or Fisher’s exact test (3) Wilcoxon test

* Phenotypes were mutually exclusive. Biopsies with AMR plus borderline TCMR were classified as AMR (derivation cohort: N=0; EU validation cohort: N=12; US validation cohort: N=88). Biopsies with DSAnegC4dneg MVI plus borderline TCMR were classified as DSAnegC4dneg MVI (derivation cohort: N=0; EU validation cohort: N=9; US validation cohort: N=0). Probable AMR with borderline TCMR was classified as probable AMR (derivation cohort: N=2; none in validation cohorts), and probable AMR with TCMR was classified as TCMR (derivation cohort: N=2; EU validation cohort: N=1). Mixed rejection was defined as AMR or DSAnegC4dneg MVI co-occurring with TCMR, including 13 AMR+TCMR and 6 DSAnegC4dneg MVI+TCMR cases in the derivation cohort; 2 AMR+TCMR and 2 DSAnegC4dneg MVI+TCMR cases in the EU validation cohort; and 100 AMR+TCMR and 2 DSAnegC4dneg MVI+TCMR cases in the US validation cohort. In the EU validation cohort, PyVAN co-occurred with TCMR (N=2), DSAnegC4dneg MVI (N=3), and AMR+TCMR (N=1); no PyVAN co-occurrence was observed in the other cohorts.

Table 2 | Discrimination of molecular indices across cohorts. AUROC values with 95% confidence intervals are shown for AMR/MVI, TCMR/TI, activity and chronicity indices in the derivation and validation cohorts. Raw and prevalence-adjusted AUPRCs are reported in Supplementary Table S3.

Molecular index	Discrimination	AUC [95% CI] Derivation cohort N=415	AUC [95% CI] European validation cohort N=471	AUC [95% CI] US validation cohort N=1,192
AMR/MVI index	AMR vs. no AMR	0.93 [0.90-0.96]	0.94 [0.88-0.99]	0.86 [0.84-0.89]
	AMR/MVI vs. no AMR/MVI	0.94 [0.90-0.97]	0.94 [0.87-0.99]	0.86 [0.84-0.88]
	Probable AMR/AMR/MVI vs. no probable AMR/AMR/MVI	0.91 [0.87-0.95]	0.92 [0.86-0.98]	NA*
TCMR/TI index	TCMR vs No TCMR	0.90 [0.86-0.95]	0.89 [0.83-0.95]	0.86 [0.84-0.88]
	Isolated v/borderline TCMR/TCMR vs. no isolated v/borderline TCMR/TCMR	0.84 [0.80-0.88]	0.83 [0.77-0.88]	0.87 [0.85-0.89]
	PyVAN vs. no PyVAN	0.44 [0.32-0.57]	0.76 [0.64-0.88]	0.53 [0.46-0.60]
Activity index	No rejection vs. any abnormal biopsy	0.82 [0.78-0.86]	0.80 [0.74-0.85]	0.96 [0.94-0.97]
	No rejection vs. AMR/MVI/TCMR/mixed rejection	0.88 [0.83-0.91]	0.87 [0.82-0.93]	0.91 [0.88-0.93]
	No rejection vs. mixed rejection	0.93 [0.88-0.98]	0.99 [0.98-1.00]	0.82 [0.78-0.85]
Chronicity index	No rejection vs. any abnormal biopsy	0.53 [0.47-0.58]	0.55 [0.48-0.63]	0.73 [0.70-0.76]
	No rejection vs. AMR/MVI/TCMR/mixed rejection	0.59 [0.52-0.69]	0.62 [0.51-0.73]	0.67 [0.64-0.70]
	No rejection vs. mixed rejection	0.52 [0.38-0.67]	0.56 [0.15-0.97]	0.68 [0.63-0.74]

NA* Not applicable, as the probable AMR phenotype was not observed in validation cohort 2.

Abbreviations: AMR, antibody-mediated rejection; AUC, area under the ROC curve; AUPRC, area under the precision-recall curve; bTCMR, borderline TCMR; MVI, microvascular inflammation without DSA or C4d; pAMR, probable AMR; PyVAN, polyomavirus-associated nephropathy; TCMR, T cell-mediated rejection.

Figures

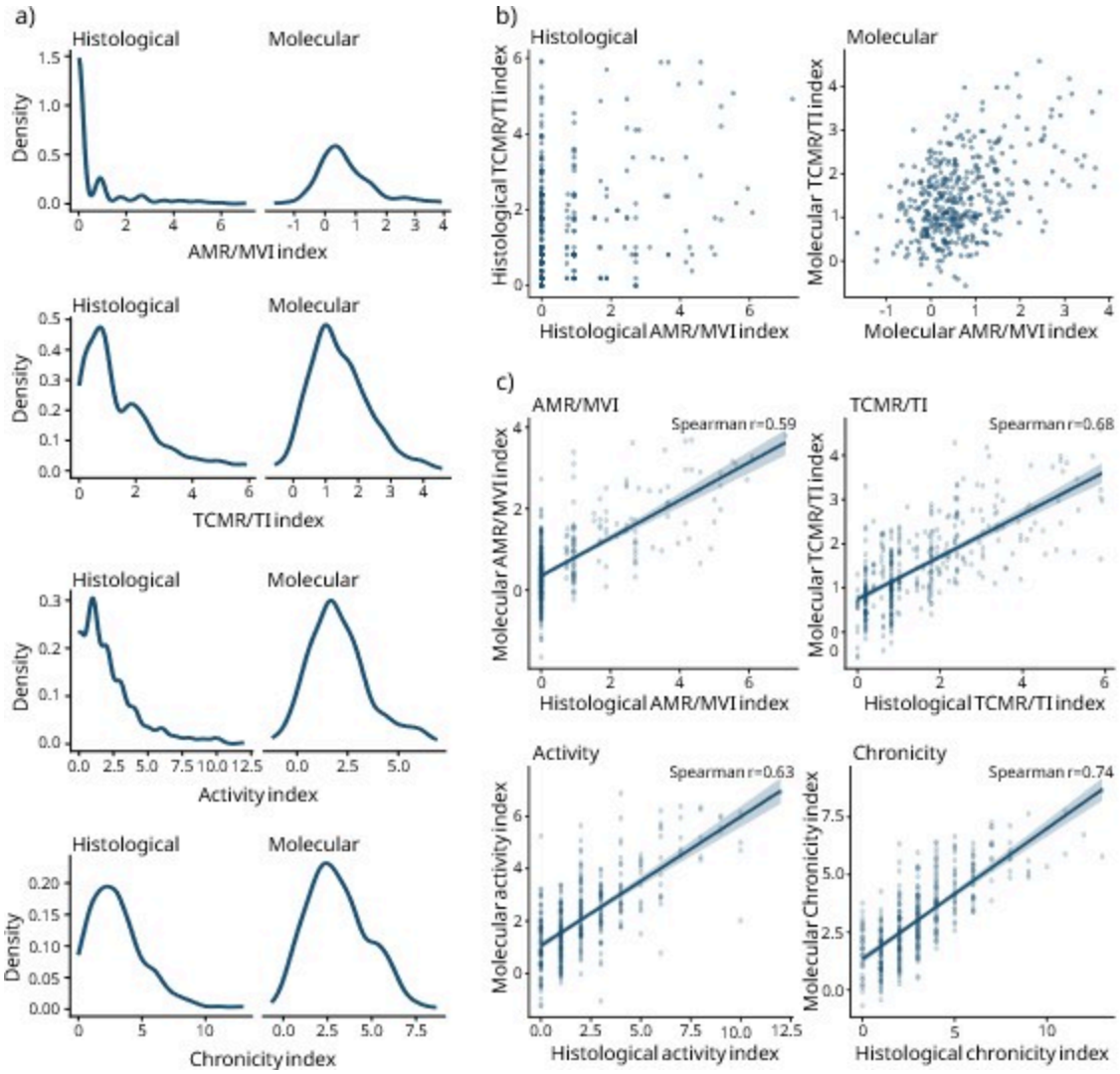

Figure 1

Molecular indices quantify continuous spectra of rejection in kidney allograft biopsies.

a, Distribution of continuous histological and molecular indices in the derivation cohort.
Density plots show the distributions of histological (left) and molecular (right) indices for antibody-mediated microvascular injury (AMR/MVI), T cell–mediated tubulointerstitial inflammation (TCMR/TI), inflammatory activity and chronicity in the derivation cohort (N=415).

b, Joint distribution of AMR/MVI and TCMR/TI indices.
Two-dimensional scatter plots depict the joint distribution of histological (left) and molecular (right) AMR/MVI and TCMR/TI indices in the derivation cohort (N=415).

c, Correlation between continuous histological and molecular indices.
Scatter plots show the relationship between each molecular index and its corresponding continuous histological index for AMR/MVI, TCMR/TI, inflammatory activity and chronicity in the derivation cohort (N=415). Points represent individual biopsies. Lines indicate linear fits with shaded 95% confidence intervals. Spearman correlation coefficients are shown in each panel.

Abbreviations: AMR, antibody-mediated rejection; MVI, microvascular inflammation; TCMR, T-cell mediated rejection; TI, tubulo-interstitial inflammation.

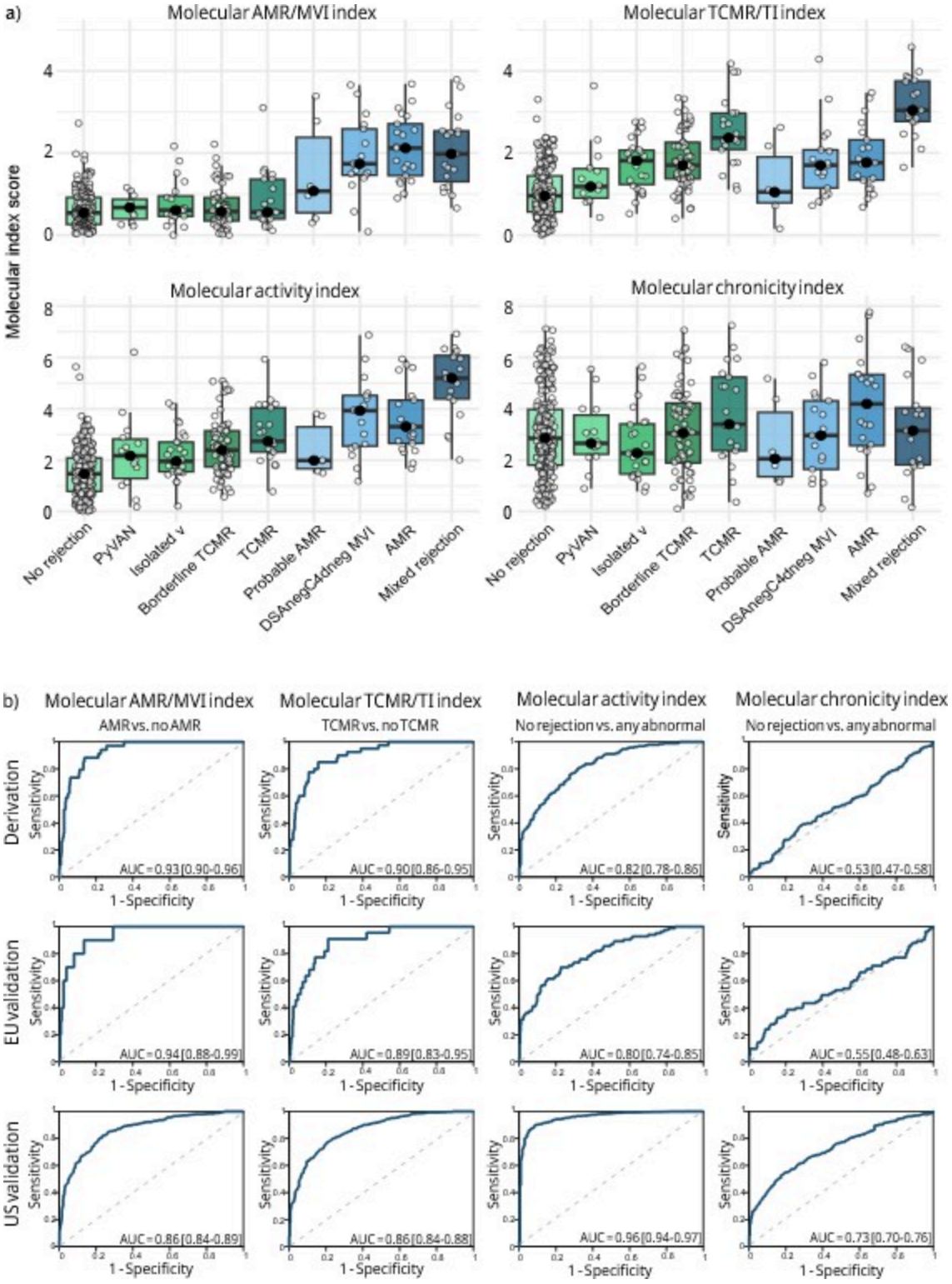

Figure 2

Relationship between molecular indices and Banff-defined rejection phenotypes across.

a, Distribution of molecular indices across Banff phenotypes in the derivation cohort (N=415). Boxplots show the distributions of the molecular AMR/MVI, TCMR/TI, activity and chronicity indices across Banff

2022 diagnostic phenotypes. Boxes indicate the interquartile range, center lines denote medians, and whiskers extend to 1.5× the interquartile range; individual points represent biopsies. Exact index values and pairwise comparisons are shown in Supplementary Table S2.

b, Discrimination of rejection phenotypes across cohorts.
Receiver-operating characteristic (ROC) curves for the molecular AMR/MVI, TCMR/TI, activity and chronicity indices in the derivation cohort (N=415), European validation cohort (N=471) and US validation cohort (N=1,190). AMR/MVI and TCMR/TI indices were evaluated for discrimination of their corresponding rejection phenotypes, whereas activity and chronicity indices were evaluated for discrimination of no rejection versus any abnormal biopsy. Area under the curve (AUC) values are shown in each panel; additional phenotype comparisons and prevalence-adjusted AUPRCs are provided in Table 2.

Abbreviations: AMR, antibody-mediated rejection; AUC, area under the curve; bTCMR, borderline TCMR; DSAnegC4dneg MVI, microvascular inflammation without DSA or C4d; pAMR, probable AMR; PyVAN, polyomavirus-associated nephropathy; ROC, receiver-operating characteristic; TCMR, T-cell mediated rejection.

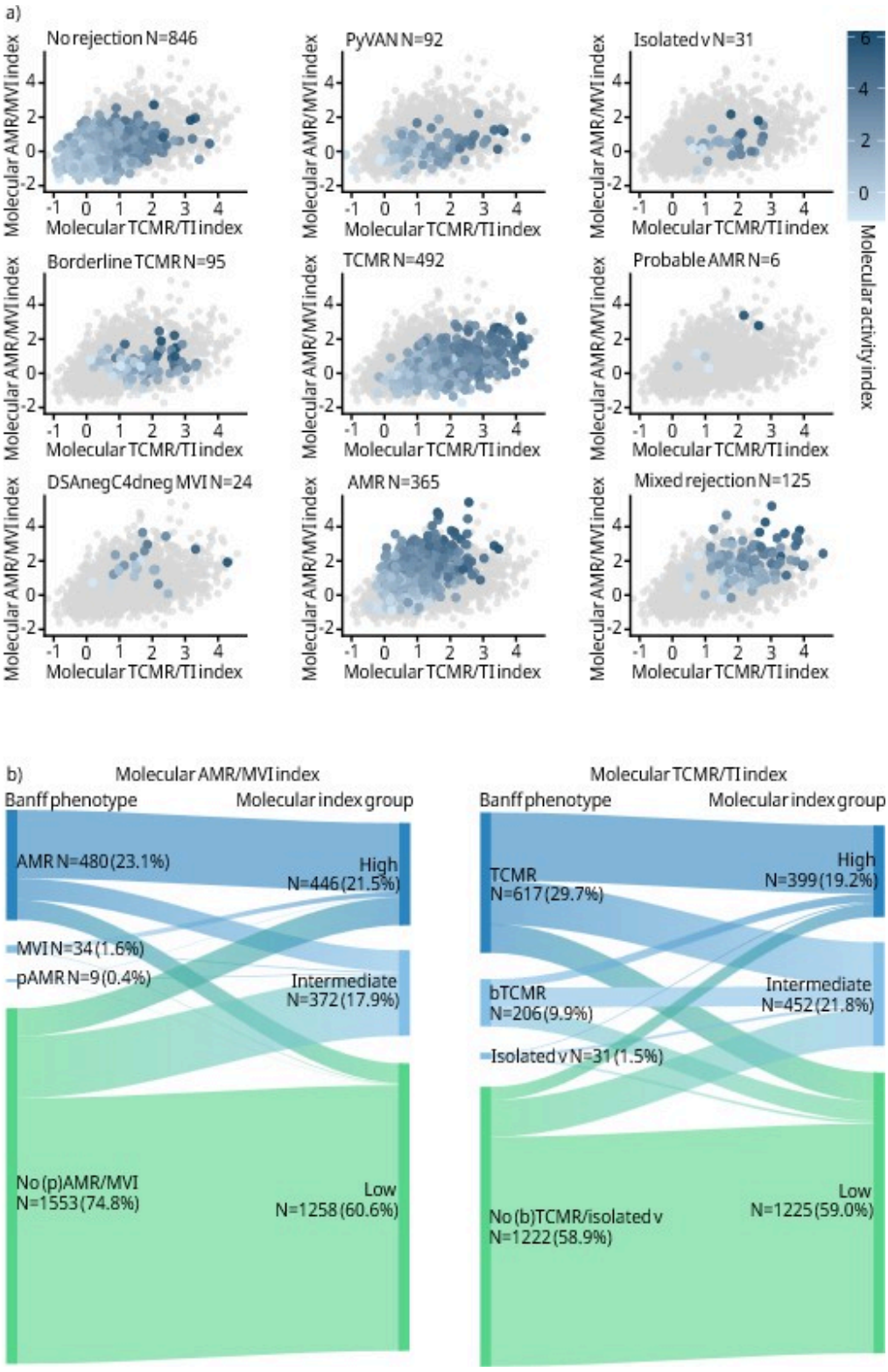

Figure 3

Molecular stratification of Banff rejection phenotypes.

a, Three-dimensional distribution of rejection phenotypes in all cohorts (N=2,076). Scatterplots show the distribution of Banff 2022 diagnostic categories within a three-dimensional molecular space defined by

the AMR/MVI index, the TCMR/TI index and the activity index across all biopsies from the three cohorts. Each point represents a single biopsy.

b, Mapping of Banff phenotypes to molecular index strata.
Alluvial plots show the distribution of Banff phenotypes across low, intermediate and high molecular index strata defined by the 75th and 95th percentiles of the molecular AMR/MVI index (left) and molecular TCMR/TI index (right).

For AMR/MVI, AMR mapped to high/intermediate/low groups in 299/480 (62.3%), 94/480 (19.6%) and 87/480 (18.1%) biopsies, respectively; DSAnegC4dneg MVI in 22/34 (64.7%), 6/34 (17.6%) and 6/34 (17.6%); and probable AMR in 2/9 (22.2%), 5/9 (55.6%) and 2/9 (22.2%). Biopsies without (probable) AMR/MVI were predominantly low-score (1,160/1,553; 74.7%), with 123/1,553 (7.9%) high-score. For TCMR/TI, TCMR mapped to high/intermediate/low groups in 298/617 (48.3%), 192/617 (31.1%) and 127/617 (20.6%) biopsies; borderline TCMR in 37/206 (18.0%), 86/206 (41.7%) and 85/206 (41.3%); and isolated v in 6/31 (19.4%), 13/31 (41.9%) and 12/31 (38.7%). Biopsies without (borderline) TCMR or isolated v were predominantly low-score (1,003/1,222; 82.1%), with 58/1,222 (4.7%) high-score.

Abbreviations: AMR, antibody-mediated rejection; bTCMR, borderline T-cell mediated rejection; DSAnegC4dneg MVI, microvascular inflammation without DSA or C4d; MVI, microvascular inflammation; pAMR, probable AMR; PyVAN, polyomavirus-associated nephropathy; TCMR, T-cell mediated rejection; TI, tubulointerstitial inflammation; v, intimal arteritis.

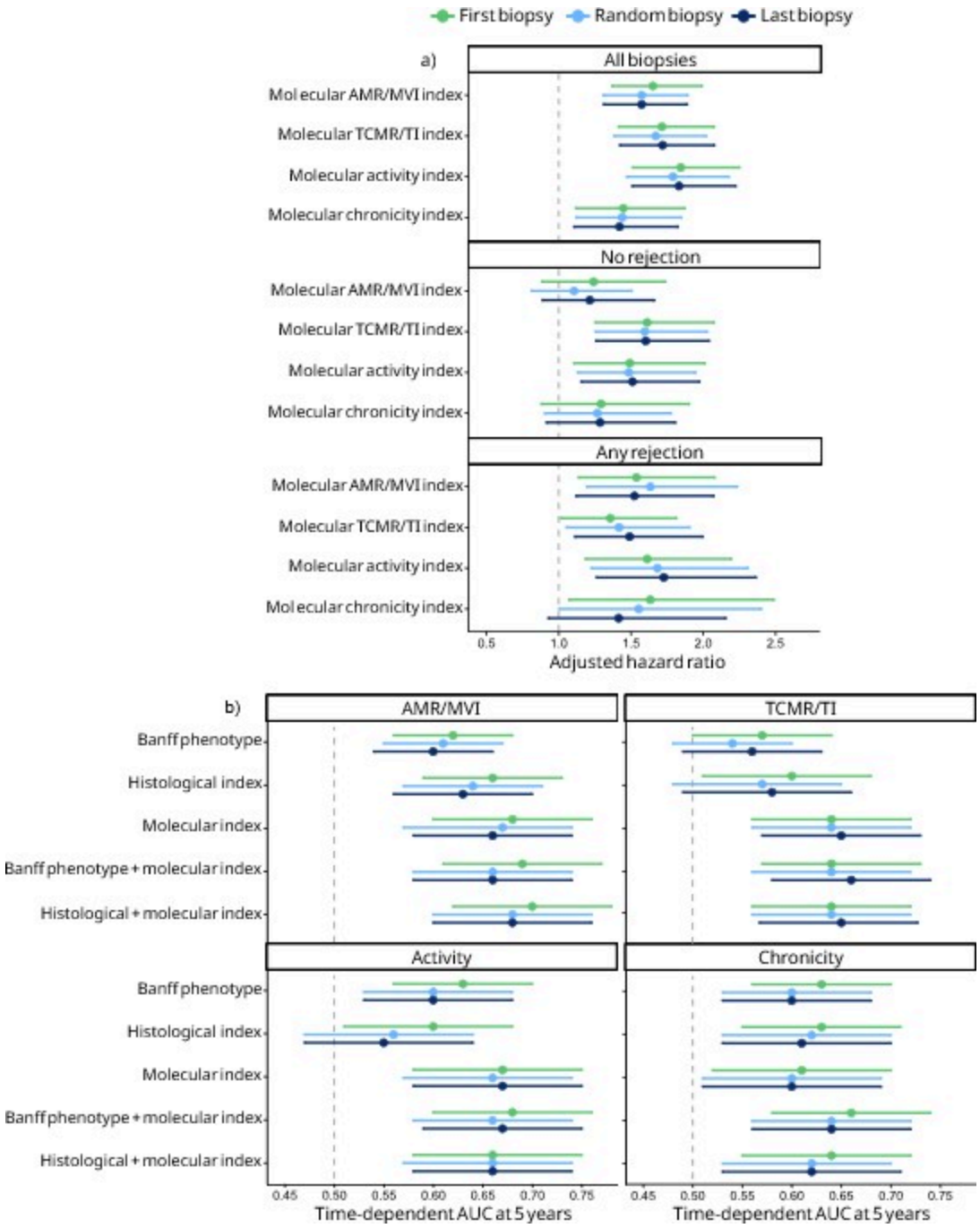

Figure 4

Molecular indices and graft failure in the combined derivation and European validation cohorts (N=886).

a, Association between molecular indices and five-year graft failure. Forest plots show hazard ratios (HRs) and 95% confidence intervals from multivariable Cox proportional-hazards models assessing graft failure within 5 years after biopsy, per 1 standard deviation increase in each molecular index. Results are presented for three biopsy-selection strategies per transplant (first available biopsy, randomly selected biopsy and last available biopsy) and stratified by phenotype (no rejection versus any rejection). Models were adjusted for time post-transplantation at biopsy, donor age, recipient age and recipient sex. Exact HR estimates are provided in Supplementary Table S10.

b, Discrimination of graft failure prediction models at five years. Forest plots show time-dependent AUC values at 5 years comparing molecular indices, Banff phenotypes, histological indices and combined models across biopsy-selection strategies. Exact AUC estimates are provided in Supplementary Table S11.

Abbreviations: AMR, antibody-mediated rejection; AUC, area under the curve; CI, confidence interval; eGFR, estimated glomerular filtration rate; HR, hazard ratio; MVI, microvascular inflammation; TCMR, T cell–mediated rejection; TI, tubulointerstitial inflammation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx