BASIC RESEARCH www.jasn.org

Data-Driven Chronic Allograft Phenotypes: A Novel and Validated Complement for Histologic Assessment of Kidney Transplant Biopsies

Thibaut Vaulet,[1] Gillian Divard,[2] Olivier Thaunat,[3,4] Priyanka Koshy,[5] Evelyne Lerut,[5] Aleksandar Senev,[6,7] Olivier Aubert,[2] Elisabet Van Loon,[6] Jasper Callemeyn,[6] Marie-Paule Emonds,[6,7] Amaryllis Van Craenenbroeck,[6,8] Katrien De Vusser,[6,8] Ben Sprangers,[6,8] Maud Rabeyrin,[9] Valérie Dubois,[10] Dirk Kuypers,[6,8] Maarten De Vos,[1,11] Alexandre Loupy,[2] Bart De Moor,[1] and Maarten Naesens[6,8]

Due to the number of contributing authors, the affiliations are listed at the end of this article.

ABSTRACT

Background No validated system currently exists to realistically characterize the chronic pathology of kidney transplants that represents the dynamic disease process and spectrum of disease severity. We sought to develop and validate a tool to describe chronicity and severity of renal allograft disease and integrate it with the evaluation of disease activity.

Methods The training cohort included 3549 kidney transplant biopsies from an observational cohort of 937 recipients. We reweighted the chronic histologic lesions according to their time-dependent association with graft failure, and performed consensus *k*-means clustering analysis. Total chronicity was calculated as the sum of the weighted chronic lesion scores, scaled to the unit interval.

Results We identified four chronic clusters associated with graft outcome, based on the proportion of ambiguous clustering. The two clusters with the worst survival outcome were determined by interstitial fibrosis and tubular atrophy (IFTA) and by transplant glomerulopathy. The chronic clusters partially overlapped with the existing Banff IFTA classification (adjusted Rand index, 0.35) and were distributed independently of the acute lesions. Total chronicity strongly associated with graft failure (hazard ratio [HR], 8.33; 95% confidence interval [CI], 5.94 to 10.88; $P<0.001$), independent of the total activity scores (HR, 5.01; 95% CI, 2.83 to 7.00; $P<0.001$). These results were validated on an external cohort of 4031 biopsies from 2054 kidney transplant recipients.

Conclusions The evaluation of total chronicity provides information on kidney transplant pathology that complements the estimation of disease activity from acute lesion scores. Use of the data-driven algorithm used in this study, called RejectClass, may provide a holistic and quantitative assessment of kidney transplant injury phenotypes and severity.

JASN 33: 2026–2039, 2022. doi: https://doi.org/10.1681/ASN.2022030290

Although kidney transplantation is considered the first-choice therapy for kidney failure and provides better outcomes than remaining on the wait list in dialysis,[1,2] graft failure remains a clinical problem.[3] The etiology of kidney transplant failure is complex and multifactorial.[4] Although some (acute) inflammatory lesions are observed mainly early after transplantation, others accumulate over time and reflect chronic injury processes, sometimes already ongoing

Received March 11, 2022. Accepted July 24, 2022.

Published online ahead of print. Publication date available at www.jasn.org.

See related editorial, "The ABCD of Kidney Allograft Pathology—The Beginning of the Beginning," on pages 1960–1963.

Correspondence: Dr. Maarten Naesens, Department of Microbiology, Immunology, and Transplantation, KU Leuven, University of Leuven, Leuven, Belgium. Email: maarten.naesens@kuleuven.be

in the donor before transplantation.[5,6] Late post-transplantation biopsies are commonly affected by nonspecific progressive fibrosis,[7,8] whereas early biopsies mostly contain inflammatory lesions related to allo-immune risk factors.[9] It results that early graft failures are dominated by T cell–mediated rejection (TCMR) and antibody-mediated rejection (ABMR), recurrent disease, and failure due to early perioperative problems, whereas late graft failures are more multifactorial, not only related to allo-immune processes, but also driven by nonspecific chronic injury such as graft quality and calcineurin inhibitor toxicity.[4,10]

The international Banff classification for kidney transplant pathology stratifies kidney transplant biopsies according to the combinations of two different sets of semiquantitative histologic lesions, acute and chronic, which distinguish both the rejection subtype (ABMR versus TCMR) and rejection stage (acute/active versus chronic). ABMR and TCMR, and acute/active and chronic lesions, can co-occur at the same time (mixed rejection; chronic/active rejection).[11] This compartmentalization of the histologic picture of kidney transplants leads to a plenitude of diagnoses that are not mutually exclusive, and that may hamper clinical decision making.

We recently demonstrated the potential of RejectClass, a data-driven algorithm for phenotyping kidney transplant biopsies leveraging semisupervised clustering on the acute Banff lesion scores.[12] This mathematical approach resulted in biopsy clusters that were both histologically sound (reflecting known associations between lesion scores and donor-specific human leukocyte antigen [HLA] antibodies) and clinically relevant. In particular, the data-driven clusters demonstrated a better association with graft failure compared with the Banff categories and eliminated the intermediate (borderline) and mixed phenotypes. Moreover, a mathematical estimation of inflammation severity was provided, again predicting the risk of graft failure. Those findings were validated on an extensive external cohort and made available for research use through an online platform.[13]

Despite its promise for reclassification of acute rejection and for assessing rejection severity, this RejectClass system focused solely on the set of acute lesions and donor-specific HLA antibodies.[12] Therefore, this approach offered only a partial view of the heterogeneous causes of graft failure. It is now well established that chronic lesions constitute an independent risk factor for late graft failure,[14] and therefore are included in multivariable models such as the iBox prediction score. Moreover, it was previously suggested that there is heterogeneity in the histologic presentation of early[15] and late[16,17] indication biopsies, which could be summarized into clusters of chronic lesions with different associations with outcome. Therefore, in the evaluation of kidney transplant histology, chronic phenotypes should also be assessed, in complement to the acute rejection subtypes. Despite the suggestion of relevance of previously suggested clusters of chronic rejection,[16,17] hierarchical clustering cannot be applied on new biopsies, is typically unstable, and yields a larger number of smaller clusters that have less clinical relevance.

Significance Statement

Chronic lesions constitute an independent risk factor for late kidney graft failure. However, there is currently no validated data-driven system to realistically describe the chronic pathology of kidney transplants. The authors describe the application of clustering methods to characterize the chronicity and severity of renal allograft disease, finding that kidney transplant biopsies cluster into four chronic phenotypes, validated on data from an independent cohort. These chronic phenotypes are distributed independently of the acute rejection phenotypes, demonstrating independent histologic information. Although chronicity is time dependent, the pattern of chronic lesions is more important than the timing of the biopsy in predicting graft failure. These findings highlight the need to assess both the acute and the chronic components of a biopsy to provide a holistic view of kidney transplant histology.

Given the potential clinical utility of the *k*-means reclassification approach on the acute Banff lesions,[12,18] together with the suggestion that also the chronic lesions are highly relevant for assessment of disease severity and prognosis, we derived and externally validated data-driven chronic phenotypes of kidney transplant biopsies, leveraging the chronic Banff lesion scores. We also demonstrated the complementarity of assessing both the acute and the chronic components to provide a holistic view on kidney transplant histology. This system can be used as a supplement to the Banff classification, and provides a more global and also quantitative assessment of disease activity and chronicity.

METHODS

Data

Training Cohort

The training cohort consisted of all kidney transplantations performed at the University Hospitals Leuven between March 2004 and February 2013 (n=1137). Data extraction was done in March 2018, thus leading to a minimum of 5 years of post-transplant follow-up. Recipients of combined transplants (n=113) and kidney transplantation after another solid organ transplantation (n=24) were excluded. The transplants were performed with negative complement-dependent cytotoxicity crossmatches. The routine clinical data were collected in electronic health records and associated with a SAS database from which the research data were extracted. The standard immunosuppressive maintenance protocol consisted of tacrolimus, mycophenolate, and corticosteroids.[19] The histologic data consisted of all 3622 kidney transplant biopsies performed at the Leuven University Hospitals between April 2004 and February 2015 in 949 patients. Biopsies were performed upon

medical indication (indication biopsies at time of graft dysfunction) or as part of an established follow-up protocols (protocol biopsies).[20] Biopsies with missing lesion scores were excluded (n=73). A total of 3549 biopsies from 937 recipients were available for analysis. Ethnicity was identified and classified by investigator observation. Ethnical information was collected to enable discussions on the generalizability of the findings. This study was approved by the Ethical Committee of the University Hospitals Leuven (S64006).

Validation Cohort

The electronic databases of Lyon University Hospitals (registration #AC-2016-2706) and the Necker Hospital in Paris were screened with the same selection criteria as detailed above. The independent validation set was composed of 4031 complete biopsies from 2054 transplants, of which 1399 biopsies (from 698 transplants) were performed between January 2007 and December 2015 in Lyon, and 2632 biopsies (from 1356 transplants) were performed between March 2009 and October 2019 in Paris. All biopsies were performed either for indication or as part of the routine follow-up at 3 months and 12 months post-transplantation. The clinical and histologic content of those databases was anonymized and transmitted to Leuven as external validation cohort.

Histologic Scoring

All post-transplant kidney allograft biopsies from the training cohort were reviewed by a single pathologist (E. Lerut). The severity of the histologic lesions was assessed semiquantitatively according to the Banff guidelines.[21] Next to the set of acute Banff lesions, tubulitis (t), interstitial inflammation (i), glomerulitis (g), intimal arteritis (v), C4d deposition in peritubular capillaries (C4d), peritubular capillaritis (ptcitis), and thrombotic microangiopathy,[12] in this study, we focused on the following set of histologic chronic lesions: glomerular basement membrane double contours (cg, from 0 to 3), interstitial fibrosis (ci, from 0 to 3), tubular atrophy (ct, from 0 to 3), vascular fibrous intimal thickening (cv, from 0 to 3), mesangial matrix expansion (mm, from 0 to 3), and arteriolar hyalinosis (ah, from 0 to 3). The degree of interstitial fibrosis and tubular atrophy (IFTA) was calculated from the ci and ct scores: IFTA 3: ci=3 and ct=3; IFTA 2: ci≥2 and ct≥2; IFTA 1: ci≥1 and ct≥1. The number of sclerosed glomeruli (glomerulosclerosis; gs score) was scored separately (0 = <10%; 1 = 10%–25%; 2 = 25%–50%; 3 = >50% glomerulosclerosis).[22] The choice of completely disentangling the sets of acute and chronic lesions was motivated by the scientific evidence that chronicity constitutes an independent risk factor in itself.[14,23] The histologic scoring of the validation datasets are provided for the Lyon[24] and the Paris[14] datasets.

Data Analysis

Semisupervised Clustering with Time-Dependent Weighting

We used a similar semisupervised methodology as described recently.[12] Individual lesion scores were first standardized and reweighted based on their association with graft failure. Then, a traditional unsupervised clustering algorithm was applied on the reweighted data using a consensus approach to derive stable clusters. However, because chronic injury is a dynamic process, it was important to account for time dependency when deriving lesion weights, instead of constant weights. We therefore developed a two-step empirical approach to derive time-dependent weights of the chronic lesions. First, local Cox models at time post-transplant t were computed using all of the biopsies performed during a time window of 120 days centered around t. A minimal amount of 50 biopsies from unique patients in the time window were needed to compute a local Cox model, else no score was computed for this time t. By including only one biopsy per patient, we avoided artificially inflating the event rates with repeated measurements. In the second stage, a continuous function was fit to approximate the longitudinal evolution of the univariate Cox model's coefficient over time. Using continuous functions avoids local fluctuation of the weights from unstable estimates and allows weights to be derived even if no biopsies were available at that time post-transplant. We applied winsorizing to limit the effect of outliers' coefficients, defined here as below the 2.5th or above the 97.5th percentiles, on the function approximation. The time-dependent weights were modeled with one of the following functions: linear, quadratic, square root, or logarithmic. The choice of the functional form was determined by the Bayesian information criterion. With such dynamic weighting, chronic lesion scores do not carry the same importance throughout time. Therefore, although some biopsies can be considered similar in a time-independent approach, they can be treated differently once adjusted for time dependency.

Consensus Clustering

Stable clustering was produced through consensus clustering with 400 partitions according to a similar methodology as used in Vaulet *et al.*,[12] with k-means as core clustering algorithm. For each of the 400 partitions, a k-means algorithm with a random initialization was trained on a subset (80%) of the original data similar to Monti *et al.*[25] For the remaining 20% out-of-bag biopsies, a cluster label was assigned based on the nearest cluster centroid. The final clustering was obtained through majority voting consensus over the 400 partitions. The optimal number of clusters was chosen based on the proportion of ambiguous clustering (PAC). Each biopsy was weighted based on its total chronicity score (*i.e.*, the sum of all chronic lesions). Thereby, less frequent biopsies with high chronicity scores acquire more importance during the clustering process

compared with the large set of biopsies with mild chronicity scores. The final clusters were obtained with majority voting over the whole set of clustering partitions.

For the validation phase, the chronic lesion score from new biopsies were standardized and weighted using the parameters derived on the training cohort. The biopsies were then classified to the nearest cluster centroid of the training data. The degree of similarity between two data partitions was assessed with the adjusted Rand index (ARI), which accounts for potential overlapping partitions due to chance. A score of 1 indicates a perfect overlap of the two partitions, whereas a score of 0 indicates a random partitioning. ARI can take values from −1 to 1. A decision tree was trained on the cluster-labeled data to get insight into the clustering process. The tree was generated using the Gini impurity criterion with a minimum of ten biopsies per leaf and a maximum depth of 6.

Survival Analysis

Graft survival times are reported as post-biopsy times in days until graft failure. Patients were censored at the last follow-up date or at time of death. Survival curves are plotted with Kaplan–Meier estimators along with the 95% confidence interval (CI). In case several biopsies from a given patient were classified to the same cluster, their survival times were averaged to avoid overinflating the transplant failure events rate with repeated biopsies. Hazard ratios (HRs) with 95% CI from Cox models were calculated for pairwise comparison of survival curves. In addition, restricted mean survival times (RMSTs)[26] with a 95% CI at 5 and 10 years were evaluated to assess the mean survival time without event within a predefined time range. Unlike HR, RMST is not affected by proportional hazard assumption violations and provides an intelligible measure of mean survival time. Finally, we reported the differences in RMSTs from the baseline cluster, which estimates the difference in average event-free survival, in years, between a given cluster and the baseline category. Joint modeling was used to assess the effect of continuous variables on graft survival, while accounting for repeated biopsies. The joint model consisted of a combination of a mixed model with random slope for the longitudinal response jointly trained with a survival model.[27]

Visualization

We followed a similar approach as described in Vaulet *et al.*[12] to visualize the biopsies on a two-dimensional polar plot. The radius was calculated as the sum of the weighted chronic lesion scores, scaled to the unit interval. It represents the total chronicity score of the biopsy. The theta axis was derived from a linear discriminant analysis (LDA). We trained an LDA on standardized unweighted chronic lesion scores using the biopsies from the two clusters with the worst survival outcome. We then derived the unidimensional score for the whole dataset. Because LDA aims to find the best separation between two classes, it produces a better separation of the chronic phenotypes than unsupervised approaches such as principal component analysis. A marginal model on the basis of generalized estimating equations (with an autoregressive correlation structure) was used to estimate overall association measures in presence of repeated measurements.

All analyses were performed with Python 3.6.6, except the joint modeling which used the JMbayes2 R package.[28] A web application, where other teams can upload their own data and derive the chronic and acute clusters from the individual Banff lesion scores is available at https://rejectionclass.eu.pythonanywhere.com.

RESULTS

Training Cohort Characteristics

Descriptive statistics of the training cohort at the patient (n=937) and biopsy (n=3549) levels are reported in Table 1. On average, 3.79 biopsies (range: 1–11, 74 patients [7.8%] had only one biopsy) were performed per patient; there were 786 indication biopsies (22.1%) and 2763 protocol biopsies (77.9%). Overall, biopsies were performed at a median post-transplant time of 365 days (interquartile range [IQR]: 91–742). The biopsies were categorized according IFTA grade in the following proportions: IFTA 0 (n=1907, 53.7%), IFTA 1 (n=1170, 33.0%), IFTA 2 (n=328, 9.2%) and IFTA 3 (n=144, 4.1%).

Semisupervised Clustering of Chronic Injury Phenotypes

Based on the PAC metric, the optimal number (k) of clusters was 4, although k=3 had very close performance (Supplemental Table 1). Compared with a fully unsupervised approach, which favored three clusters, the clusters from the semisupervised approach were more stable and better differentiated graft outcome (Supplemental Table 1). The four clusters obtained with the semisupervised clustering were labeled from 1 to 4 based on their association with graft failure (Figure 1). Most of the biopsies (n=2876, 81.0%) were assigned to cluster 1. With a median of 349 days post-transplant (IQR: 89–729 days) (Supplemental Figure 1), cluster 1 was dominated by early biopsies with zero or mild lesion scores (Supplemental Table 2). Biopsies in cluster 2 (n=420, 11.8%) had higher grades of ci and ct compared with cluster 1. The median time post-transplant was 718 days (IQR: 367–787), reflecting later biopsies. Cluster 3 contained biopsies (n=195, 5.5%) from a later stage of chronicity with a median time post-transplant of 1097 days (IQR: 742–1826 days). Those biopsies were dominated by severe ci and ct lesion scores, and to a lesser extent, mild to severe cv, ah, and glomerulosclerosis lesions,

Table 1. Demographic, clinical, and histologic characteristics of the patients and biopsies included in the training cohort

Characteristics	Results
Cohort characteristics	Total (n=937)
Donor demographics	
Donor type	
Donation after brain death, *n* (%)	727 (77.6)
Donation after cardiac death, *n* (%)	153 (16.3)
Living donation, *n* (%)	57 (6.1)
Age (yr), mean±SD	47.7±14.7
Male, *n* (%)	498 (53.1)
Diabetes, *n* (%)	24 (2.6)
Recipient demographics	
Age (yr), mean±SD	53.5±13.3
Male, *n* (%)	573 (61.2)
Ethnicity	
Asian, *n* (%)	3 (0.3)
Black, *n* (%)	12 (1.3)
Hispanic, *n* (%)	1 (0.1)
White, *n* (%)	921 (98.3)
Body mass index (kg/m^2), mean±SD	25.4±4.5
Pretransplant donor-specific HLA antibodies, *n* (%)	107 (11.4)
Repeat transplantation, *n* (%)	141 (15)
Cold ischemia time (h), mean±SD	14.2±5.7
Total number of HLA A/B/DR mismatches, mean±SD	2.7±1.3
Biopsy characteristics	Total (n=3549)
Banff 2019 diagnosis	
No rejection, *n* (%)	2690 (75.8)
Borderline changes, *n* (%)	331 (9.3)
TCMR, *n* (%)	287 (8.1)
ABMR, *n* (%)	124 (3.5)
Mixed rejection (ABMR + TCMR), *n* (%)	90 (2.5)
Mixed borderline rejection (ABMR + borderline changes), *n* (%)	27 (0.8)
Interstitial fibrosis and tubular atrophy (IFTA) grade	
IFTA 0, *n* (%)	1907 (53.7)
IFTA 1, *n* (%)	1170 (33.0)
IFTA 2, *n* (%)	328 (9.2)
IFTA 3, *n* (%)	144 (4.1)
Indication biopsies, *n* (%)	786 (22.1)
Days since transplantation, median (interquartile range)	22 (8–105)
eGFR at day of biopsy (ml/min per 1.73 m^2), median (interquartile range)	19.4 (10.8–28.8)
Protocol biopsies, *n* (%)	2763 (77.9)
3 months, *n* (%)	827 (29.9)
12 months, *n* (%)	767 (27.8)
24 months, *n* (%)	644 (23.3)
36 months, *n* (%)	210 (7.6)
48 months, *n* (%)	23 (0.8)
60 months, *n* (%)	292 (10.6)
Days since transplantation, median (interquartile range)	377 (101–752)
eGFR at day of biopsy (ml/min per 1.73 m^2), median (interquartile range)	46.3 (36.5–57.8)

altogether with a large majority of cg0 scores. Finally, biopsies in cluster 4 (n=58, 1.6%) were dominated by mild to severe cg lesion scores. Compared with cluster 3, severe ci and ct scores were less present in cluster 4. The median time post-transplant was 765 days (IQR: 703–1136 days). Biopsies in clusters 1, 2, and 3 had a large majority of cg0 lesions (>97% for all three clusters) whereas cluster 4 gathered all of the remaining cg1, cg2, and cg3 but none of cg0. From a marginal model analysis, the year post-transplant was associated with the total chronicity in kidney transplant biopsies (coefficient, 0.030; 95% CI, 0.028 to 0.032; P<0.001). Although there was an association with time, the clustering of the biopsies solely based on the time post-transplant variable produced clusters without discrimination in graft survival (Supplemental Figure 2 for various values of k). This indicates that the pattern of the chronic lesions is more important than the timing of the biopsy in predicting graft survival.

Because the PAC was similar between k=3 and k=4, we also evaluated the composition of k=3 clusters (Supplemental Figure 3). Compared with an approach with three clusters, having four clusters facilitated the discrimination

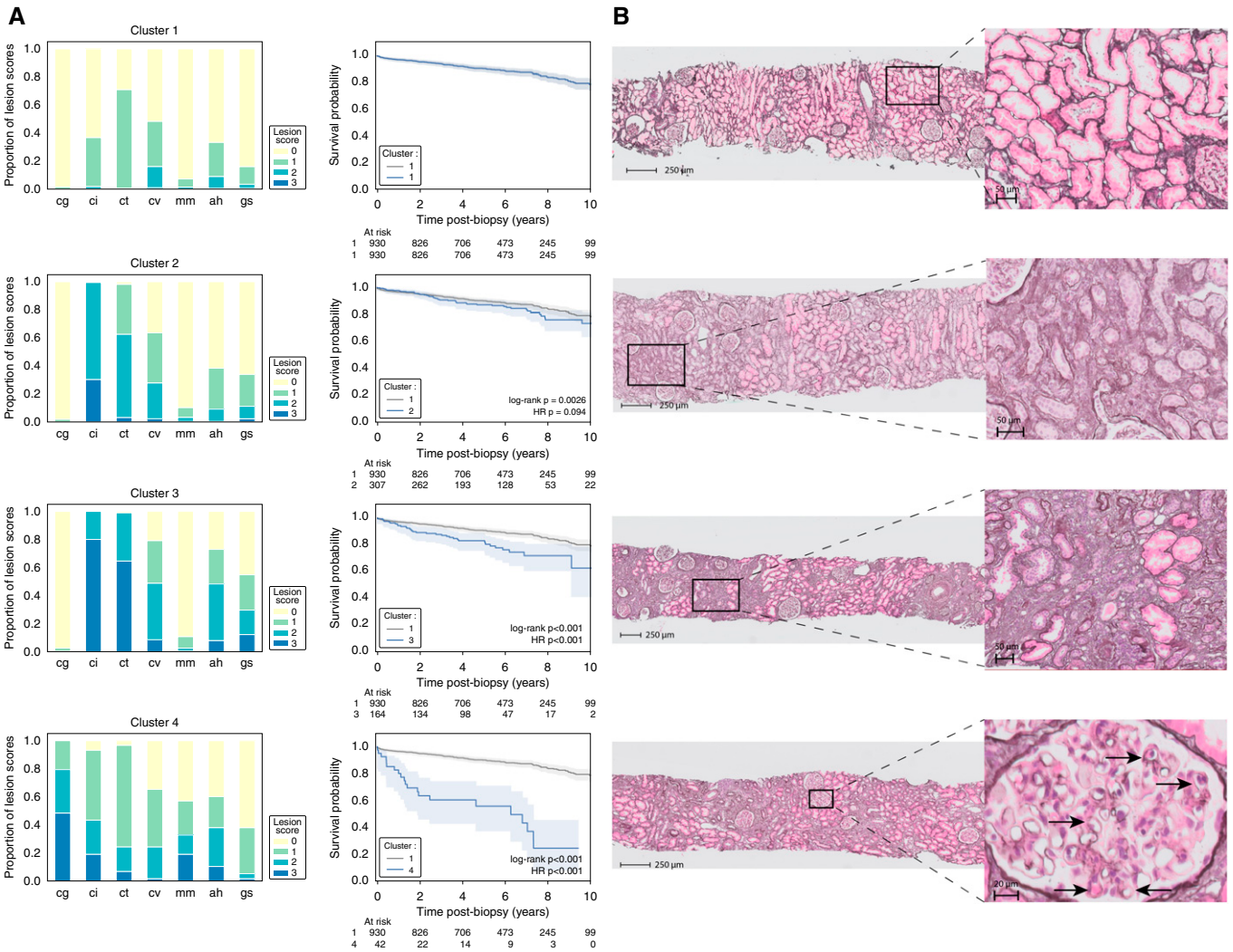

Figure 1. Distribution of the individual chronic lesion scores in the different chronic clusters, and post-biopsy Kaplan–Meier graft survival curves relative to cluster 1 of the training cohort (*n*=3549 biopsies), along with exemplary biopsies. (A) Cluster 1 was dominated by biopsies (*n*=2876, 81.0%) with zero or mild lesion scores and had the best graft survival. Biopsies in cluster 2 (*n*=420, 11.8%) had higher grades of ci and ct compared with cluster 1. Cluster 3 contained biopsies (*n*=195, 5.5%) dominated by severe ci and ct lesion score, and to a lesser extent, mild to severe cv, ah and glomerulosclerosis (gs score) lesions, altogether with a large majority of cg0 scores. The biopsies in cluster 4 (*n*=58, 1.6%) were dominated by mild to severe cg lesion scores. Unlike cluster 3, severe ci and ct scores were less present in cluster 4. Biopsies in clusters 1, 2, and 3 had a large majority of cg0 lesions (>97% for all three clusters) whereas cluster 4 gathered all the remaining cg1, cg2, and cg3 but none of cg0. The biopsies in clusters 3 and 4 demonstrated worse graft survival ($P<0.001$) than biopsies in cluster 1. (B) Exemplary biopsies for each cluster. The biopsy labeled as cluster 1 did not demonstrate any chronic Banff lesions except ah1. The biopsy labeled as cluster 2 was scored with the following lesions: ci2, cv1, ct1, and ah1. The biopsy labeled as cluster 3 was scored with the following lesions: ci3, ct3, cv1, ah1, and gs1. The biopsy labeled as cluster 4 was scored with the following lesions: cg2 (arrows), ci3, ct2, and ah2. cg, glomerular basement membrane double contours; ci, interstitial fibrosis; ct, tubular atrophy; cv, vascular fibrous intimal thickening; mm, mesangial matrix expansion; ah, arteriolar hyalinosis; gs, glomerulosclerosis; HR, hazard ratio.

between a cluster driven by ci and ct (cluster 3) and a cluster dominated by cg (cluster 4), while adding an intermediate phenotype (cluster 2). Supplemental Figure 4 displays the distribution of acute lesions among the k=4 clusters. Chronic clusters 1, 2, and 3 demonstrated similar acute lesion profiles, whereas cluster 4 exhibited a higher rate of i, t, g, C4d, ptcitis, and donor-specific HLA antibodies compared with the other clusters. A decision tree trained to predict the cluster labels was not able to entirely reproduce the whole clustering process (balanced accuracy of 0.875). Although prototypical cases (*e.g.*, nonrejection cases) were easily classified, more intricate cases could not always be correctly classified, which limits the tree's clinical usefulness.

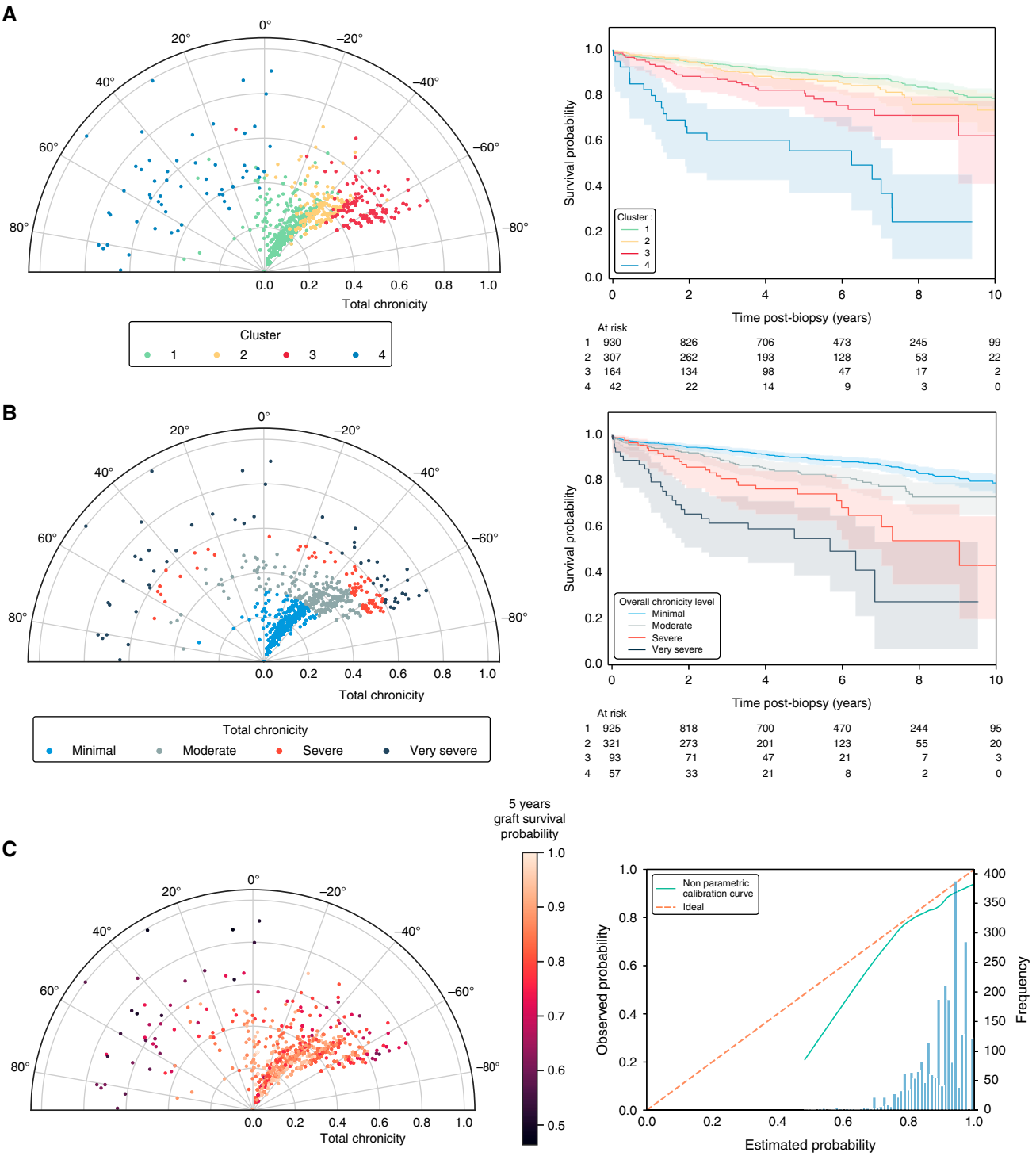

Figure 2. Visualization of the chronic clusters on the polar plot and association with graft survival (*n*=3549 biopsies of the training cohort). (A) Polar plot of the 3549 biopsies of the training cohort overlaid by the four chronic clusters. The radius represents the total chronicity computed as the sum of reweighted chronic lesion scores, scaled to the unit interval (from 0 to 1). The theta angle is directly related to the chronic rejection phenotypes and differentiates between cluster 4 and cluster 3 based on linear discriminant analysis. Biopsies from cluster 1 occupy a central location with limited total chronicity, and biopsies from cluster 2 have an intermediate position between cluster 1 and cluster 3. Biopsies from clusters 3 and 4 share similar range of chronicity scores. However, the patterns of lesions composing the clusters are drastically different, projecting the biopsies in opposite direction on the theta angle of the polar plot. The corresponding Kaplan–Meier curves are displayed on the right. (B) Association of the polar plot

Figure 2. (*Continued*) radius with graft survival in the training cohort. The biopsies were stratified along the radius axis in four strata. The corresponding Kaplan–Meier survival curves are plotted on the right. This demonstrates that the total chronicity (as the sum of the reweighted chronic lesion scores, scaled to the unit interval from 0 to 1), represented by the radius of the polar plot, is positively associated with the risk of graft failure. The total chronicity score was discretized into four different levels of chronicity, corresponding to the following radiuses: "Minimal": radius ≤0.3; "Moderate": radius 0.31–0.5; "Severe": radius 0.51–0.60; and "Very severe": radius >0.6. (C) Estimated graft survival probability at 5 years post-biopsy, computed from the 40 nearest biopsies with the corresponding calibration curve displayed on the right panel.

Visualization of the Clusters

A two-dimensional polar plot visualization of the biopsies based on the chronic cluster attribution is displayed in Figure 2. Most of the biopsies from cluster 4, driven by cg lesion scores, are located on the left half of the plot (positive theta angle), whereas biopsies from other clusters mostly have a negative theta angle. The total chronicity score, represented by the radius, provides a good discrimination in graft survival association as demonstrated by the survival curves from different radius strata (Figure 2B). As expected, biopsies from cluster 1 occupy a central location with a smaller total chronicity, and biopsies from cluster 2 have an intermediate position between cluster 1 and cluster 3. Biopsies from clusters 3 and 4 have a similar range of chronicity scores (*i.e.*, similar radius distribution) (Supplemental Figure 5). However, the patterns of lesions composing the clusters are drastically different, projecting the biopsies in opposite direction on the theta angle of the polar plot. This difference of patterns is also evident in Supplemental Figure 6 which displays various combinations of lesion scores and the time post-transplant on the same polar plots.

Association between Chronic Clusters and Graft Failure

Overall, 125 grafts failed during follow-up. From clusters 1–4, respectively 9.9%, 12.8%, 17.5%, and 44.0% of the grafts failed within 5 years post-biopsy respectively. Clusters 3 and 4 were associated with an increased risk of graft failure in comparison with cluster 1 (Figure 1 and Table 2). HRs of clusters 2, 3, and 4 compared with cluster 1 were 1.33 (95% CI, 0.95 to 1.87), 2.07 (95% CI, 1.41 to 3.04), and 6.57 (95% CI, 4.11 to 10.49), respectively. When we predicted graft failure for each biopsy separately, by using local Kaplan–Meier estimates based on the 40 nearest neighbors, solely based on the chronic lesion scores, the areas under the receiver operating characteristic curve (AUCs) of the probability for graft failure at 2 and 5 years post-biopsy were 0.70 (95% CI, 0.67 to 0.73) and 0.65 (95% CI, 0.63 to 0.68), respectively (Figure 2C).

Comparison between Chronic Clusters and Acute Rejection Phenotypes

Unsupervised dimensionality reduction with principal component analysis on the whole set of lesions separated the acute and the chronic lesions in two almost orthogonal directions (Supplemental Figure 7), suggesting a high degree of independence. The ARI between the acute Banff categories and the chronic clusters was −0.0025, which reflects an almost random partition of the Banff categories among the chronic clusters. Similarly, the ARI between the acute rejection clusters (obtained from the work of Vaulet *et al.*[12]) and the chronic clusters was 0.0133, again reflecting low agreement. Supplemental Figure 8 (top) displays the acute rejection clusters on the chronic polar plot and did not reveal obvious patterns. Additionally, there was no correlation between the radius based on acute lesions and the radius based on chronic lesions (Spearman correlation, 0.0016; $P=0.923$). From this, it can be concluded that acute and chronic radiuses represent independent information.

Association between (Acute and Chronic) Radius and Graft Failure

From the survival part of the multivariable joint model, the HR for graft failure of the acute radius (calculated according to Vaulet *et al.*[12]) and the chronic radius were 5.01 (95% CI, 2.83 to 7.00; $P<0.001$) and 8.33 (95% CI, 5.94 to 10.88; $P<0.001$) respectively, indicating that both radiuses remained strongly associated with graft failure, even when combined in a single model. Compared with univariate models (*i.e.*, the models using either the acute or the chronic radius alone) the combined model demonstrated better discriminative performance over a larger range of time post-transplantation (Table 3) whereas a model solely trained on the acute radius discriminated mostly early biopsies and a model trained on chronic radius best predicted outcome of late biopsies.

Comparison between Chronic Clusters and IFTA Grade

The chronic clusters demonstrate a partial overlap with the existent IFTA categories, with an ARI of 0.35. However, the IFTA categories are less discriminatory with regard to the risk of graft failure than the chronic clusters, as demonstrated by the larger differences in RMST and larger HR of the chronic clusters when compared with the baseline category (*i.e.*, cluster 1 and IFTA 0) (Figure 1, Supplemental Figures 9 and 10, Table 2). The major difference with IFTA is the creation of a cluster driven by cg lesions, which demonstrated the highest risk of graft failure. Supplemental Figure 8 (bottom) displays the IFTA categories on the chronic polar plot.

External Validation

Based on the features' weights and the cluster centroids derived from the consensus clustering process on the training

Table 2. Graft survival, restricted mean survival time (RMST), and difference in RMST (DMRST) at 5- and 10-years post-biopsy, according to each IFTA category and each cluster in the training cohort (*n*=3549 biopsies)

IFTA Category and Cluster Number	% Graft Survival at 5 Years Post-Biopsy	% Graft Survival at 10 Years Post-Biopsy	RMST at 5 Years Post-Biopsy (95% CI)	RMST at 10 Years Post-Biopsy (95% CI)	HR versus Cluster 1 (95% CI)	HR *P* value versus IFTA 0 or Cluster 1	DRMST at 5 Years versus IFTA 0 or Cluster 1 (95% CI)	DRMST at 10 Years versus IFTA 0 or Cluster 1 (95% CI)
IFTA 0	90.7	80.3	4.72 (4.65 to 4.79)	9.02 (8.84 to 9.2)	—	—	—	—
IFTA 1	89.4	79.9	4.70 (4.62 to 4.78)	8.92 (8.71 to 9.14)	1.09 (0.82 to 1.45)	0.547[a]	−0.02 (−0.12 to 0.09)[a]	−0.09 (−0.37 to 0.19)[a]
IFTA 2	83.5	70.8	4.50 (4.34 to 4.66)	8.29 (7.86 to 8.72)	1.82 (1.30 to 2.56)	<0.001[a]	0.22 (0.05 to 0.4)[a]	0.73 (0.26 to 1.19)[a]
IFTA 3	79.6	58.2	4.39 (4.14 to 4.65)	7.76 (7.07 to 8.44)	2.50 (1.66 to 3.79)	<0.001[a]	0.33 (0.07 to 0.59)[a]	1.26 (0.55 to 1.97)[a]
Cluster 1	90.1	78.8	4.71 (4.61 to 4.81)	8.97 (8.56 to 9.38)	—	—	—	—
Cluster 2	87.2	73.8	4.65 (4.45 to 4.86)	8.69 (7.86 to 9.52)	1.33 (0.95 to 1.87)	0.094[b]	0.05 (−0.08 to 0.19)[b]	0.28 (−0.11 to 0.66)[b]
Cluster 3	82.5	62.6	4.43 (4.13 to 4.73)	8.02 (6.67 to 9.36)	2.07 (1.41 to 3.04)	<0.001[b]	0.28 (0.06 to 0.5)[b]	0.95 (0.34 to 1.56)[b]
Cluster 4	56.0	24.9	3.42 (2.66 to 4.18)	5.27 (3.52 to 7.02)[c]	6.57 (4.11 to 10.49)	<0.001[b]	1.29 (0.66 to 1.91)[b]	3.37 (2.17 to 4.57)[b,c]

IFTA, interstitial fibrosis and tubular atrophy.
[a]Versus IFTA 0.
[b]Versus cluster 1.
[c]RMST and DRMST computed for 9.4 years (no data available after this period).

data, new biopsies can be classified in one of the four clusters derived on the training data from Leuven. We applied the trained algorithm to an external set of 4031 biopsies (Supplemental Table 3). A comparison of the cluster proportion between the training and the validation sets is reported in Supplemental Figure 11. Although the cluster ranking in terms of numbers of biopsies was the same between both cohorts, with a dominant cluster 1, proportionally, the other clusters were more prevalent in the validation set. Like the training data, the biopsies of the external cohort that classified as cluster 1 were dominated by low lesion scores, although cv and ah lesions had higher scores compared with the training data. As observed on the training cohort, the biopsies from cluster 3 were also driven by ci and ct with very few cg, whereas the biopsies in cluster 4 all presented mild to severe cg lesion scores (Figure 3A). Biopsies in cluster 2 demonstrated higher lesion scores compared with the same cluster from the training data. All clusters from the validation set were significantly associated with an increased risk of graft failure compared with cluster 1 (Figure 3A) and the two-dimensional polar plot demonstrated similar distribution of the biopsies compared with the training cohort (Figure 3B). Using local Kaplan–Meier estimates based on the 40 nearest neighbors, solely based on the chronic lesion scores, the AUC of the probability for graft failure at 2- and 5-years post-biopsy were 0.72 (95% CI, 0.69 to 0.75) and 0.68 (95% CI, 0.65 to 0.70) respectively (Figure 3, C and D). The total chronicity score also provided a good discrimination in graft survival association on the validation cohort, as demonstrated by the survival curves from different radius strata (Supplemental Figure 12).

Table 3. Discriminative performance in AUC of joint models on the training cohort based on the acute and chronic radius on the polar plot

Starting Time (No. of Patients Still at Risk)	Models								
	Acute Radius Only			Chronic Radius Only			Acute + Chronic Radiuses		
	Delta=1 Year	Delta=2 Years	Delta=5 Years	Delta=1 Year	Delta=2 Years	Delta=5 Years	Delta=1 Year	Delta=2 Years	Delta=5 Years
1 year (*n*=881)	0.799	0.777	0.714	0.492	0.598	0.616	0.756	0.759	0.668
2 years (*n*=856)	0.711	0.691	0.649	0.815	0.785	0.703	0.858	0.814	0.702
5 years (*n*=715)	0.636	0.707	0.589	0.747	0.736	0.752	0.766	0.752	0.741

The joint modeling approach combined a survival model with a longitudinal model which accounts for the repeated biopsies at the patient level. Using dynamic predictions, it was possible to leverage information from prior biopsies up to a predefined starting time to estimate the survival outcome within a delta time window at the patient level. The deltas correspond to the time window after the starting point in which a graft failure is predicted. For instance, with starting time = 2 years and delta = 5 years, the AUC corresponds to the discriminative performance of the model using biopsy data from the first 2 years post-transplant to predict graft failure between 2 and 7 years post-transplant.

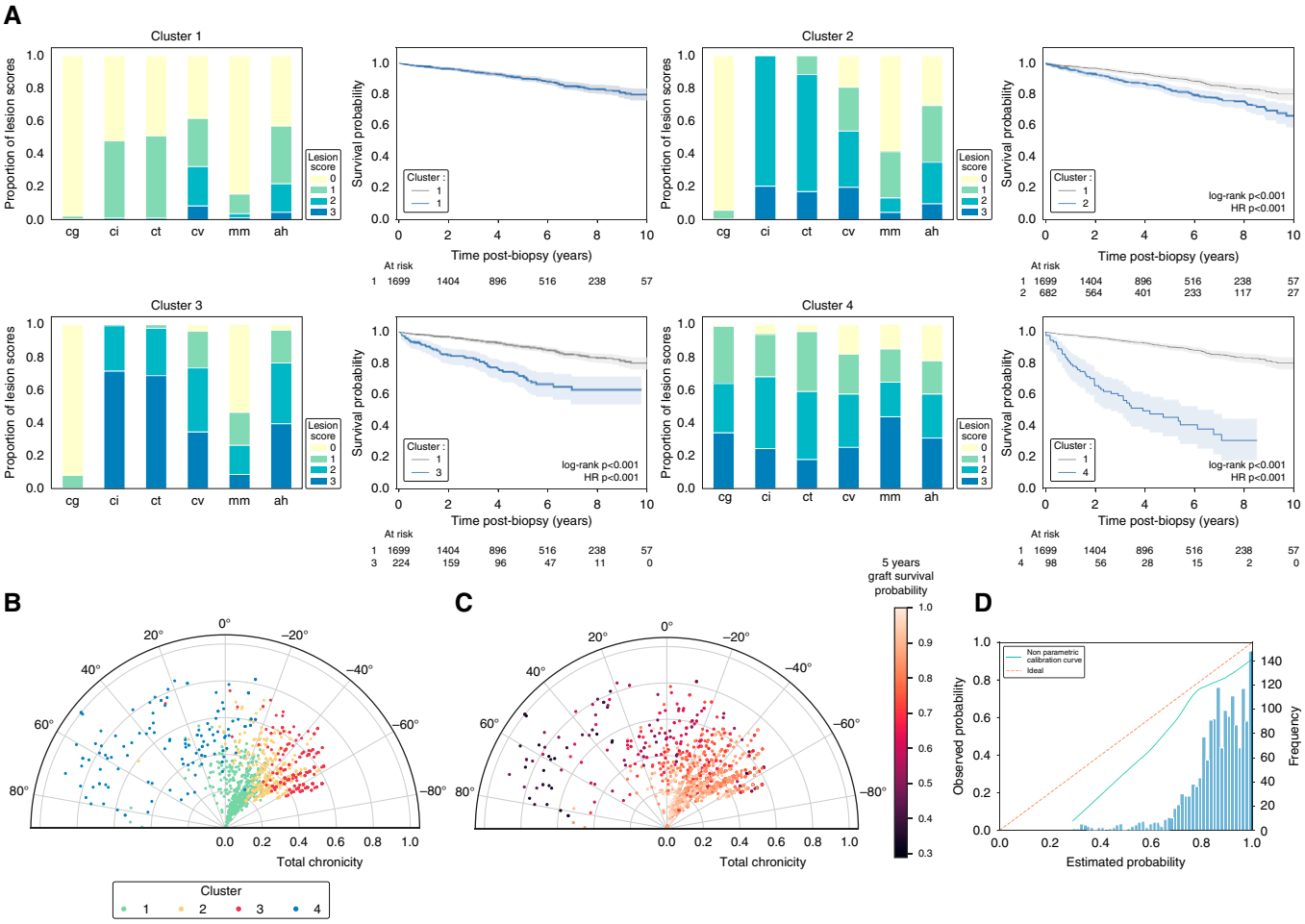

Figure 3. Graft survival association and visualization of the chronic clusters from the external validation cohort (*n*=4031 biopsies). (A) Distribution of the individual chronic lesion scores in the different clusters, and post-biopsy Kaplan–Meier graft survival curves. cg, glomerular basement membrane double contours; ci, interstitial fibrosis; ct, tubular atrophy; cv, vascular fibrous intimal thickening; mm, mesangial matrix expansion; ah, arteriolar hyalinosis; gs, glomerulosclerosis. *P* values refer to HR from the Cox models. (B) Polar plot of the validation cohort overlaid by the four chronic clusters. The radius represents the total chronicity computed as the sum of reweighted chronic lesion scores, scaled to the unit interval (from 0 to 1). The theta angle is directly related to the chronic rejection phenotypes and differentiates between cluster 4 and cluster 3 based on linear discriminant analysis. (C) Estimated graft survival probability at 5 years post-biopsy, computed from the 40 nearest biopsies, and (D) the corresponding calibration curve.

DISCUSSION

In this study, we trained and validated a semisupervised consensus clustering algorithm to cluster post-transplant kidney biopsies based on their chronic Banff lesion scores. Our data-driven algorithm, trained on 3549 biopsies, produced four clusters of different degrees of chronicity. The two clusters with the worst survival outcome were mostly driven by ci and ct (cluster 3) and cg (cluster 4). The clusters obtained on the chronic lesions were distributed independently of the acute lesions, highlighting the interest to assess both the acute and chronic systems in parallel. The total chronicity score and the polar plot approach provide a more granular visualization through a continuous assessment of the overall severity of the chronic injury. These results were validated on an extensive cohort of 4031 external biopsies.

Our finding that primarily ci and ct on the one hand and cg on the other hand were most informative for the clustering approach concurs with previous studies that showed the independent association of these two chronic lesions with graft failure, in multivariable models that included also acute lesions and graft functional characteristics.[14,23] Notwithstanding the dominance of these lesions, also other chronic lesions such as arteriolar hyalinosis, vascular intimal thickening, and mesangial matrix expansion correlate with them, but seem less relevant for predicting outcome and driving the clustering.

The total chronicity in kidney transplant biopsies associated with time post-transplant, as late biopsies tended to have higher chronicity scores than early biopsies. Naturally, this relation to time was also observed with the four chronic clusters. Cluster 1 mostly consisted of early biopsies, cluster 2 contained biopsies performed at intermediate time post-transplant, and clusters 3 and 4 were attributed later after transplantation. Nevertheless, we demonstrated that the clustering of the biopsies did not solely reflect time post-transplantation. Clustering based on time after transplantation alone in and of itself did not discriminate the risk for graft failure, whereas the pattern of chronic lesions associated strongly with graft outcome.

Previous work by Matas *et al.* also evaluated the application of clustering techniques to the Banff histologic lesions.[17] In the Deterioration of Kidney Allograft Function (DeKAF) study consisting of 265 post-transplant biopsies, the authors used unsupervised hierarchical clustering on a joint set of acute and chronic lesions. The clustering was only performed on indication biopsies without adjusting for the time post-transplant and resulted in 13 clusters of which six main clusters contained more than 13 biopsies. The resulting clusters, as derived from both acute and chronic lesions, described different entities with various degrees of fibrosis, inflammation, and/or chronicity. The long-term follow-up of the study[16] confirmed that the graft failure rate was significantly different per main cluster, as observed in the preliminary results published in 2010.[17] The nature of the algorithm chosen by Matas *et al.* and their choice to include a mixture of acute and chronic lesions prevented us from performing meaningful comparisons with this study.

In our approach, we completely disentangled the sets of acute and chronic lesions, following the scientific evidence that chronicity constitutes an independent risk factor in itself.[14,23] Additionally, unsupervised dimensionality reduction separated the acute and the chronic lesions in two almost orthogonal directions in the principal component analysis; ARIs between acute and chronic clusters were low; and correlations between acute and chronic radius scores were absent. All this suggests a high degree of independence of the acute from the chronic lesions (when assessed on the same biopsy). These findings highlight the interest to assess both the acute and chronic systems in parallel. It also illustrates that the acute lesions in a biopsy cannot be used to infer exclusive causal relations to the chronic lesions observed in the same biopsies. The latter are potentially related to earlier injury processes unrelated to the active inflammation sometimes observed simultaneously. This can only be assessed by considering the histologic evolution over time, in repeated biopsies, which requires more advanced and complex time-dependent models.[29] Our chronic reclassification can be seen as a supplement to the existing Banff phenotypes. By disentangling the chronicity component from the acute phenotypes, our approach can potentially simplify the current conundrum of having multiple versions of each Banff phenotype (acute versus chronic versus chronic active).

By demonstrating the discriminative power of the chronic lesions on graft failure, independently of inflammation, this study confirms the clinical interest of exploiting a more extensive assessment of the chronicity of kidney transplant biopsies, which remains underappreciated in the current Banff classification. With a multivariate joint modeling approach, we demonstrated that a continuous assessment of both inflammation and chronicity severities contributed complementarily to predict graft outcome. Therefore, we suggest that both the acute and the chronic components of a biopsy should be assessed in parallel and presented jointly to clinicians. Finally, additional clinical interests lie in the potential discriminative power of chronic profiles for therapeutic interventions, as a high degree of chronicity might illustrate lesser treatment response in contrast to acute inflammation without chronicity, which could be more amenable to interventions. The demonstration of such clinical utility requires additional validation studies. By providing a freely accessible web application, we encourage other research groups to evaluate the generalizability of our results on different patient populations.

Because the clustering algorithm was trained on the Banff lesion scores, the results directly depend on the quality of the histologic scoring, which is subject to inter- and intra-observer variability. Despite this essential drawback of human interpretation of lesion scores, the algorithm does not appear to have suffered from this, as similar results were obtained on an external cohort, assessed by different pathologists. The human dependence of the histologic scoring warrants the analysis of alternative and automated data extraction approaches such as molecular expression or computerized imaging data. Although these were not available in the training dataset, recently proposed lesion scores such as ti[30] could further improve our reclassification framework in producing more granular phenotypes. Although the chronic clusters bear some predictive value, the present reclassification framework, by focusing solely on a subset of histologic lesion scores, is not intended for prognostication. For accurate prognostication, additional variables beyond histology should be taken into account in multivariable models such as iBox.[14,31]

Like all statistical and machine learning methods, the current findings do not guarantee the causality, and therefore pathophysiologic mechanisms cannot be deducted from any clusters. Finally, although the choice of the core algorithm was motivated by several obvious advantages such as easy classification of new data points, more elaborate clustering approaches could further improve the field. Especially, precise modeling of the clustering time dependency with time-dependent model-based clustering could be valuable and warrants additional studies.

In conclusion, based on semisupervised clustering, we identified and validated four clusters of chronic kidney transplant injury. We demonstrated the complementarity of assessing total chronicity on top of the estimation of disease activity from acute lesion scores. This approach delivers a holistic and quantitative assessment of kidney transplant injury phenotypes and severity, as supplement to the current Banff classification. Separating the degree (often nonspecific) of chronicity from rejection activity can significantly reduce the complexity of the description of kidney transplant histology. Further validation of the system (*e.g.*, *via* the web application) is necessary to gain more widespread adoption by the transplant community, on the way to the eventual integration of these algorithms into accessible clinical decision support systems.

DISCLOSURES

M. De Vos reports consultancy fees from Albus Health; reports ownership interest with Circadian Therapeutics; and reports research funding from Janssen Pharmaceuticals. M. Emonds reports advisory or leadership role with Eurotransplant tissue typing (advisory committee). D. Kuypers reports consultancy fees from Astellas, CSL Behring, GlaxoSmithKline, Hansa Biopharma, Sangamo Therapeutics, Takeda, and UCB; reports research funding from Astellas; reports honoraria from Astellas, CSL Behring, GlaxoSmithKline, Hansa Biopharma, Takeda, and UCB; reports advisory or leadership role with *Current Clinical Pharmacology*, *Therapeutic Drug Monitoring*, *Transplantation* (associate editor), and *Transplantation Reviews* (editorial board member); and is on the speakers bureau for Astellas and Hikma. M. Naesens reports research funding from CareDx; and reports patents or royalties: inventor of two patents related to the Research Foundation Flanders (FWO) SBO application: the first, EP19152365.3: mRNA-based biomarkers for antibody-mediated transplant rejection (this biomarker was licensed in September 2020 to CareDx, a precision medicine solutions company focused on solutions for transplant patients); the second, PCT/EP2018/097044: Biomarkers for typing allograft recipients (patent application submitted December 2018). O. Thaunat reports consultancy fees from Biotest, AstraZeneca, and Novartis; reports research funding from BioMérieux, Bristol-Myers Squibb, and Immucor; reports honoraria from Astellas, Biotest, and Novartis; and reports advisory or leadership role with European Society for Organ Transplantation. E. Van Loon reports research funding from Research Foundation Flanders fellowship grant; and reports patents or royalties for European patent EP19152365.3: mRNA-based biomarkers for antibody-mediated transplant rejection. All remaining authors have nothing to disclose.

FUNDING

This work was supported by the Research Foundation Flanders (FWO) and the Flanders Innovation and Entrepreneurship Agency (VLAIO), with a TBM project (IWT.150199), and by the KU Leuven C3 internal grant (C32/17/049). M. Naesens and B. Sprangers are senior clinical investigators of the FWO (supported by grants 1844019N and 1842919N, respectively). T. Vaulet, E. Van Loon, and J. Callemeyn are supported by the FWO through holding fellowship grants (1S93918N, 1143919N, and 1196119N, respectively). O. Thaunat is supported by the Agence Nationale de la Recherche (ANR-16-CE17-0007-01), the Fondation pour la Recherche Médicale (PME20180639518), and the Etablissement Français du Sang. B. De Moor is supported by KU Leuven Research Fund (projects C16/15/059, C3/19/053, C24/18/022, C3/20/117, C3I-21-00316), Industrial Research Fund (fellowships 13-0260, IOFm/16/004, IOFm/20/002), and Leuven Research and Development bilateral industrial projects: FWO (EOS project 30468160 [SeLMA] and SBO project I013218N). B. De Moor also received funding from the Flemish Government (AI Research Program), VLAIO (City of Things COT.2018.018), Industrial Projects (HBC. 2018.0405), and the European Commission. This project has received funding from the H2020 European Research Council (885682). A. Loupy is supported by Institut National de la Santé et de la Recherche Médicale (ATIP Avenir). O. Aubert is supported by a Fondation Bettencourt Schueller fellowship grant. G. Divard is supported by a Fondation pour la Recherche Médicale fellowship grant.

ACKNOWLEDGMENTS

The authors thank the participating transplant centers, clinicians, surgeons, nursing staff, and patients.

AUTHOR CONTRIBUTIONS

M. Naesens and T. Vaulet conceptualized the study, were responsible for the methodology, and wrote the original draft; P. Koshy and T. Vaulet were responsible for visualization; T. Vaulet was responsible for formal analysis and software; M. Naesens was responsible for project administration; B. De Moor, D. Kuypers, and M. Naesens were responsible for funding; B. De Moor, D. Kuypers, and M. Naesens, and O. Thaunat were responsible for resources; B. De Moor, M. De Vos, and M. Naesens were responsible for supervision; O. Aubert, G. Divard, V. Dubois, A. Loupy, M. Rabeyrin, and O. Thaunat were responsible for data curation and validation; J. Callemeyn, K. De Vusser, M.-P. Emonds, E. Lerut, M. Naesens, A. Senev, B. Sprangers, A. Van Craenenbroeck, and E. Van Loon were responsible for data curation; M. De Vos and T. Vaulet were responsible for validation; all authors reviewed and edited the manuscript.

DATA SHARING STATEMENT

The clinical data that support the findings of this study are available on request from the corresponding author, M. Naesens. The data are not publicly available due to privacy regulations.

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2022030290/-/DCSupplemental.

Supplemental Table 1. Performance indices for a range of k values used to define the optimal number of clusters in the semisupervised clustering algorithm (n=3549 biopsies of the training cohort).

Supplemental Table 2. Details of the individual Banff lesion scores of the training cohort according to the chronic clusters (n=3549).

Supplemental Table 3. Demographic, clinical, and histological characteristics of the patients and biopsies included in the validation cohort.

Supplemental Figure 1. Density distribution of time post-transplantation of the biopsies in the training cohort (n=3549), per cluster.

Supplemental Figure 2. Association check for between ... and (not between ... to) graft failure and biopsy clustering based solely on the time post-transplant variable, irrespective of the lesion scores, for various k (n=3549 biopsies of the training cohort).

Supplemental Figure 3. Results for the semisupervised consensus clustering for k=3 on the training cohort (n= 3549 biopsies).

Supplemental Figure 4. Distribution of the individual acute lesion scores in the different chronic clusters, and post-biopsy Kaplan–Meier graft survival curves relative to cluster 1 of the training cohort (n=3549 biopsies).

Supplemental Figure 5. Distribution of total chronicity scores per cluster (n=3549 biopsies of the training cohort).

Supplemental Figure 6. Various combinations of lesion scores and year post-transplant displayed on the polar plots (n=3549 biopsies of the training cohort).

Supplemental Figure 7. Unsupervised dimensionality reduction with principal component analysis (n=3549 biopsies of the training cohort).

Supplemental Figure 8. Acute rejection clusters and IFTA grades on the chronic polar plot (n=3549 biopsies of the training cohort).

Supplemental Figure 9. Distribution of the chronic lesion scores among the IFTA categories (n=3549 biopsies of the training cohort).

Supplemental Figure 10. Association of the IFTA phenotypes and graft failure.

Supplemental Figure 11. Comparison of the proportion of clusters in the biopsies of the training (n=3549) and the validation (n=4031) sets.

Supplemental Figure 12. Association of the polar plot radius with graft survival in the validation cohort (n=4031 biopsies).

Python code for semisupervised consensus clustering.

REFERENCES

1. Rana A, Gruessner A, Agopian VG, Khalpey Z, Riaz IB, Kaplan B, et al.: Survival benefit of solid-organ transplant in the United States. *JAMA Surg* 150: 252–259, 2015
2. Hellemans R, Kramer A, De Meester J, Collart F, Kuypers D, Jadoul M, et al.: Does kidney transplantation with a standard or expanded criteria donor improve patient survival? Results from a Belgian cohort. *Nephrol Dial Transplant* 36: 918–926, 2021
3. Coemans M, Süsal C, Döhler B, Anglicheau D, Giral M, Bestard O, et al.: Analyses of the short- and long-term graft survival after kidney transplantation in Europe between 1986 and 2015. *Kidney Int* 94: 964–973, 2018
4. Mayrdorfer M, Liefeldt L, Wu K, Rudolph B, Zhang Q, Friedersdorff F, et al.: Exploring the complexity of death-censored kidney allograft failure. *J Am Soc Nephrol* 32: 1513–1526, 2021
5. Nankivell BJ, Borrows RJ, Fung CL, O'Connell PJ, Allen RD, Chapman JR: The natural history of chronic allograft nephropathy. *N Engl J Med* 349: 2326–2333, 2003
6. Stegall MD, Cornell LD, Park WD, Smith BH, Cosio FG: Renal allograft histology at 10 years after transplantation in the tacrolimus era: Evidence of pervasive chronic injury. *Am J Transplant* 18: 180–188, 2018
7. Van Loon E, Lerut E, Naesens M: The time dependency of renal allograft histology. *Transpl Int* 30: 1081–1091, 2017
8. Chand S, Atkinson D, Collins C, Briggs D, Ball S, Sharif A, et al.: The spectrum of renal allograft failure. *PLoS One* 11: e0162278, 2016
9. Van Loon E, Lerut E, Senev A, Coemans M, Pirenne J, Monbaliu D, et al.: The histological picture of indication biopsies in the first 2 weeks after kidney transplantation. *Clin J Am Soc Nephrol* 15: 1484–1493, 2020
10. Van Loon E, Senev A, Lerut E, Coemans M, Callemeyn J, Van Keer JM, et al.: Assessing the complex causes of kidney allograft loss. *Transplantation* 104: 2557–2566, 2020
11. Loupy A, Haas M, Roufosse C, Naesens M, Adam B, Afrouzian M, et al.: The Banff 2019 Kidney Meeting Report (I): Updates on and clarification of criteria for T cell- and antibody-mediated rejection. *Am J Transplant* 20: 2318–2331, 2020
12. Vaulet T, Divard G, Thaunat O, Lerut E, Senev A, Aubert O, et al.: Data-driven derivation and validation of novel phenotypes for acute kidney transplant rejection using semi-supervised clustering. *J Am Soc Nephrol* 32: 1084–1096, 2021
13. RejectClass: RejectClass algorithm for classification and quantification of kidney transplant rejection. Available at https://rejectionclass.eu.pythonanywhere.com. Accessed September 2, 2022
14. Loupy A, Aubert O, Orandi BJ, Naesens M, Bouatou Y, Raynaud M, et al.: Prediction system for risk of allograft loss in patients receiving kidney transplants: International derivation and validation study. *BMJ* 366: l4923, 2019
15. Naesens M, Kuypers DR, De Vusser K, Vanrenterghem Y, Evenepoel P, Claes K, et al.: Chronic histological damage in early indication biopsies is an independent risk factor for late renal allograft failure. *Am J Transplant* 13: 86–99, 2013
16. Matas AJ, Fieberg A, Mannon RB, Leduc R, Grande J, Kasiske BL, et al.: Long-term follow-up of the DeKAF cross-sectional cohort study. *Am J Transplant* 19: 1432–1443, 2019
17. Matas AJ, Leduc R, Rush D, Cecka JM, Connett J, Fieberg A, et al.: Histopathologic clusters differentiate subgroups within the nonspecific diagnoses of CAN or CR: Preliminary data from the DeKAF study. *Am J Transplant* 10: 315–323, 2010
18. Vasquez-Rios G, Menon MC: Kidney transplant rejection clusters and graft outcomes: Revisiting Banff in the era of "big data". *J Am Soc Nephrol* 32: 1009–1011, 2021
19. Senev A, Coemans M, Lerut E, Van Sandt V, Daniëls L, Kuypers D, et al.: Histological picture of antibody-mediated rejection without donor-specific anti-HLA antibodies: Clinical presentation and implications for outcome. *Am J Transplant* 19: 763–780, 2019
20. Coemans M, Van Loon E, Lerut E, Gillard P, Sprangers B, Senev A, et al.: Occurrence of diabetic nephropathy after renal transplantation despite intensive glycemic control: An observational cohort study. *Diabetes Care* 42: 625–634, 2019
21. Roufosse C, Simmonds N, Clahsen-van Groningen M, Haas M, Henriksen KJ, Horsfield C, et al.: A 2018 reference guide to the Banff classification of renal allograft pathology. *Transplantation* 102: 1795–1814, 2018
22. Sethi S, D'Agati VD, Nast CC, Fogo AB, De Vriese AS, Markowitz GS, et al.: A proposal for standardized grading of chronic changes in native kidney biopsy specimens. *Kidney Int* 91: 787–789, 2017
23. Naesens M, Lerut E, Emonds MP, Herelixka A, Evenepoel P, Claes K, et al.: Proteinuria as a noninvasive marker for renal allograft histology and failure: An observational cohort study. *J Am Soc Nephrol* 27: 281–292, 2016
24. Koenig A, Chen CC, Marçais A, Barba T, Mathias V, Sicard A, et al.: Missing self triggers NK cell-mediated chronic vascular rejection of solid organ transplants. *Nat Commun* 10: 5350, 2019
25. Monti S, Tamayo P, Mesirov J, Golub T: Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn* 52: 91–118, 2003
26. Royston P, Parmar MK: Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 13: 152, 2013
27. Asar Ö, Ritchie J, Kalra PA, Diggle PJ: Joint modelling of repeated measurement and time-to-event data: An introductory tutorial. *Int J Epidemiol* 44: 334–344, 2015
28. Rizopoulos D: The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J Stat Softw* 72: 1–46, 2016
29. Coemans M, Senev A, Van Loon E, Lerut E, Sprangers B, Kuypers D, et al.: The evolution of histological changes suggestive of antibody-mediated injury, in the presence and absence of donor-specific anti-HLA antibodies. *Transpl Int* 34: 1824–1836, 2021
30. Mengel M, Reeve J, Bunnag S, Einecke G, Jhangri GS, Sis B, et al.: Scoring total inflammation is superior to the current Banff inflammation score in predicting outcome and the degree of molecular disturbance in renal allografts. *Am J Transplant* 9: 1859–1867, 2009
31. Naesens, M, Budde K, Hilbrands L, Oberbauer R, Bellini MI, Glotz D, et al.: Surrogate endpoints for late kidney transplantation failure. *Transpl Int* 35: 10136, 2022

AFFILIATIONS

[1]ESAT Stadius Center for Dynamical Systems, Signal Processing, and Data Analytics, KU Leuven, Leuven, Belgium
[2]Paris Translational Research Center for Organ Transplantation, Université de Paris, INSERM, PARCC, Paris, France; Kidney Transplant Department, Necker Hospital, Assistance Publique–Hôpitaux de Paris, Paris, France
[3]CIRI, INSERM U1111, Université Claude Bernard Lyon I, CNRS UMR5308, Ecole Normale Supérieure de Lyon, Univ. Lyon, Lyon, France
[4]Department of Transplantation, Nephrology, and Clinical Immunology, Hospices Civils de Lyon, Edouard Herriot Hospital, Lyon, France
[5]Department of Imaging and Pathology, University Hospitals Leuven, Leuven, Belgium
[6]Department of Microbiology, Immunology, and Transplantation, KU Leuven, Leuven, Belgium
[7]Histocompatibility and Immunogenetics Laboratory, Belgian Red Cross–Flanders, Mechelen, Belgium
[8]Department of Nephrology and Kidney Transplantation, University Hospitals Leuven, Leuven, Belgium
[9]Department of Pathology, Hospices Civils de Lyon, Bron, France
[10]Human Leukocyte Antigen (HLA) Laboratory, French National Blood Service (EFS), Décines-Charpieu, France
[11]Department of Development and Regeneration, University Hospitals Leuven, KU Leuven, Leuven, Belgium