Novel Phenotypes for Acute Kidney Transplant Rejection Using Semi-Supervised Clustering

We note that Vaulet et al.\(^1\) confirmed the Deterioration of Kidney Allograft Function (DeKAF) studies that (1) identified Banff histologic clusters with differing postbiopsy graft survival; and (2) showed that markers of antibody-mediated rejection affect graft survival, even after adjusting for inflammation (e.g., Banff inflammation, tubulitis, glomerulitis \(g\), peritubular capillaritis scores).\(^2\)\(^-\)\(^4\) In spite of methodologic similarities, Vaulet et al. neglected to cite DeKAF publications that provide insights into cluster phenotypes of late allograft pathologies. The clustering studies differed as follows: (1) DeKAF considered all histologic lesions (acute and chronic) for clustering, whereas Vaulet et al. only considered those associated with acute inflammation plus C4d, donor-specific antibodies, and thrombotic microangiopathy; (2) DeKAF included indication biopsy specimens, whereas Vaulet et al. included both indication and surveillance biopsy specimens; (3) for DeKAF, the median time from transplant to biopsy was 5.7 years, whereas 83.3% of indication biopsies were in the first year in the Vaulet et al. study (median, 22 days post-transplant); and (4) DeKAF’s unsupervised clustering identified six primary clusters, whereas Vaulet et al. identified four. However, because they found the “histologic and clinical relevance” of their unsupervised clusters to be unclear, the authors then used a semisupervised clustering approach—weighing the histologic features with survival information—and identified six clusters.

There are striking similarities between clusters in the studies. Both studies identified a cluster with mild fibrosis but little inflammation, both identified a cluster consistent with acute T cell–mediated rejection, and both identified a cluster with “\(g\)” as the predominant histologic score. For Vaulet et al., these three clusters were similar to the remaining three, except for the absence of donor-specific antibodies; for DeKAF, differences in both acute and chronic lesions differentiated clusters.

There are aspects of Vaulet et al.’s findings that limit their clinical relevance. First, the identified clusters were not purely data driven, but were determined on the basis of phenotypes they expected to exist. They chose to disregard results with three or fewer clusters because they were “not helpful to describe different phenotypes.” Indeed, the authors rejected their unsupervised clusters because they did not “reflect the clinical reality and previous knowledge on the relevance of these lesions [g, peritubular capillaritis, and C4d] and [antibody-mediated rejection].” Hence, the resulting clusters can only be interpreted within the rigid framework imposed by the authors. Secondly, we question the validity of their conclusion that they “showed statistically improved prediction of graft failure with the clustering approach compared with using the Banff categories.” The cluster features were weighted on the basis of graft survival, and their prognostic significance can only be evaluated in external data. Although we agree with the authors that phenotypes beyond the Banff categories may exist, the clusters identified by Vaulet et al. only confirm the importance of lesions already known to be associated with graft failure.

DISCLOSURES

R.B. Mannon reports serving on the American Society of Nephrology (ASN) Grants Committee, as chair of the ASN Policy and Advocacy Committee, as the chair of the data safety monitoring board for the National Institute of Diabetes and Digestive and Kidney Diseases/National Institutes of Health, on the program committee for The Transplantation Society 2020 and 2022, as cochair of the Scientific Registry of Transplant Recipients Review Committee, and as chair of Women in Transplantation; receiving research funding from Astellas, CareDx, CSL Behring, Mallinckrodt, Quark Pharmaceuticals, and Transplant Genomics Inc.; receiving honoraria from CSL Behring, Hansa, Novartis, Sanofi, and Vitaeris; and serving on the steering committee for the Vitaeris VTX01 IMAGINE Trial. A.J. Matas reports receiving research funding from Alexion, Astellas, Bristol Myers Squibb, CareDx, Shire, and Veloxis; receiving honoraria from Astellas, CareDx, CSL Behring, and Veloxis; serving as a scientific advisor for, or member of, CareDx, CSL Behring, and Jazz Pharma; and having consultancies with Veloxis. D. Rush reports receiving research funding and honoraria from, serving as a scientific advisor for/member of, serving on a speakers bureau for, and having other interests/relationships (as a meeting steering committee member) in Astellas Canada Inc. The remaining author has nothing to disclose.

FUNDING

None.

REFERENCES


Arthur J. Matas,1 Roslyn B. Mannon,2 David Rush,3 and Erika Helgeson4

1Transplantation Division, Department of Surgery, University of Minnesota, Minneapolis, Minnesota
2University of Nebraska Medical Center and Nebraska–Western Iowa Veterans Affairs Medical Center, Omaha, Nebraska
3Department of Internal Medicine, University of Manitoba, Winnipeg, Manitoba, Canada
4Biostatistics Division, School of Public Health, University of Minnesota, Minneapolis, Minnesota

doi: https://doi.org/10.1681/ASN.2021040572

Authors’ Reply

We thank Matas et al. for their interest in our algorithm for reclassification of acute kidney transplant rejection.1 We agree that data-driven approaches may add value to the iteratively developed Banff classification.2 Moreover, our approach not only provides an easily accessible system for reclassification, it also overcomes the limitations of intermediate and mixed phenotypes, and adds information on disease severity, which is potentially relevant for clinical decision making.

In contrast to what Matas et al. conclude, we do not think that our algorithm’s value is diminished by the similarities between our clusters1 and Banff categories.2 On the contrary, our clusters and visual presentation1 mathematically support the validity of the Banff classification, because both the Banff system and our approach distinguish rejection subtypes primarily on the basis of tubulointerstitial versus microcirculation inflammation, and donor-specific HLA antibodies (HLA-DSA). Furthermore, the similarities between our clusters and the Banff phenotypes enable clinical interpretation of our clusters (clusters 1 and 4 representing no rejection, cluster 3 representing T cell–mediated rejection, cluster 5 representing antibody-mediated rejection). The main, and clinically interesting, differences between our clusters and Banff are (1) the elimination of the disputed borderline category; (2) the addition of “mixed rejection” cluster 6; and (3) the addition of the novel cluster 2, representing the intriguing phenotype of HLA-DSA–negative microcirculation inflammation.3

How our clusters relate to the unvalidated clusters of the Deterioration of Kidney Allograft Function (DeKAF) study4 is more difficult to assess. In contrast to our k-mean clusters, specifically intended for reclassification of acute rejection on the basis of acute lesion scores,1 DeKAF hierarchical clusters were essentially developed for differentiation of chronic injuries, mostly on the basis of features of chronicity.4 Moreover, we neither have access to the exact DeKAF algorithm nor data on the TATR (t-IFTA) lesion, which seems necessary for the DeKAF clustering.

Therefore, we concur that our algorithm, as any other, is not purely data driven but is dependent on the following human choices: (1) the clinical use, i.e., reclassification of acute rejection1 (versus chronic injury4); (2) the method, i.e., k-means clustering, which allows reclassifying new biopsy specimens without the need for retraining1 (versus hierarchic clustering,4 which is often unstable and leads to larger numbers of smaller clusters that are difficult to reproduce); (3) the included features, i.e., acute lesions and HLA-DSA1 (versus selected, mainly chronic,4 lesions but not HLA-DSA4); and (4) the biopsy specimens included in training, i.e., early and late protocol and indication biopsy specimens5 (versus later indication biopsy specimens4).

Additionally, and importantly, we chose semisupervised clustering. In our first unsupervised approach, only two main parameters contributed: (1) presence of HLA-DSA, and (2) the extent of tubulointerstitial inflammation (Supplemental Figure 1 and Supplemental Table 1).1 In unsupervised clustering, the importance of microcirculation inflammation thus seemed underestimated, as compared with extensive literature. By weakly informing the clustering on the relation between individual lesions and graft outcome, which was our final semisupervised approach, we further improved cluster stability and predictive performance and clinical interpretability. We derived the number of clusters from the proportion of ambiguous clustering; we also did not interfere with the features or their weights to compute the clusters.1

Obviously, our semisupervised approach led to some circular argument in the association between the clusters and outcome in our training cohort (Leuven); it was therefore essential to demonstrate the validity of our locked clustering algorithm, the visual presentation, and the added value beyond Banff, on totally independent cases (Paris and Lyon, in total 3835 biopsy specimens from 1989 patients; see detailed