SUPPLEMENTARY MATERIAL

Contents

Supplemental Results

S1 Table. Performance indices for a range of k values used to define the optimal number of clusters in the semi-supervised clustering algorithm

S2 Table. Details of cluster composition according to the individual Banff lesions scores and donor-specific HLA antibodies (N=3510 biopsies of the derivation cohort).

S3 Table. Distribution of biopsies among clusters and stratification into protocol vs indication biopsies (N=3510 biopsies of the derivation cohort).

S4 Table. Demographic, clinical and histological characteristics of the patients and biopsies included in the validation dataset.

S5 Table. Contingency tables comparing the Banff 2019 diagnosis and the 6 clusters obtained on the external validation dataset.

S1 Fig. Distribution of the individual acute lesion scores in the clusters using an unweighted approach, and post-biopsy Kaplan-Meier graft survival curves relative to cluster 1 of the derivation cohort

S2 Fig. Distribution of chronic lesions in the 6 acute lesion clusters

S3 Fig. Relative distances to the closest clusters' boundary

S4 Fig. Decision tree of the clustering process

S5 Fig. Various combinations of lesions scores displayed on the polar plots.

S6 Fig. Comparison of indication vs. protocol biopsies, as superposed on the polar plot.

S7 Fig. Post-biopsy graft survival in the three DSA-/DSA+ pair of clusters.

S8 Fig. Post-biopsy graft survival in the six clusters, according to the adjustment method for repeated biopsies per patient.

S9 Fig. Comparison of cluster proportion per center

S10 Fig. Distribution of the individual acute lesion scores in the different clusters, and post-biopsy Kaplan-Meier graft survival curves relative to cluster 1 of the external validation cohort

S11 Fig: Overlay of the data from Leuven and the external dataset in the polar plot, according to the six clusters identified in the derivation cohort.

S12 Fig. Distribution of the radius from the polar plot

S13 Fig. Association with graft survival in the polar plot visualization tool based on the validation data

Supplemental Results

Unsupervised clustering of rejection phenotypes

Fully unsupervised clustering of our biopsy cohort (N=3510) yielded an optimum of 4 different clusters, based on the proportion of ambiguous clustering (S1 Fig). Compared to cluster 1 (essentially normal biopsies), the three other clusters associated significantly with impaired graft survival. However, their histological and clinical relevance were less clear, as none of these three clusters were defined based on microcirculation inflammation and antibody activity (glomerulitis, peritubular capillaritis and C4d), suggesting that the number of clusters was insufficient to reflect the clinical reality and previous knowledge on the relevance of these lesions and ABMR. Increasing the number of clusters created clusters that were no longer associated with impaired graft survival compared to cluster 1 (S1 Table).

Comparison of disease clusters with Banff 2019 rules

There was important overlap between the clusters and the Banff categories with an ARI of 0.48 (Table 2). 89.9% of the biopsies that were classified as No rejection according to Banff, were classified as cluster 1 by semi-supervised learning. 53 biopsies (2%) of the initially labeled as No rejection by the Banff criteria were assigned to cluster 2. A detailed examination of these reclassified cases revealed that they all shared g2 or g3 lesions scores, sometimes with one additional acute lesion of score 1. Similarly, the 4 (0.2%) No rejection cases reclassified to cluster 3 had isolated i3 lesions, and the 215 biopsies (8.1%) of No rejection cases reclassified to cluster 4 had HLA-DSA positivity with low acute lesion scores. Borderline changes according to Banff were reclassified mostly to cluster 1 and 4 (79.8% and 8.0%), although in 13.2% of cases with Borderline changes, the biopsies were reclassified to clusters with significant inflammation (cluster 2, 3 and 6). ABMR biopsies were mostly distributed between cluster 4 (45.9%) and 5 (43.4%). In cluster 4, those biopsies had no or minimal g (g0 or g1) scores while they had g2 or g3 scores in cluster 5. TCMR biopsies according to Banff were distributed in 4 different clusters, although the majority (64.6%) ended up in cluster 3. TCMR biopsies in cluster 2 (8.8%) had all g2-g3 scores with i1/t1 score whereas TCMR biopsies in cluster 3 had all i2-i3 and t2-t3 scores, mostly with g0 score. TCMR biopsies in cluster 6 (8.1%) all shared HLA-DSA positivity with medium to high t and i lesions scores. Some cases of Banff-grade rejection were assigned to cluster 1: 48 TCMR (16.8%), 8 ABMR (6.6%) and 5 mixed rejection biopsies (5.6%). None of the main driving lesions of the other clusters (HLA-DSA, g2-g3 or i2-i3/t2-t3) were present in these biopsies. Rather, they had several low score lesions (e.g. t1-i1-v1) or a combination of a low score lesion with a higher score lesion less relevant to the algorithm (e.g.

g1 with C4d3). The specific combinations of lesions yielded biopsies closer to the centroid of cluster 1 than to any other cluster.

Comparison of clustering vs. Banff classification for association with graft failure

The biopsies that were diagnosed as No rejection or Borderline changes according to Banff but reclassified to inflamed clusters 2 and 3 (N=92), had worse outcome compared to patients reclassified to non-inflamed cluster 1 (HR 1.98, CI 95% 1.18-3.34; p=0.01). Low numbers of reclassified cases with No rejection according to Banff in the HLA-DSA positive patients obviated to perform this analysis in this group. No rejection biopsies according to Banff had significantly worse graft survival when clustered to cluster 2, 3, 5, or 6 (taken together N=100) compared to non-inflamed cluster 1 or 4 (N=2886)(HR 1.99, 95% CI 1.03-3.84; p=0.04, adjusted for the presence of HLA-DSA). Restricted to the DSA-negative clusters, No rejection biopsies according to Banff had a significantly worse survival when classified in cluster 2 or 3 (n=57) vs. cluster 1 (N=2387) (HR: 1.96, 95% CI 1.01- 3.77; p= 0.04). An analogous trend was observed in the less numerous cases of Banff-defined Borderline changes, when classified in cluster 1 (n=261) vs. in clusters 2-3 (n=35) (HR = 2.052, CI 95% 0.93-4.55; p= 0.077). Similarly, rejection biopsies according to the Banff classification (ABMR, TCMR or mixed phenotypes; N=524) had significantly better survival when clustered to non-inflamed cluster 1 or 4 (N=131) compared to cluster 2, 3, 5, or 6 (N=393) (HR 0.583, 95% CI 0.385-0.881); p=0.01, adjusted for the presence of HLA-DSA). The Banff classification did not contribute to graft outcome prediction within the disease clusters. Non-inflamed biopsies (cluster 1 or 4) did not have different outcome when classified as either No rejection (N=2602) or rejection (N=135) according to the Banff classification (HR 1.11, 95% CI 0.58-2.13; p=0.75). Likewise, biopsies from clusters 2,3,5 or 6 did not have different outcome when classified as No rejection (N=57) vs. rejection (N=393) according to the Banff classification (HR 0.83, 95% CI 0.41-1.69; p=0.60).

External validation

Using the features weights and the cluster centroids obtained from the consensus clustering process, we are able to classify any new biopsy into one of the 6 previously described clusters. We applied this algorithm, starting from the lesion scores and HLA-DSA status only, without information on graft survival, to an external dataset of 3835 biopsies from Lyon University Hospital (N=1356) and the Paris Transplant Group (N=2479)(S4 Table). Note that this dataset did not include thrombi in its variables. We therefore imputed this feature from the mean value of our training data. A comparison of the final clusters proportions between the two centers is presented in S9 Fig. Similar to the training set, biopsies from the

external validation set were largely dominated by non-inflamed cluster 1. The main difference in cluster distribution was a higher proportion of cluster 4 biopsies in the external dataset compared to the Leuven dataset (26.0% vs 8.7%, p<0.0001), explained by a larger prevalence of HLA-DSA positive biopsies in the external data. Logically, the proportion of lesions within each clusters of the external validation set were very similar to the clusters obtained from the original data. There was also a similar association of the clusters with graft failure (S10 Fig).

A polar plot illustrates the full overlap in the histological presentations between the training and validation cohorts (S11 Fig). Although the proportion of biopsies performed upon indication was notably higher in the validation cohort (22.0% vs 37.7%, χ^2 test p<0.0001), the overall distribution of inflammation, estimated using the radius on the polar plot, was comparable between the training and validation datasets (S12 Fig). Comparing the clusters obtained on the validation dataset with the Banff categories, we obtained an ARI of 0.35. While maintaining a large overlap between the clustering method and the Banff classification (S5 Table), it demonstrates a higher reclassification rate in the validation dataset. We observed that biopsies that were diagnosed as No rejection or Borderline changes according to Banff but reclassified to inflamed clusters 2 and 3 (N=165), had worse outcome compared to patients reclassified to non-inflamed cluster 1 (HR 2.55, 95% CI 1.76-3.67; p<0.001). Low numbers of reclassified cases with No rejection according to Banff in the HLA-DSA-positive patients obviated to perform this analysis in this group. No rejection biopsies according to Banff had significantly worse graft survival when clustered to cluster 2, 3, 5, or 6 (taken together N=57) compared to non-inflamed cluster 1 or 4 (taken together N=2826)(HR 4.84, 95% CI 2.90-8.07; p=0<0.0001), adjusted for the presence of HLA-DSA). Rejection biopsies according to the Banff classification (ABMR, TCMR or mixed phenotypes; N=704), demonstrated better survival when clustered to non-inflamed cluster 1 or 4 (N=227) compared to cluster 2, 3, 5, or 6 (N=477) (HR 1.59, 95% CI 1.08-2.35; p=0.018, adjusted for the presence of HLA-DSA).

S1 Table. Performance indices for a range of k values used to define the optimal number of clusters in the semi-supervised clustering algorithm (N=3510 biopsies of the derivation cohort).

Performance index	**k=4**	**k=5**	**k=6**	**k=7**	**k=8**
Unsupervised clustering (unweighted lesion scores)					
Proportion of Ambiguous Clustering (PAC)	0.115	0.168	0.154	0.141	0.129
Log-rank p-value between best and 2nd best survival cluster	0.034	0.102	0.155	0.245	0.980
Minimal distance between centroids*	0.942	0.941	0.981	0.954	0.837
Adjusted Rand Index (ARI)	0.490	0.478	0.486	0.495	0.473
Semi-supervised clustering (weighted lesion scores)					
Proportion of Ambiguous Clustering (PAC)	0.112	0.152	0.083	0.133	0.144
Log-rank p-value between best and 2nd best survival cluster	0.005	0.025	0.012	0.970	0.996
Minimal distance between centroids*	0.130	0.134	0.126	0.126	0.120
Adjusted Rand Index (ARI)	0.475	0.490	0.480	0.469	0.455

**Distances are specific to the space used and should not be compared between unsupervised and semi-supervised methods.*

S2 Table. Details of cluster composition according to the individual Banff lesions scores and donor-specific HLA antibodies (N=3510 biopsies of the derivation cohort).

Banff lesion	***Lesion score***	***Cluster 1* (N=2710)**	***Cluster 2* (N=307)**	***Cluster 3* (N=231)**	***Cluster 4* (N=101)**	***Cluster 5* (N=95)**	***Cluster 6* (N=66)**
Tubulitis (t)	0	72.9%	47.5%	3.5%	73.0%	46.3%	4.5%
	1	21.4%	38.6%	17.7%	23.1%	46.3%	30.3%
	2	5.0%	12.9%	48.9%	3.6%	6.3%	42.4%
	3	0.7%	1.0%	29.9%	0.3%	1.1%	22.7%
Interstitial inflammation (i)	0	91.5%	59.4%	0.0%	90.9%	61.1%	0.0%
	1	8.3%	17.8%	2.2%	9.1%	26.3%	3.0%
	2	0.2%	8.9%	49.4%	0.0%	8.4%	40.9%
	3	0.0%	13.9%	48.5%	0.0%	4.2%	56.1%
Intimal arteritis (v)	0	96.5%	71.3%	74.0%	92.2%	67.4%	66.7%
	1	3.2%	22.8%	21.2%	7.8%	30.5%	28.8%
	2	0.1%	5.0%	3.9%	0.0%	2.1%	3.0%
	3	0.1%	1.0%	0.9%	0.0%	0.0%	1.5%
Glomerulitis (g)	0	93.4%	0.0%	81.8%	78.8%	0.0%	56.1%
	1	6.6%	0.0%	13.0%	16.3%	2.1%	28.8%
	2	0.0%	56.4%	5.2%	4.9%	27.4%	13.6%
	3	0.0%	43.6%	0.0%	0.0%	70.5%	1.5%
Peritubular capillaritis (ptc)	0	93.9%	61.4%	55.8%	79.8%	28.4%	39.4%
	1	4.2%	20.8%	24.2%	11.4%	30.5%	27.3%
	2	1.7%	16.8%	17.7%	8.5%	38.9%	28.8%
	3	0.1%	1.0%	2.2%	0.3%	2.1%	4.5%
Thrombi	0	97.8%	94.1%	97.0%	96.4%	96.8%	95.5%
	1	2.2%	5.9%	3.0%	3.6%	3.2%	4.5%
C4d deposition in peritubular capillaries (C4d)	0	89.0%	71.3%	78.4%	72.3%	51.6%	59.1%
	1	9.5%	14.9%	12.6%	13.4%	9.5%	10.6%
	2	0.5%	2.0%	1.3%	3.3%	4.2%	4.5%
	3	1.0%	11.9%	7.8%	11.1%	34.7%	25.8%
Donor-specific antibodies (DSA)	0	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%
	1	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%

S3 Table. Distribution of biopsies among clusters and association with clinical demographics, transplant and biopsy characteristics (N=3510 biopsies of the derivation cohort).

Cluster	1	2	3	4	5	6
N	2710 (77.2%)	101 (2.9%)	231 (6.6%)	307 (8.7%)	95 (2.7%)	66 (1.9%)
Donor demographics*						
Donor Type						
Donation after brain death, N (%)	626/811 (77.2%)	50/65 (76.9%)	141/179 (78.8%)	92/115 (80.0%)	44/50 (88.0%)	35/45 (77.8%)
Donation after cardiac death, N (%)	137/811 (16.9%)	12/65 (18.5%)	28/179 (15.6%)	15/115 (13.0%)	3/50 (6.0%)	8/45 (17.8%)
Living donation, N (%)	48/811 (5.9%)	3/65 (4.6%)	10/179 (5.6%)	8/115 (7.0%)	3/50 (6.0%)	2/45 (4.4%)
Age (years) Mean (SD)	47.6 (+/-14.5)	52.0 (+/-13.8)	48.5 (+/-14.1)	47.0 (+/-17.2)	47.3 (+/-16.9)	46.4 (+/-18.0)
Sex male n/N(%)	434/811 (53.5%)	33/65 (50.8%)	84/179 (46.9%)	57/115 (49.6%)	28/50 (56.0%)	22/45 (48.9%)
Diabetes n/N(%)	21/639 (3.3%)	1/55 (1.8%)	4/137 (2.9%)	2/83 (2.4%)	0/34 (0.0%)	2/31 (6.5%)
Recipient demographics*						
Rec age (years) Mean (SD)	53.4 (+/-13.2)	56.0 (+/-11.9)	52.2 (+/-14.8)	52.4 (+/-14.4)	53.4 (+/-13.6)	48.0 (+/-17.1)
Sex male n/N(%)	515/811 (63.5%)	42/65 (64.6%)	112/179 (63.1%)	52/115 (45.2%)	22/50 (44.0%)	24/45 (53.3%)
BMI rec. Mean(SD)	25.3 (+/-4.39)	26.8 (+/-3.89)	25.3 (+/-4.34)	25.3 (+/-5.04)	25.6 (+/-5.16)	26.1 (+/-5.90)
Pretransplant-DSA n/N(%)	0/811 (0.0%)	0/65 (0.0%)	0/179 (0.0%)	90/115 (78.3%)	46/50 (92.0%)	33/45 (73.3%)
Graft *						
Repeat TX n/N(%)	81/811 (10.0%)	6/65 (9.2%)	14/179 (8.4%)	49/115 (42.6%)	26/50 (52.0%)	20/45 (44.4%)
Total Mismatch Mean(SD)	2.72 (+/-1.33)	3.26 (+/-1.21)	2.83 (+/-1.42)	2.90 (+/-1.25)	2.68 (+/-1.11)	2.91 (+/-1.12)
CIT (hours) Mean(SD)	14.2 (+/-5.7)	14.9 (+/-6.1)	14.1 (+/-5.3)	14.3 (+/-5.8)	14.5 (+/-4.7)	14.3 (+/-4.6)
Biopsy *						
Protocol biopsy	2269 (83.7%)	57 (56.4%)	115 (49.8%)	218 (71.0%)	53 (55.8%)	25 (37.9%)
Indication biopsy	441 (16.3%)	44 (43.6%)	116 (50.2%)	89 (29.0%)	42 (44.2%)	41 (62.1%)
Proteinuria Mean(SD)	0.26 (+/-0.54)	0.47 (+/-0.62)	0.41 (+/-0.65)	0.29 (+/-0.60)	0.44 (+/-0.96)	0.68 (+/-0.94)
Median(IQR)	0.145 (0.1-0.23)	0.22 (0.12-0.51)	0.24 (0.14-0.43)	0.15 (0.10-0.23)	0.19 (0.13-0.41)	0.39 (0.19-0.61)
eGFR Mean(SD)	44.1 (+/-18.4)	34.7 (+/-21.6)	33.4 (+/-19.4)	40.1 (+/-22.5)	35.1 (+/-21.7)	28.2 (+/-20.4)
Median(IQR)	44.0 (32.2-55.6)	32.09 (17.1-50.8)	31.4 (19.86-44.8)	37.3 (24.5-52.9)	31.8 (19.2-48.0)	20.5 (14.7-36.3)
Days since transplantation	533 (+/-541)	368 (+/-428)	389 (+/-507)	483 (+/-524)	364 (+/-455)	519 (+/-555)

Mean(SD)						
Median(IQR)	366 (92-744)	111 (14-733)	104 (13-729)	363 (89-740)	103 (17-666)	368 (17-790)

*when two or more biopsies from the same patient are classified in the same cluster, only the first biopsy is taken into account for graft-related statistics.

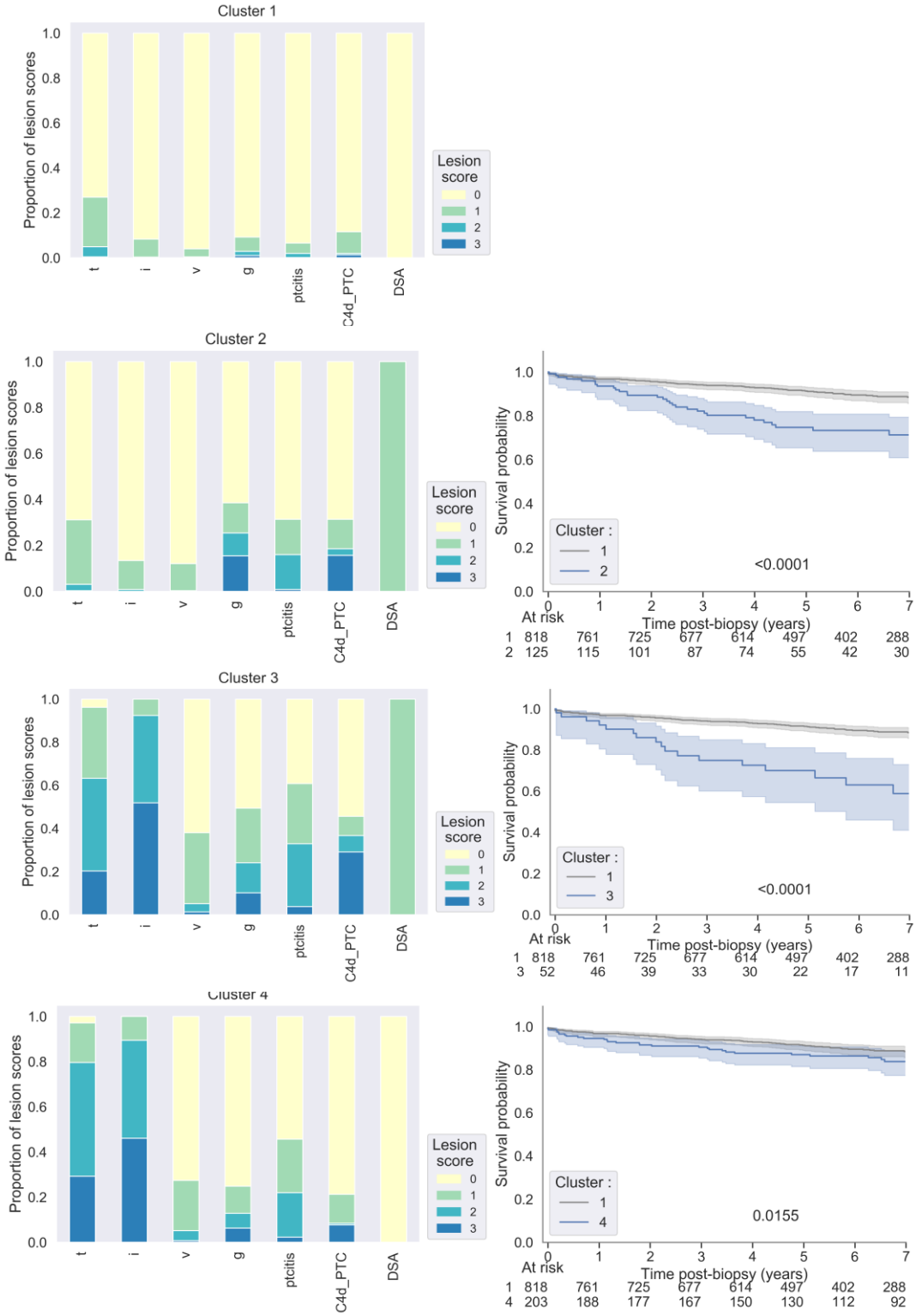
S4 Table. Demographic, clinical and histological characteristics of the patients and biopsies included in the validation dataset.

Cohort characteristics	
Total	N=1989
Donor type (%)	
Donation after Brain Death	1469 (73.8)
Donation after Cardiac Death	101 (21.1)
Living Donation	419 (5.1)
Age in years– mean (std)	53.3 (16.8)
Sex - Male (%)	1048 (52.7)
Recipient demographics	
Age in years – mean (std)	51.0 (14.5)
Sex – male (%)	1241 (62.4)
BMI (kg/m2) – mean (std)	24.4 (4.3)
Repeat transplantation (%)	271 (13.6)
CIT (h) –mean (std)	14.7 (10.5)
Mismatch total (A+B+DR) mean (std)	3.4 (1.4)
Biopsy characteristics	
Total	N=3835
Banff 2019 diagnosis	
No rejection, N (%)	2883 (75.2)
Borderline changes, N (%)	248 (6.5)
TCMR, N (%)	194 (5.1)
ABMR, N (%)	416 (10.8)
Mixed rejection (ABMR + TCMR), N (%)	33 (0.9)
Mixed borderline rejection (ABMR + borderline changes), N (%)	61 (1.6 %)
Indication biopsies, N (%)	N=1442 (37.7)
Days since transplantation, median (interquartile range)	190 (45-655)
eGFR at day of biopsy, mean (std)	36.5 (18.4)
Protocol biopsies, N (%)	N=2393 (62.3)
Days since transplant –median (interquartile range)	180 (94-375)
eGFR at day of biopsy –mean (std)	54.7 (22.4)

S5 Table. Contingency tables comparing the Banff 2019 diagnosis and the 6 clusters obtained on the external validation dataset. Proportions represent the distribution in the clusters per Banff category.

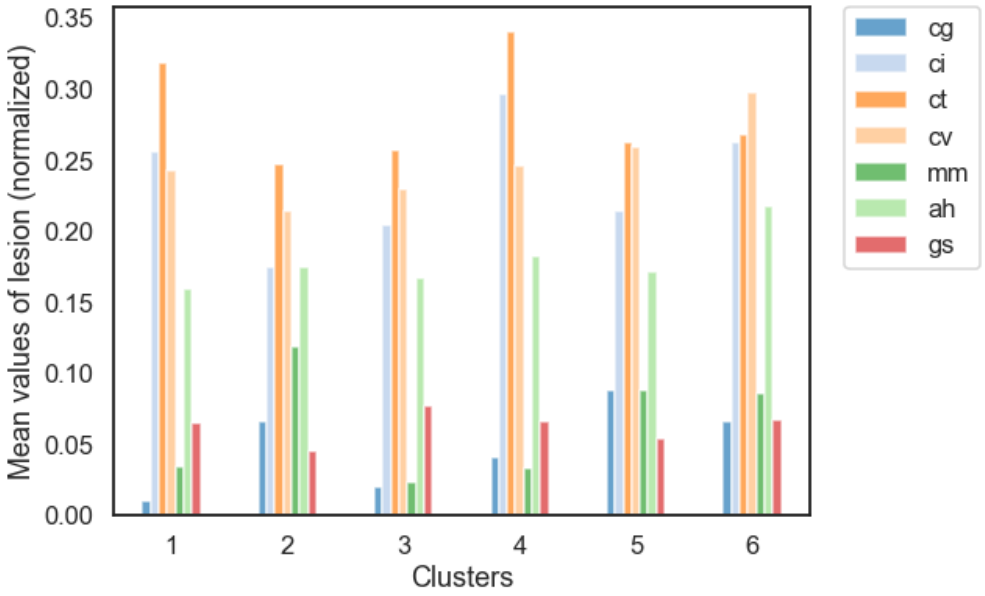
CLUSTER	n	1	2	3	4	5	6
No rejection	2883	2003 (69.5%)	54 (1.9%)	3 (0.1%)	823 (28.5%)	0 (0.0%)	0 (0.0%)
Borderline	248	106 (42.7%)	24 (9.7%)	84 (33.9%)	16 (6.5%)	0 (0.0%)	18 (7.3%)
TCMR	194	7 (3.6%)	10 (5.2%)	158 (81.4%)	2 (1.0%)	0 (0.0%)	17 (8.8%)
ABMR	416	63 (15.1%)	40 (9.6%)	0 (0.0%)	152 (36.5%)	158 (38.0%)	3 (0.7%)
Mixed borderline rejection	33	0 (0.0%)	3 (9.1%)	10 (30.3%)	2 (6.1%)	9 (27.3%)	9 (27.3%)
Mixed rejection	61	0 (0.0%)	8 (13.1%)	20 (32.8%)	1 (1.6%)	6 (9.8%)	26 (42.6%
TOTAL	**3835**	**2179 (56.8%)**	**139 (3.6%)**	**275 (7.2%)**	**996 (26.0%)**	**173 (4.5%)**	**73 (1.9%)**

S1 Fig. Distribution of the individual acute lesion scores in the clusters using an unweighted approach, and post-biopsy Kaplan-Meier graft survival curves relative to cluster 1 of the derivation cohort (N=3510 biopsies). Compared to cluster 1 (essentially normal biopsies), the three other clusters associated significantly with impaired graft survival. However, their histological and clinical relevance was more ambiguous. P-values refer to HR from the Cox models.



S2 Fig. Distribution of chronic lesions in the 6 acute lesion clusters (N=3510 biopsies).

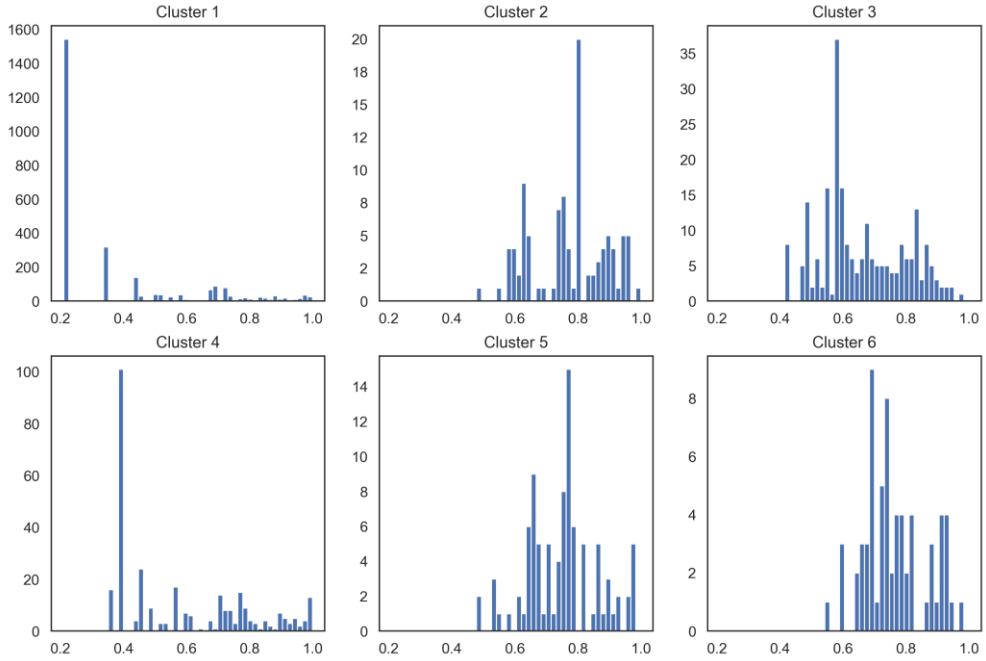
Chronic lesions shared similar profiles across the 6 clusters based on acute lesion score, translating presumably independent mechanisms. cg=transplant glomerulopathy, ci=interstitial fibrosis, ct=tubular atrophy, cv=vascular intimal thickening, mm=mesangial matrix increase, ah=arteriolar hyalinosis, gs=glomerulosclerosis



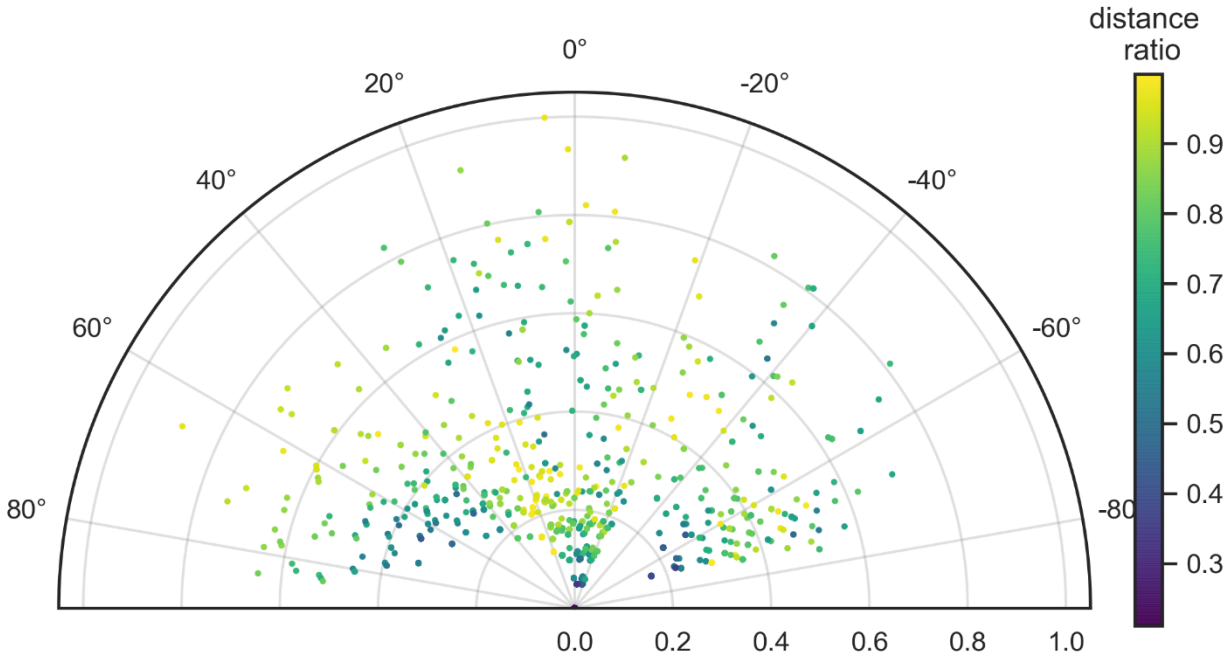
S3 Fig. Relative distances to the closest clusters' boundary

A. Distribution of the relative distance from the centroids of the biopsies included in each cluster. **B**. Visualization of the relative distances to the cluster boundaries, plotted on the polar plot. A relative distance close to one indicates an almost equidistant position from the two closest cluster centroids. Conversely, a biopsy that is much closer to a centroid than it is from the second closest centroid has a lower relative distance.

A

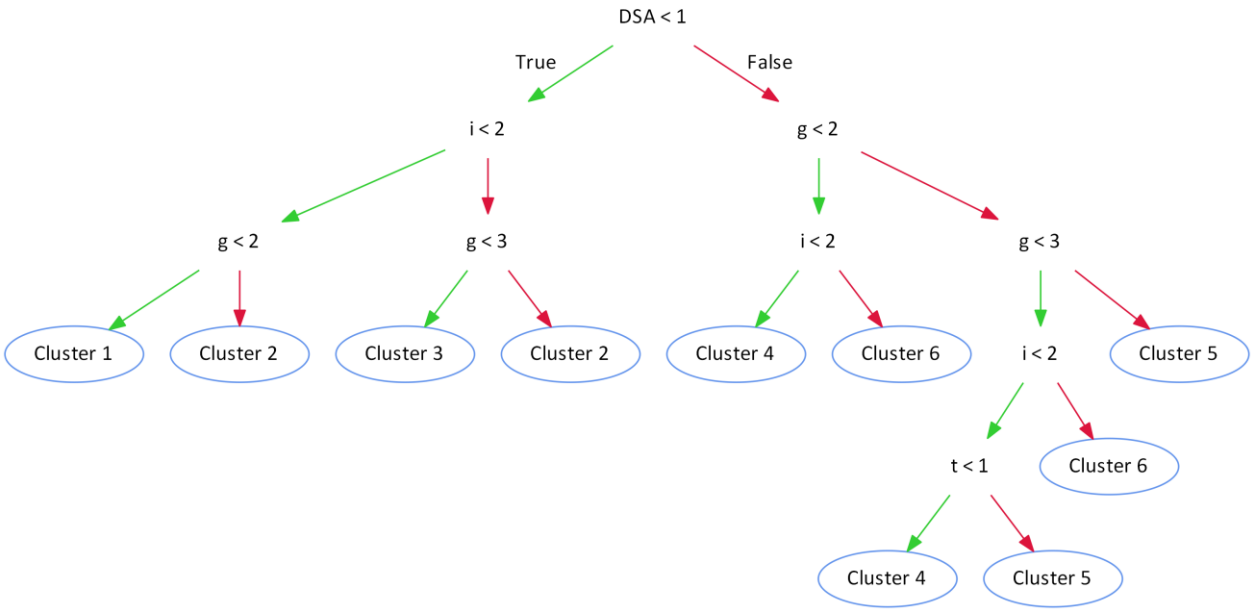


B



S4 Fig. Decision tree of the clustering process

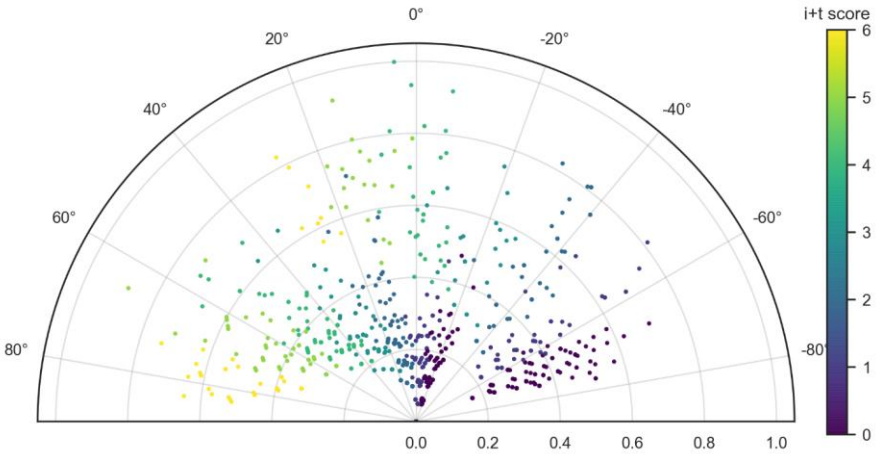
This decision tree was constructed with a minimum sample size in leaf nodes of 10 biopsies. It demonstrates that using only 4 parameters - DSA, i, t, g – is sufficient to be assigned in the correct cluster with 97% of balanced accuracy.



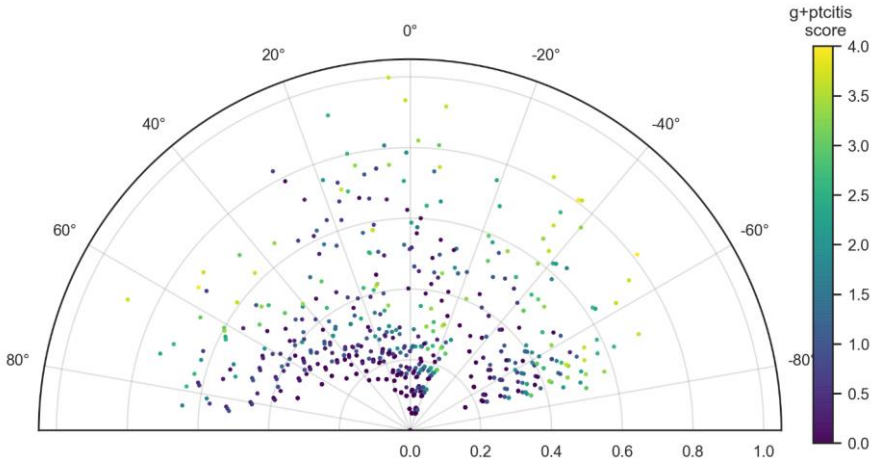
S5 Fig. Various combinations of lesions scores displayed on the polar plots.

The sum of lesions scores is represented by the color of the points. **A** Biopsies with a high combined score of t and i are predominant on the extreme left ("Pure TCMR") and on middle of the polar plot (mixed rejection). **B** Biopsies with a high combined score of g and ptc are predominant on the right side of the plot ("pure ABMR") and on the middle of the plot ("mixed rejection").

A

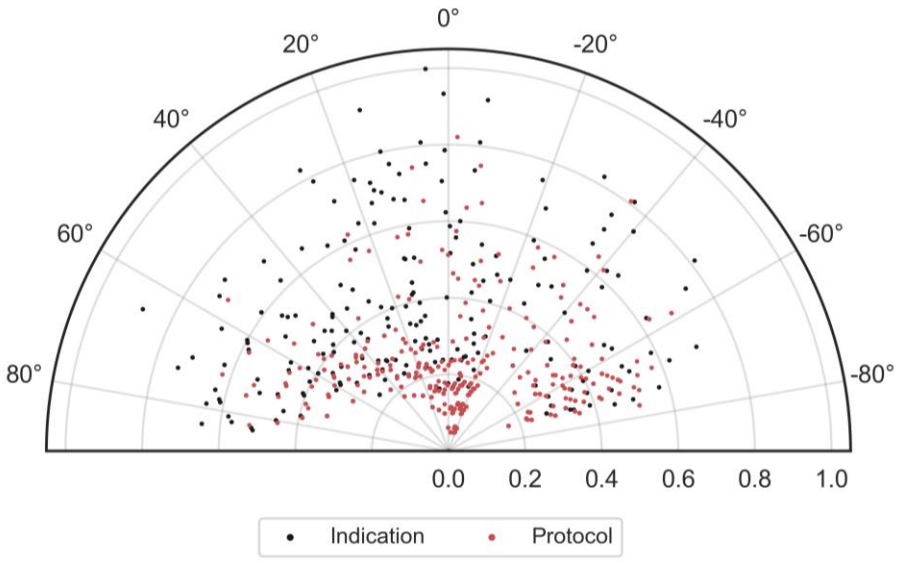


B



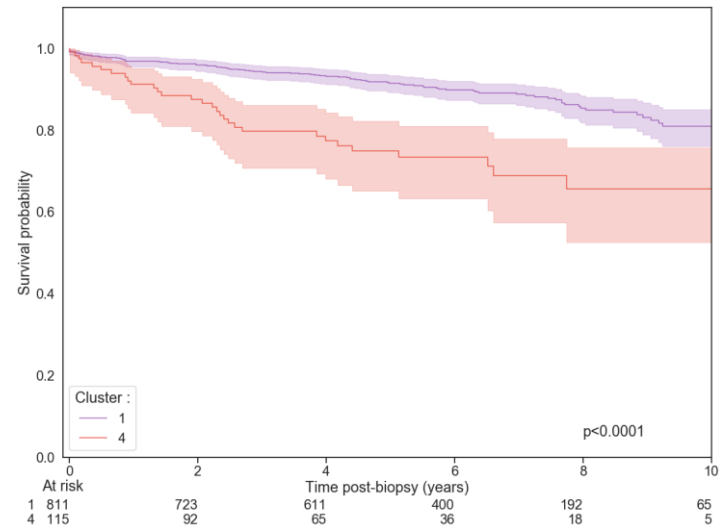
S6 Fig. Comparison of indication vs. protocol biopsies, as superposed on the polar plot.

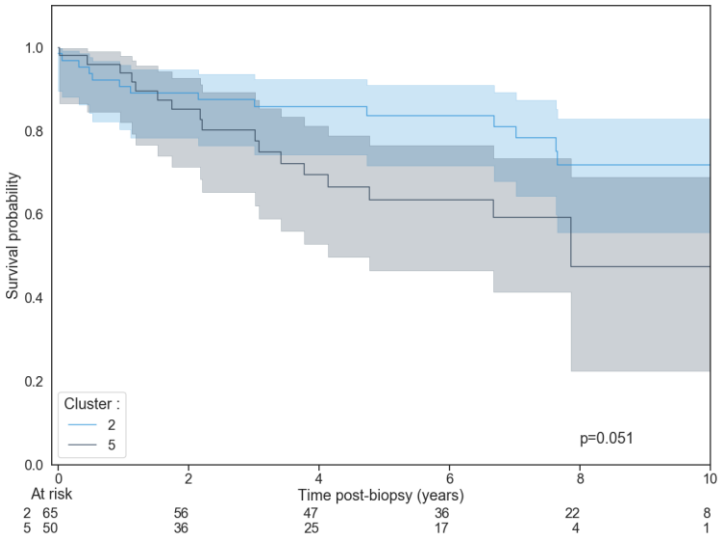
The radius from indication biopsies was on average higher compared to protocol biopsies (mean ± sd : 0.22 ± 0.23 vs. 0.08 ± 0.13 respectively, student t test $p<0.0001$), illustrating more inflamed biopsies at time of graft dysfunction than at time of stable graft function.

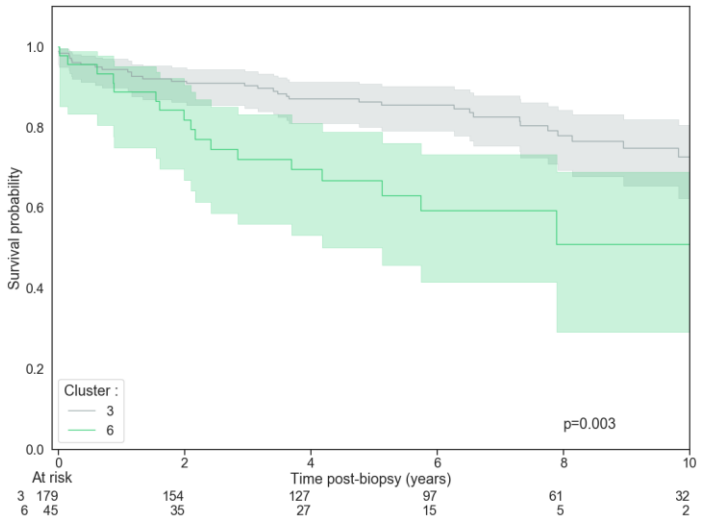


S7 Fig. Post-biopsy graft survival in the three DSA-/DSA+ pair of clusters.

A. Cluster 1 vs. cluster 4; **B** Cluster 2 vs. cluster 5; **C.** Cluster 3 vs. cluster 6). These plots demonstrate the negative impact of the presence of DSA on graft survival. P-values refer to HR from Cox models.

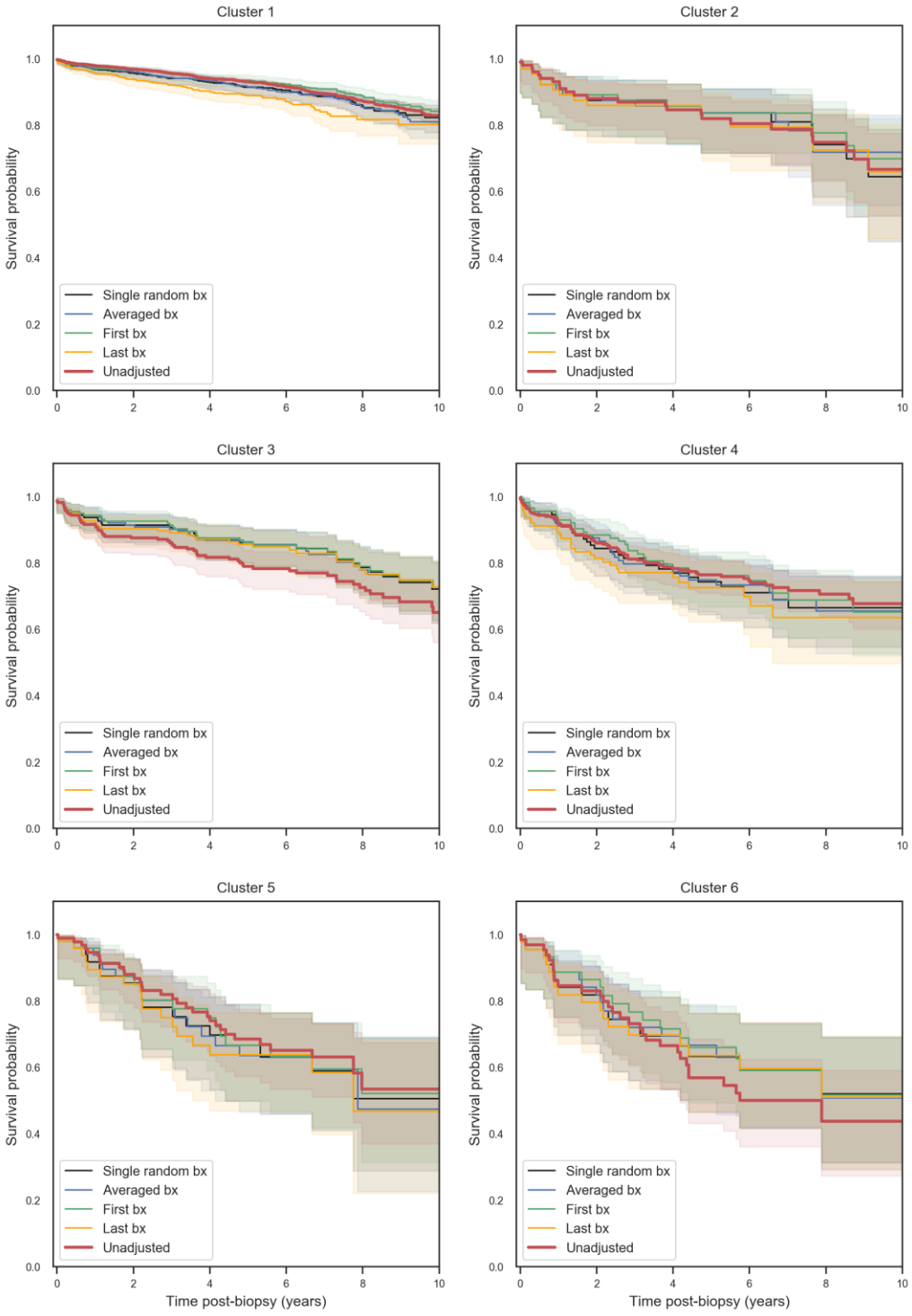






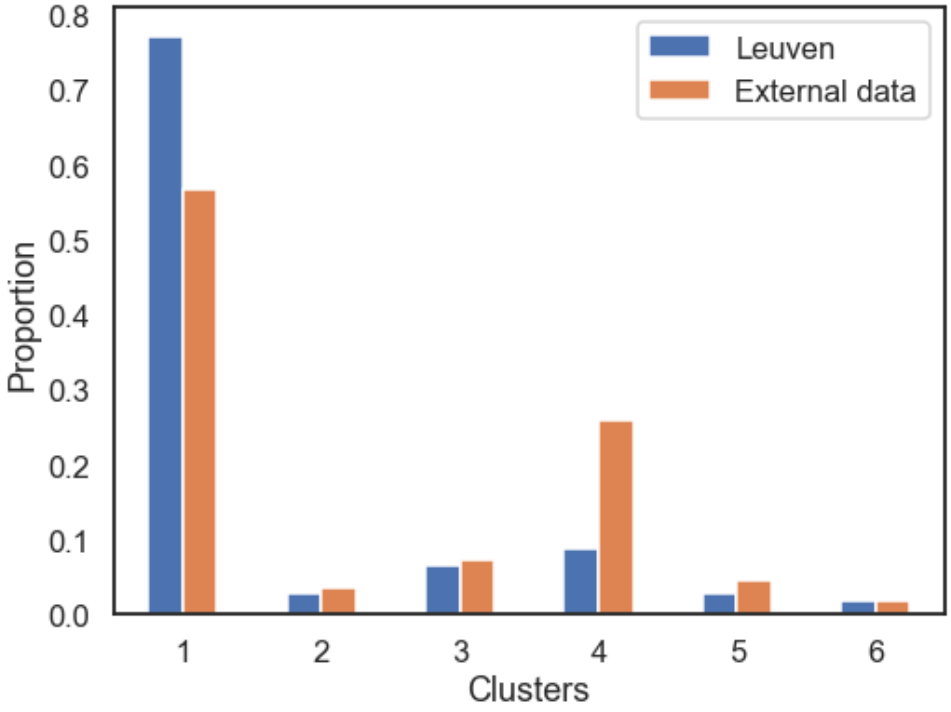
S8 Fig. Post-biopsy graft survival in the six clusters, according to the adjustment method for repeated biopsies per patient.

The following Kaplan-Meyer curves demonstrate various adjustment method for repeated biopsies per patient in Survival times from repeated biopsies were either averaged for each patient, based on a single random biopsy, using the first or last biopsy per patient. bx=biopsy.

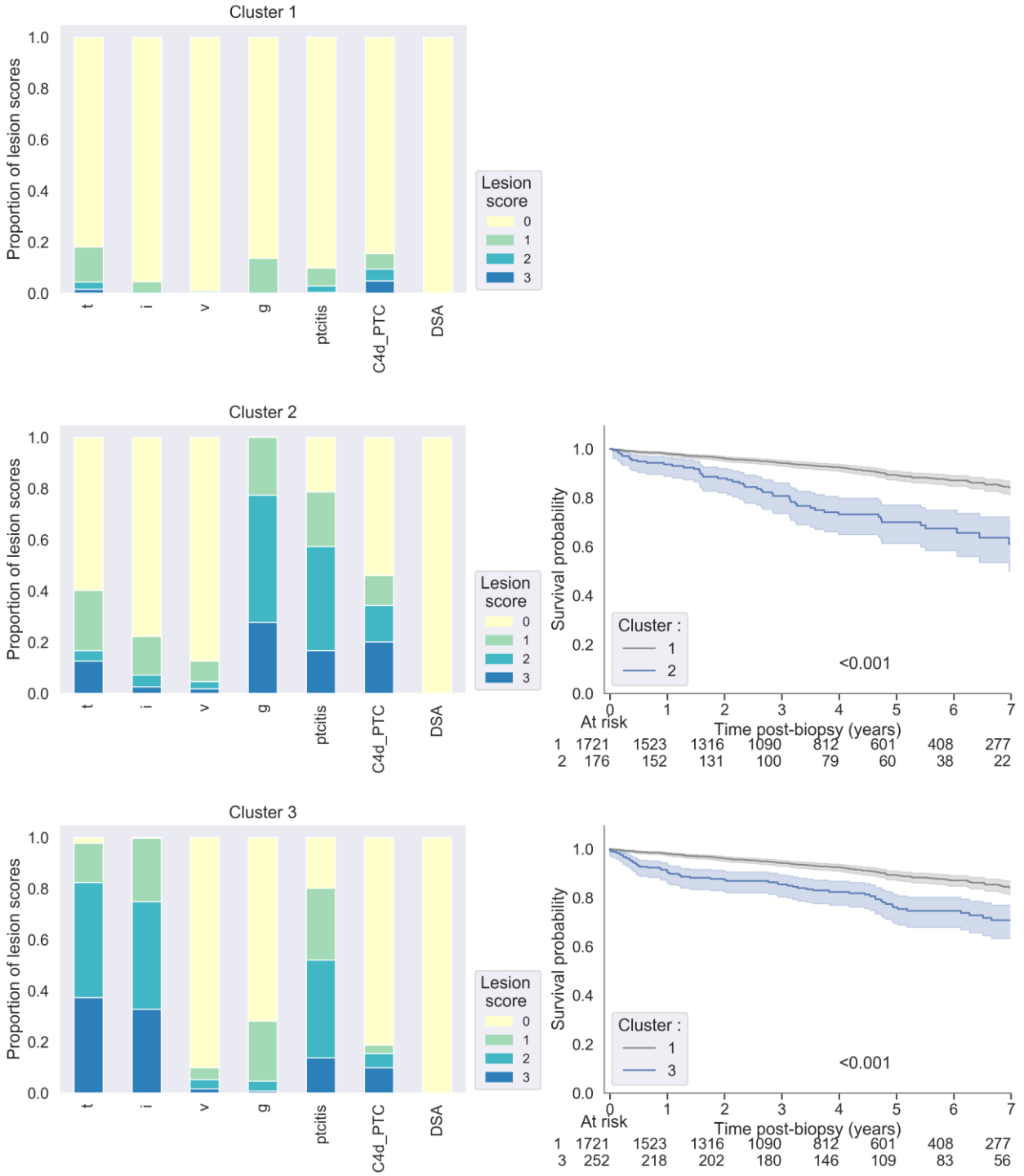


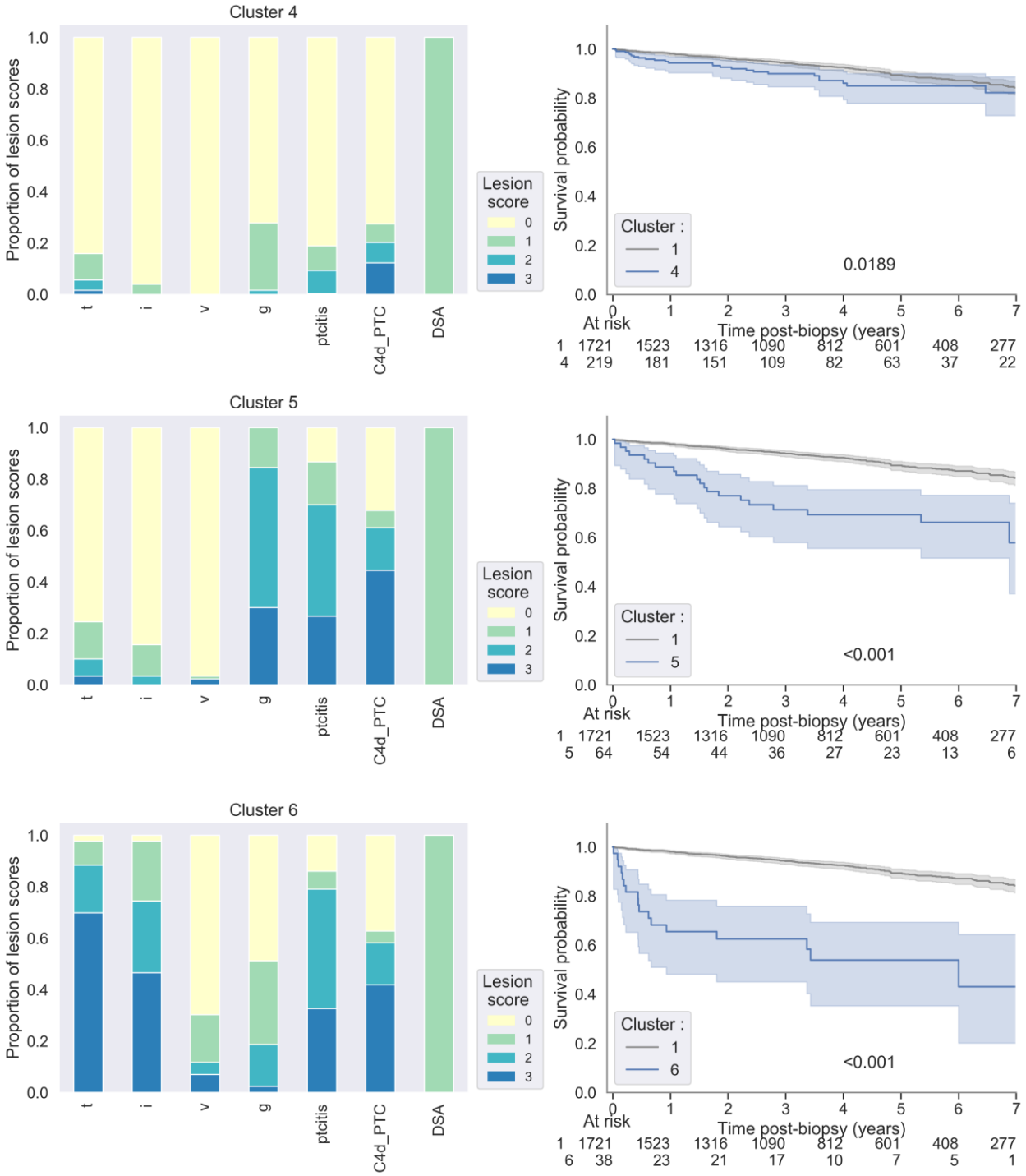
S9 Fig. Comparison of cluster proportion per center

N=3510 biopsies from the derivation cohort (Leuven) and 3835 biopsies from the external validation cohort. Non inflamed clusters 1 and 4 were the dominant clusters in both datasets. The larger prevalence of HLA-DSA positive biopsies in the external data is reflected by a higher proportion of cluster 4 biopsies in the external dataset compared to the Leuven dataset (26.0% vs 8.7%, p<0.0001). Other clusters have similar proportion in both datasets.



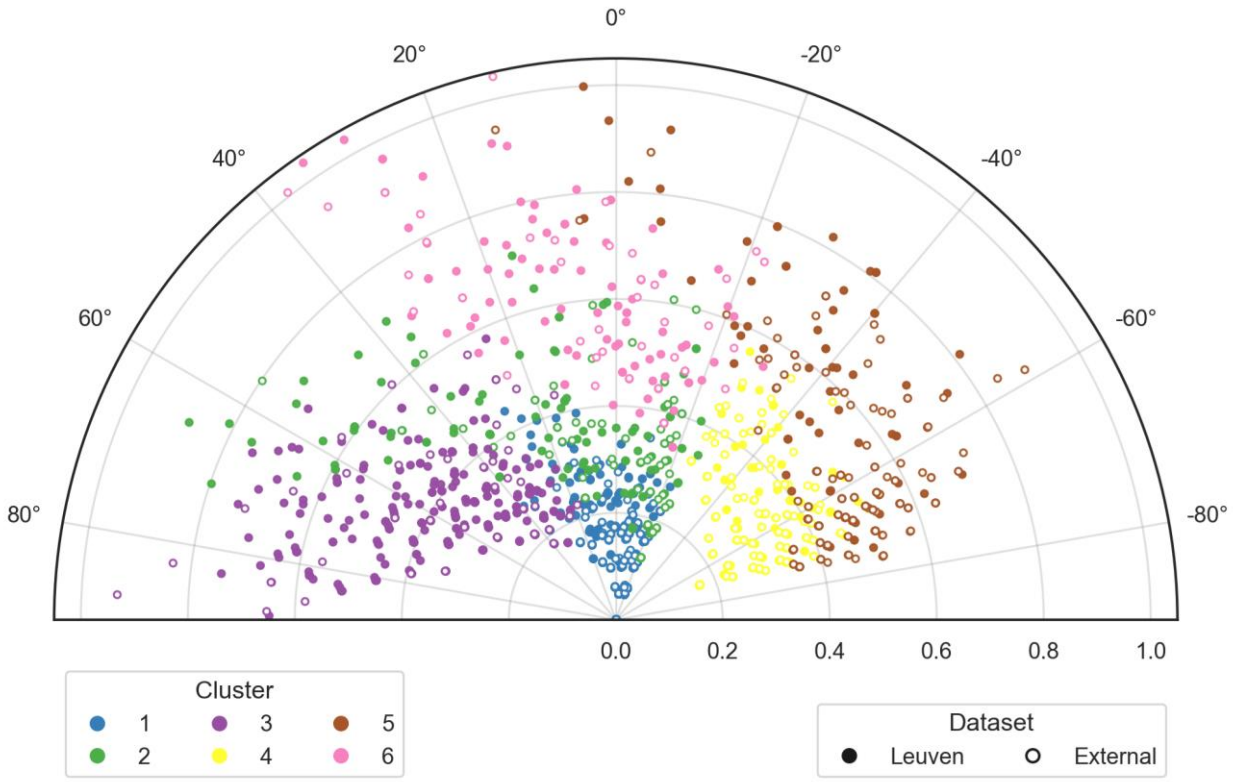
S10 Fig. Distribution of the individual acute lesion scores in the different clusters, and post-biopsy Kaplan-Meier graft survival curves relative to cluster 1 of the external validation cohort (N=3835 biopsies).




Cluster 4
Proportion of lesion scores
1.0
0.8
0.6
0.4
0.2
0.0
t
i
v
g
ptcitis
C4d_PTC
DSA
Lesion score
0
1
2
3
Survival probability
Cluster :
1
4
0.0189
At risk
Time post-biopsy (years)
1 1721 1523 1316 1090 812 601 408 277
4 219 181 151 109 82 63 37 22
Cluster 5
Proportion of lesion scores
Lesion score
Survival probability
Cluster :
1
5
<0.001
At risk
Time post-biopsy (years)
1 1721 1523 1316 1090 812 601 408 277
5 64 54 44 36 27 23 13 6
Cluster 6
Proportion of lesion scores
Lesion score
Survival probability
Cluster :
1
6
<0.001
At risk
Time post-biopsy (years)
1 1721 1523 1316 1090 812 601 408 277
6 38 23 21 17 10 7 5 1

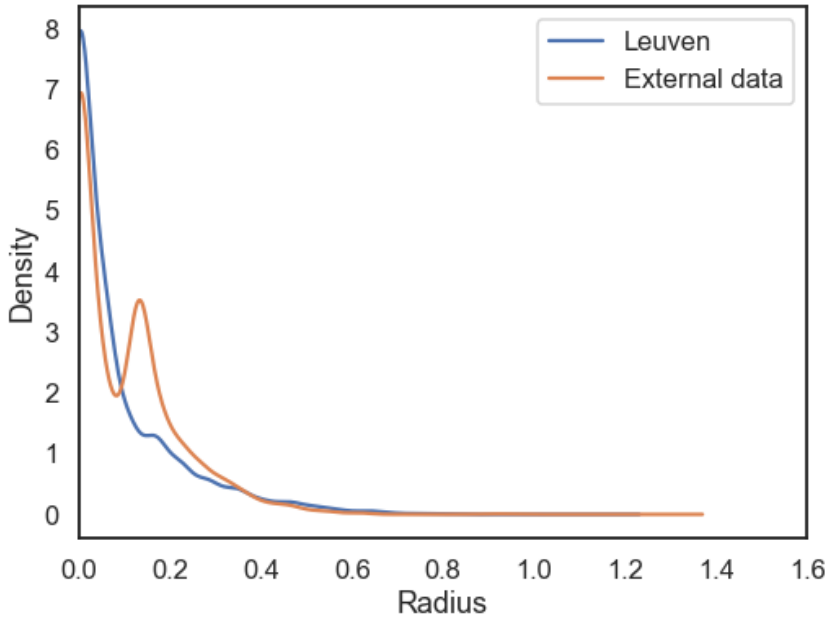
S11 Fig: Overlay of the data from Leuven and the external dataset in the polar plot, according to the six clusters identified in the derivation cohort.

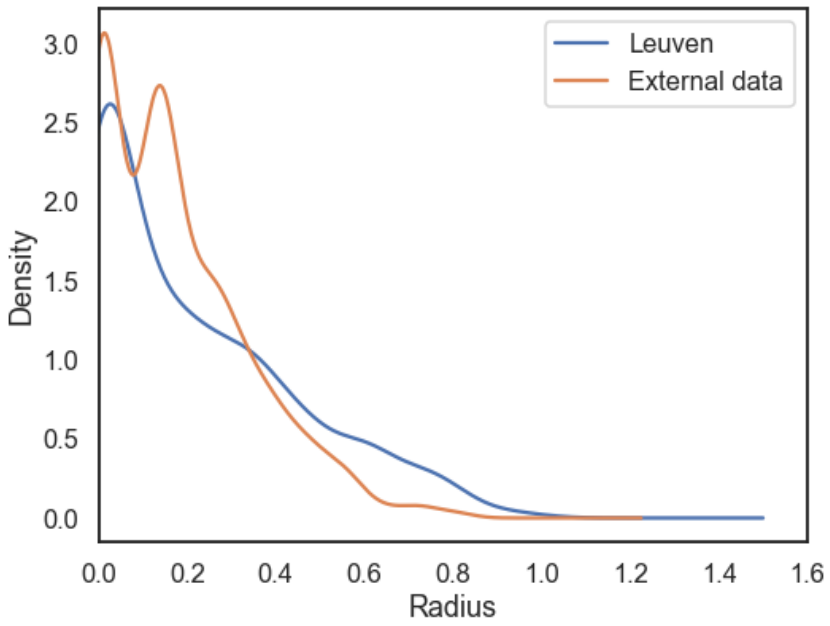
This plot demonstrates the full overlap in the histological presentations between the derivation and validation cohorts.



S12 Fig. Distribution of the radius from the polar plot

A. Distribution of the radius on the polar plot for protocol biopsies, according to center (N=2737 biopsies from Leuven; N=2393 biopsies from the external validation cohort). **B.** Distribution of the radius on the polar plot for indication biopsies, according to center (N=773 biopsies from Leuven; N=1442 biopsies from the external validation cohort). Note the extra peak in the external data corresponding to the larger proportion of DSA positive biopsies without inflammation.





S13 Fig. Association with graft survival in the polar plot visualization tool based on the validation data

A. Association of the polar plot radius with graft survival in the external validation cohort. We stratified the 3835 biopsies along the radius axis in five subgroups and plotted the corresponding Kaplan-Meier survival curves. **B.** Estimated survival probability at 5 years post-biopsy, calculated from the nearest neighborhood with k=40.