American Journal of Transplantation 2010; 10: 2066–2073
Wiley Periodicals Inc.

doi: 10.1111/j.1600-6143.2010.03240.x

Inflammation in Areas of Tubular Atrophy in Kidney Allograft Biopsies: A Potent Predictor of Allograft Failure

R. B. Mannon[a,*], A. J. Matas[b], J. Grande[c], R. Leduc[d], J. Connett[d], B. Kasiske[e], J. M. Cecka[f], R. S. Gaston[a], F. Cosio[g], S. Gourishankar[h], P. F. Halloran[h], L. Hunsicker[i] and D. Rush[j]; for the DeKAF Investigators

[a] *Division of Nephrology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL*
[b] *Department of Surgery, University of Minnesota, Minneapolis, MN*
[c] *Department of Pathology, Mayo Clinic, Rochester, MN*
[d] *Department of Biostatistics, University of Minnesota, Minneapolis, MN*
[e] *Division of Nephrology, Hennepin County Medical Center, Minneapolis, MN*
[f] *UCLA Immunogenetics Center, Los Angeles, CA*
[g] *Division of Nephrology, Mayo Clinic, Rochester, MN*
[h] *Division of Nephrology, University of Alberta, Edmonton, Alberta, Canada*
[i] *Division of Nephrology, University of Iowa, Iowa City, IA*
[j] *Department of Nephrology, University of Manitoba, Winnipeg, Manitoba, Canada*
**Corresponding author: Roslyn B. Mannon, rmannon@uab.edu*

The Banff scoring schema provides a common ground to analyze kidney transplant biopsies. Interstitial inflammation (*i*) and tubulitis (*t*) in areas of viable tissue are features in scoring acute rejection, but are excluded in areas of tubular atrophy (TA). We studied inflammation and tubulitis in a cohort of kidney transplant recipients undergoing allograft biopsy for new-onset late graft dysfunction (N = 337). We found inflammation ('*iatr*') and tubulitis ('*tatr*') in regions of fibrosis and atrophy to be strongly correlated with each other ($p < 0.0001$). Moreover, *iatr* was strongly associated with death-censored graft failure when compared to recipients whose biopsies had no inflammation, even after adjusting for the presence of interstitial fibrosis (Hazard Ratio = 2.31, [1.10–4.83]; p = 0.0262) or TA (hazard ratio = 2.42, [1.16–5.08]; p = 0.191), serum creatinine at the time of biopsy, time to biopsy and *i* score. Further, these results did not qualitatively change after additional adjustments for C4d staining or donor specific antibody. Stepwise regression identified the most significant markers of graft failure which include *iatr* score. We propose that a more global assessment of inflammation in kidney allograft biopsies to include inflammation in atrophic areas may provide better prognostic information. Phenotypic characterization of these inflammatory cells and appropriate treatment may ameliorate late allograft failure.

Key words: Banff schema, biopsy, inflammation, fibrosis, graft failure, injury

Received 22 January 2010, revised 22 June 2010 and accepted for publication 01 July 2010

Introduction

The classification of kidney allograft pathology by the Banff criteria provides reproducible diagnostic categories of allograft injury (1). Common histological features of failing allografts are interstitial fibrosis (IF) and tubular atrophy (TA), the severity of which is graded semi-quantitatively as mild, moderate and severe (2). Recent studies have demonstrated that the simple quantification of IF/TA is insufficient to identify those at greatest risk for long-term graft loss (3). Protocol biopsies, including those from living donors, at 1 year posttransplant nearly uniformly show IF/TA. In these studies, IF/TA was associated with a small decrease in allograft survival compared to normal histology; however, the combination of IF/TA with inflammation of any degree was associated with a worse prognosis than fibrosis alone (4–6).

In the current Banff schema, interstitial inflammation and tubulitis are scored only in areas of nonfibrotic interstitium and nonatrophic tubules, respectively. Subcapsular inflammation is also excluded. However, a protocol biopsy study performed in recipients of kidney–pancreas transplants showed that inflammation in areas of IF and TA (referred to as the 'chronic damage index' or 'cdi') predicted the development of progressive tubulointerstitial injury in sequential biopsies (7), and although its association with allograft failure was not determined, the finding of inflammation in areas of chronic injury appeared to indicate 'active'–i.e. progressive fibrosis. At the 2007 meeting of the Banff allograft pathology group, future study of the association between a total inflammation score ('*total i*') including all areas of the renal parenchyma, and allograft survival, was

proposed (8). Accordingly, Mengel et al. have since noted the importance of infiltrates in areas of fibrosis and reported that *total i* correlates better than the *i* score with subsequent graft deterioration (9,10). Thus, understanding the role of ongoing inflammatory injury, both in areas of preserved architecture as well as areas of chronic injury is critically important to the prognosis and management of the failing kidney grafts.

The Long-Term Deterioration of Kidney Allograft Function (DeKAF) study is a multicenter study designed to identify the causes of late allograft dysfunction (11). To date, 337 renal transplant recipients with new onset, late graft dysfunction have undergone allograft biopsies that were read, using standard Banff criteria, by a central pathologist. Additionally, semi-quantitative scoring of inflammatory cell infiltrates ('*iatr*') and tubulitis ('*tatr*') in areas of TA were obtained. We found that *iatr* was frequently present in this cohort and that it was strongly associated with time to death-censored graft failure even after adjustment for serum creatinine, Banff *i* score, and extent of IF. These results support a more comprehensive assessment of inflammatory cell infiltrates in kidney allografts than described in the current Banff system.

Methods

Patients and enrollment

The DeKAF study consists of two cohorts of kidney transplant recipients enrolled at 7 transplant centers in the US and Canada: (1) a cross-sectional cohort transplanted prior to October, 2005 and developing new onset late graft dysfunction and (2) a prospective cohort transplanted on or after January 1, 2006 (11). The study is registered at www.clinicaltrials.gov. Institutional Review Board approval was obtained at all participating sites.

For the current analysis, we studied biopsies done for new onset late graft function in the cross-sectional cohort. Recipients were eligible for enrollment if transplanted prior to October 1, 2005, having a baseline serum creatinine <2.0 mg/dL as of January 1, 2006, and subsequently developing a $\geq$25% increase in serum creatinine, or new onset proteinuria [albumin/Cr ratio >0.2 or protein/Cr ratio >0.5]) leading to an allograft biopsy. Enrollment occurred at the time of the biopsy.

Histological analysis

Allograft biopsies were read by the local pathologist and pathologic diagnosis was used to guide clinical care and immunosuppressive management per local protocols using Banff 1997 criteria (2) and the updated criteria additions of 2007 (8). Representative sections (H&E, silver, PAS, trichrome stains, and 11 unstained sections for additional studies) were submitted to a central laboratory where all biopsies were interpreted by the same pathologist in a masked fashion (N = 337; JG).

Interstitial inflammation and tubulitis were scored separately in nonatrophic and atrophic regions of the renal cortex. Inflammation and tubulitis in nonatrophic regions of the cortex was scored according to the 'standard' Banff classification scheme (2) for assessment of '*i*' and '*t*' scores, respectively. Inflammation in areas of atrophy—currently ignored in the 'standard' Banff classification scheme—was assessed as the percentage of atrophic cortex with inflammatory infiltrates ('*iatr*'): 0 = inflammation in less than 10% of atrophic regions; 1 = inflammation in 10–25% of atrophic regions; 2 = inflammation in 26–50% of atrophic regions; and 3 = inflammation in $>$50% of atrophic regions. Similarly, tubulitis in atrophic tubules ('*tatr*') was assessed in the same manner as for nonatrophic ones (0 = no mononuclear cells in tubules; 1 = foci with 1–4 cells/tubular cross section; 2 = foci with 5–10 cells/tubular cross section; and 3 = foci with $>$10 cells/tubular cross section). Illustrative examples of inflammation and tubulitis in regions of atrophy are shown in Figure 1. Total *i* score, as defined by the proportion of total cortical surface area involved by inflammation, whether atrophic or nonatrophic, was assessed as previously described by Mengel (9).

C4d staining was performed using standard immunohistochemical methods. In brief, antigen recovery was carried out by heat treatment in EDTA for 30 min using a vegetable steamer. Endogenous biotin in the kidney was blocked by treating with 3% H_2O_2, followed by the Vector Avidin/Biotin Blocking Kit (Vector Laboratories, Burlingame, CA). Antihuman C4d antibody (C4d pAb; Alpco Diagnostics, Salem, NH) was applied for 30 min, followed by rabbit EnVision+ HRP (Dako, Carpinteria, CA) for 30 min. NovaRED (Vector Labs) was used for color development, followed by hematoxylin staining. To facilitate consistency, slides were batched and stained on a Dako autostainer. C4d stains were read in a masked fashion, without clinical or pathologic information. The estimated percentage of peritubular capillaries staining positively for C4d was recorded as negative, $\geq$10%, $\geq$25%, or $\geq$50%, using the Banff classification scheme (12) and as described by Crary et al. (submitted).

Donor specific antibody testing

Serum samples (2 mL) were collected at the time of each biopsy and tested for antidonor HLA antibodies (DSA) by a central laboratory (JMC) using

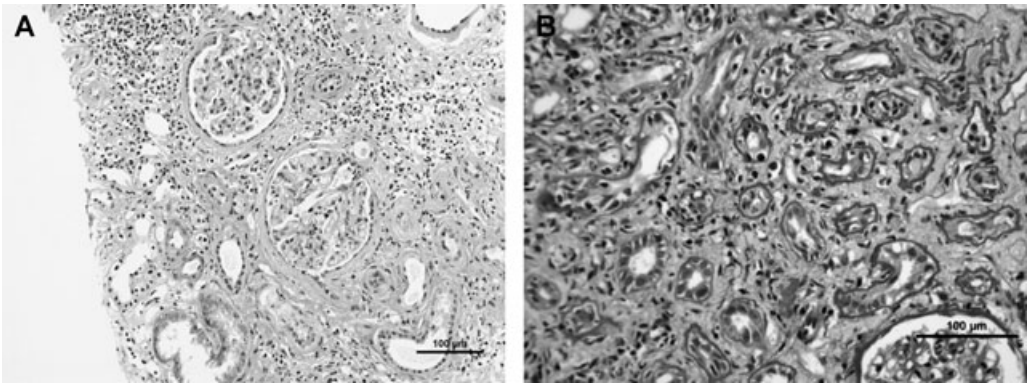

Figure 1. Representative photomicrographs showing (A) inflammation in region of atrophy and (B) tubulitis in atrophic tubules.

microparticles with individual purified HLA antigens (Ag) covalently bound as targets (One Lambda, Inc, Canoga Park, CA) on the Luminex platform. DSA was considered positive (+) if antibodies were detected against donor HLA-A, B, DR or DQ. DSA was negative (−) if no DSA was present as well as no antibody against HLA-Cw or DP (as donor typing at these loci was not reported).

Clinical data management

For all enrolled subjects data were collected every 6 months and at the time of allograft loss. These data included: serum creatinine, urine protein and creatinine, demographics, current immunosuppressive medications, other intervening illnesses, date and cause of graft loss (return to dialysis, retransplant or recipient death).

Statistical analysis

Associations of *iatr*, *tatr*, *i* and *total i* were analyzed by Chi-square testing. Tests of association between diagnoses of acute rejection and *iatr* were assessed by Fisher's Exact Test. Strength of association between categorical variables was measured by Kendall's Tau-b coefficient. Analysis of variance was used to compare mean time from transplant to biopsy by level of *iatr*. Nonparametric methods (Kaplan–Meier graphs, log-rank test) were used to analyze time to death-censored graft failure for biopsies with and without inflammation, in areas of fibrosis and with and without tubulitis, in regions of atrophy. To adjust for center specific effects, log-rank tests to compare these survival curves were also stratified by clinical center. To estimate relative hazards and adjust for the presence of other covariates, Cox proportional hazard models were developed for time to death-censored graft failure stratified by clinical center, and adjusting for time from transplant to biopsy, serum creatinine, inflammation (Banff *i* score), total *i* score, TA (Banff ct score), allograft fibrosis (Banff *ci* score), C4d staining, donor specific antibody and treatment for acute rejection. Stepwise regression methods were used to select an optimal subset of these covariates with an entry-criterion for variable selection of a $p < 0.15$ and a retention criterion of a $p < 0.10$. Because not all subjects had both C4d and donor specific antibody available for analysis, donor specific antibody was not included in the stepwise regression procedures for considerations of sample size.

Results

The cross-sectional cohort began enrollment in February 2006 and to date, 496 recipients have been enrolled. Of these enrollees, 53% are female; 78% are Caucasian, and 14% African American. Serum creatinine for this cohort (mean ± SE) prior to January 2006 was 1.4 ± 0.3 mg/dL (median, 1.4 mg/dL) with mean time (±SD) from transplant to biopsy 7.1 ± 5.9 years (median, 5.7 years). Of 496 patients enrolled in this study to date, 337 consecutive patients have had biopsies reviewed by the central pathology lab. Of these 337, death-censored graft failure occurred in 77 recipients and 16 additional recipients died with graft function, consistent with the high-risk nature of the population under study. Overall, the mean time to death-censored allograft failure after biopsy was 306 ± 262 days. Of the 337 consecutive biopsies reviewed, 291 were classified as IF/TA based on a definition of $ci > 0$ or $ci = 0$ (n = 290) and $ct \geq 2$ (n = 1). Conversely, there were 43 cases that were not counted as IF/TA with $ci = 0$, $ct = 1$ and 3 cases with $ci = ct = 0$.

There was a strong association between *iatr* and *tatr* scores (Kendall's Tau-b 0.59 ± 0.03; Table 1). The presence of both *iatr* and *tatr* was also strongly associated with the centrally evaluated *i* scores ($p < 0.0001$). In spite of these associations, of 210 recipients with Banff *i* score of 0, 108 (51%) had *iatr* scores ≥ 1 and 151 (72%) had *tatr* scores ≥ 1, demonstrating the relatively moderate association of inflammatory cell infiltrates in areas of IF and TA with inflammation in areas of viable tubule-interstitial parenchyma which is part of the criteria for acute cellular rejection. As a measure of the strength of this association, the Kendall's Tau-b statistic between *iatr* and *tatr* and the Banff *i* score was 0.54 ± 0.04 and 0.41 ± 0.04, respectively. By comparison, the Kendall Tau-b measuring the association of the Banff *i* and *t* scores was 0.73 ± 0.03.

High *iatr* and *tatr* scores were significantly associated with decreased graft survival as demonstrated by log-rank tests (*iatr* Log-rank = 43.91 3 df, $p < 0.0001$ and *tatr* Log-rank = 9.0 3 df, $p = 0.0293$; Figure 2). Similarly, log-rank tests comparing the presence of *iatr* (*iatr*≥1) versus no *iatr* (*iatr* = 0), demonstrated that allograft failure was significantly more common when *iatr* was detected, regardless of its severity (Log-rank 11.73 1 df; $p = 0.0006$; Figure 3). Similar results were obtained for *tatr* (not shown) but the intensity of this relationship was less strong (Log-rank = 4.88 1 df; $p = 0.0272$).

In multivariate Cox proportional hazards models stratified by clinical center and adjusted for creatinine at time of biopsy, scores of *iatr* or *tatr* of 1 (mild) were not associated with allograft outcome ($p = 0.075$, $p = 0.066$, respectively). However, with *iatr* score = 2, the risk of allograft failure was 2.52 ($p = 0.009$), compared to biopsies with *iatr* = 0 (Table 2). This risk increased dramatically as *iatr* score increased to 3 (HR = 6.35 $p < 0.001$). When further adjusted for inflammation (*i* score) within the biopsy, *iatr*

Table 1: Two-way frequency distribution of 'iatr' and 'tatr' scores in allograft biopsies

	'*tatr*' Scores number and % of total N (337)				
'*iatr*' Scores	0	1	2	3	Total
0	53 (15.7%)	42 (12.5%)	8 (2.4%)	2 (0.6%)	105
1	11 (3.3%)	47 (14.0%)	50 (14.8%)	5 (1.5%)	113
2	1 (0.3%)	13 (3.9%)	65 (19.3%)	14 (4.2%)	93
3	0 (0.0%)	1 (0.3%)	18 (5.3%)	7 (2.1%)	26
Total	65	103	141	28	337

Chi-square 176.1761; $p < 0.0001$.

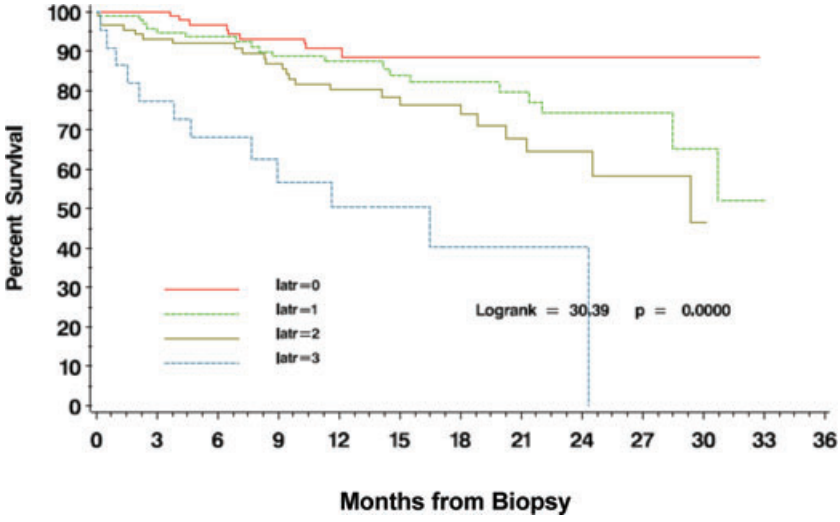

Figure 2. Time to death censored graft failure after allograft biopsy is dependent on the severity of scoring for *iatr*.

score was associated with an even greater risk of allograft loss (Table 2), suggesting that the effect of *iatr* is independent of the extent of inflammation in viable areas of the kidney. While time to allograft biopsy could confound the analysis of survival and *iatr,* time to allograft biopsy was not statistically significant when added to the models in Table 2 (data not shown). There was no difference in mean time from transplant to biopsy between grades of *iatr*, with the shortest period of time occurring in the *iatr* = 3 group (6.6 ± 5.4 years) and the longest in the *iatr* = 0 (7.4 ± 5.6 years; p = 0.76).

Moderate to severe allograft fibrosis is associated with worsened outcomes compared to nonfibrotic kidneys and may thus confound the results of either *iatr* or *tatr* (7). However, we found that when adjusted for the extent of IF ('*ci*') within the biopsy, *iatr* scores of 2 or 3 remained significantly associated with the risk of allograft failure (hazard ratios 2.12 and 3.36, respectively) (Table 2), suggesting that extent of inflammation which is normally not accounted for by conventional Banff scoring, contributes to the demise of allografts with fibrosis present. When adjusting for all biopsy fibrosis, atrophy, and inflammation, *iatr* scores of ≥ 2 demonstrated a 3.4-fold to over fivefold increase in allograft loss (Table 2), demonstrating that *iatr* is strongly predictive of allograft loss even when holding other relevant Banff score variables constant.

One recent report emphasizes the assessment of inflammation present throughout the biopsy using a total inflammation score ('*total i*'; 9) and its relationship to predicting graft failure, particularly in biopsies with IF/TA. To address whether this variable would affect our models, we also scored biopsies using the *total i* schema. Not surprisingly, *total i* score and *iatr* are strongly associated (Kendal Tau-b 0.78 ± 0.024; p < 0.0001) as *iatr* is included in the *total i* assessment. Regression analysis which included of any level of *total i* score was not significantly associated with graft loss (data not shown, p = 0.8766), although this model contained an additional 4 df. Inclusion of *total i* score

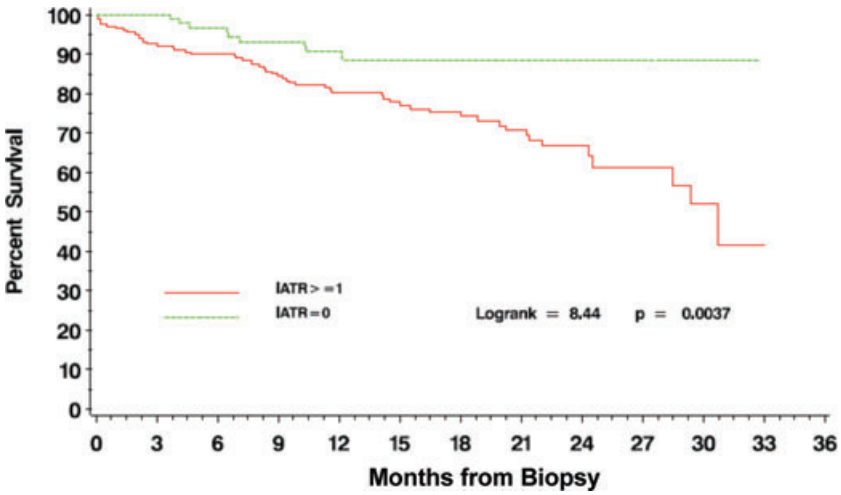

Figure 3. Time to death censored graft failure after allograft biopsy based on the presence or absence of *iatr*.

Table 2: Proportional hazards regression of time to death-censored graft failure based on inflammation in areas of interstitial fibrosis (*iatr*) adjusted for serum creatinine at time of biopsy, and extent of inflammation (i), tubular atrophy (ct) and fibrosis (ci), C4d positivity ($C4d^+$) and donor specific antibody positivity (DSA^+)

	Hazard ratio [95% Confidence Interval]; p-Value					
Group	Model 1 adjusted for creatinine	Model 2 adjusted for i and creatinine	Model 3 adjusted for ci and creatinine	Model 4 adjusted for ct and creatinine	Model 5 adjusted for ci and ct and creatinine	Model 6 adjusted for i, ci, ct, $C4d^+$, DSA^+ and creatinine
iatr = 0	REF[1]	REF	REF	REF	REF	REF
iatr = 1	1.91 [0.95,3.90]; 0.075	2.47 [1.17,5.20]; 0.018	1.59 [0.77,3.30]; 0.212	1.68 [0.81,3.48]; 0.161	1.60 [0.77,3.32]; 0.207	3.36 [1.05,10.68]; 0.0403
iatr = 2	2.52 [1.26,5.02]; 0.009	4.38 [1.95,9.82] <0.001	2.12 [1.02,4.38]; 0.043	2.00 [0.96,4.16]; 0.065	2.07 [0.99,4.35]; 0.053	5.11 [1.44,18.07]; 0.0114
iatr = 3	6.35 [2.91,13.85]; <0.001	12.0 [4.4,32.61]; <0.001	3.36 [1.39,8.13]; 0.007	3.44 [1.42,8.33]; 0.006	3.23 [1.29,8.06]; 0.012	8.07 [1.71,38.07]; 0.0083
Overall p-value for *iatr*	<0.0001	<0.0001	0.0441	0.0543	0.0756	0.0450

[1]REF = reference group.

did reduce the level of hazard ratio and significance of *iatr* with graft failure to 1.83 for *iatr* = 1 (p = 0.195) and 2.55 for *iatr* = 2 (p = 0.0919). However, hazard ratio for *iatr* = 3 remained strong and significant (5.19; p = 0.0098). When *i* score was also included in the analysis, resulting in 11 df, significance for the prediction of graft failure was not observed for *total i*, *iatr* or *i* score, possibly due to correlation between these covariates. Our data suggest that inflammation in atrophic areas of the allograft biopsy is an important component of the *total i* score as a marker of allograft failure.

To address the differential impact of inflammation in various compartments on cases with IF/TA, we have performed additional subgroup analyses on the 291 biopsies classified as IF/TA, assessing the effect of Banff *i* and *iatr* on graft-survival while adjusting for serum creatinine. In these models, *iatr* was a significant predictor of outcome with hazard ratios ranging from 1.78 to 12.35 ($p < 0.0001$) when *ct* was not adjusted and 1.7–7.1 with *ct* adjustment (p = 0.0093). However, *i* was of marginal significance regardless of the adjustment of ct scores (p = 0.061 without and p = 0.085 with adjustment, respectively). The corresponding models for *total i* are significant without adjustment for *ct* (p = 0.004) but are of marginal significance after adjusting for *ct* (p = 0.058). Thus, after adjusting for serum creatinine at the time of biopsy, time to allograft failure in biopsies with IF/TA is strongly associated with *iatr* while marginally associated with other inflammation.

In contrast to *iatr*, *tatr* was less strongly associated with allograft loss. When controlled for serum creatinine, *tatr* was associated with graft failure (p = 0.0308; Table 3). However, when the extent of TA or IF was controlled, *tatr* was not significantly associated with graft failure (Table 3).

As late allograft failure has been linked to donor specific antibody and endothelial injury (9,10), we also assessed the effects of the presence of central C4d staining and donor specific antibody (Tables 2 and 3). Adjustment for these factors did not weaken the ability for *iatr* to predict allograft failure.

We also analyzed the relationship between *iatr* and acute cellular rejection, a diagnosis made in nonatrophic portions of cortex. While the association of *iatr* with a central diagnosis of cellular rejection is statistically significant ($p < 0.0001$), 94.8% of biopsies with a central diagnosis of acute cellular rejection had *iatr* $\geq$ 1 compared with 63.4% of biopsies without rejection. Correspondingly, the strength of association as measured by Kendall's Tau-b is low (0.34 $\pm$ 0.04). In multivariate proportional hazards regression for time to death-censored graft failure, there was no significant impact of treatment of concomitant acute rejection adjusted for TA on graft outcome [HR 0.90 (0.49–1.67), p = 0.74], after adjustment for *iatr*, central Banff *ct*, and creatinine at biopsy. Thus, while acute rejection may be associated with *iatr*, the impact of *iatr* on graft failure is not primarily dependent on the presence of a diagnosis of acute rejection and its treatment.

To further assess the independent effect of *iatr* from the Banff *i*-score on survival, biopsies were categorized into four groups: (1) those with no inflammation (*iatr* = *i* = 0, n = 102), (2) those with inflammation solely in regions of atrophy (*i* = 0 and *iatr*$\geq$1, n = 108), (3) those with inflammation in both atrophic and nonatrophic interstitium (*i*$\geq$1, *iatr*$\geq$1; n = 124) and (4) those with inflammation but not in regions of atrophy (*i* $\geq$ 1, *iatr* = 0; n = 3) (Table 4). Because of the small case number in group 4, we combined groups 3 and 4 for analyses. By combining the two groups, we eliminated an underpowered estimate without

Table 3: Proportional hazards regression of time to death-censored graft failure based on tubulitis in areas of tubular atrophy (*tatr*) adjusted for serum creatinine at time of biopsy, and extent of inflammation (i), tubular atrophy (ct), interstitial fibrosis (ci), C4d positivity ($C4d^+$) and donor specific antibody positivity (DSA^+)

	Hazard ratio [95% Confidence Interval]; p-Value					
Group	Model 1 adjusted for serum creatinine	Model 2 adjusted for i and serum creatinine	Model 3 adjusted for ci and serum creatinine	Model 4 adjusted for ct and serum reatinine	Model 5 adjusted for ci and ct and serum creatinine	Model 6 adjusted for i, ci, ci, $C4d^+$, DSA^+ and serum creatinine
tatr = 0	REF[1]	REF	REF	REF	REF	REF
tatr = 1	2.26 [0.95,5.38]; 0.0656	2.44 [1.01,5.92]; 0.0482	2.09 [0.87,5.02]; 0.1013	2.20 [0.91,5.28]; 0.0787	2.09 [0.87,5.05]; 0.0997	5.76 [1.28,25.91]; 0.0224
tatr = 2	3.17 [1.46,6.85]; 0.0034	4.05 [1.69,9.70]; 0.0017	2.42 [1.08,5.41]; 0.0310	2.30 [1.03,5.14]; 0.0417	2.30 [1.02,5.16]; 0.0445	6.84 [1.51,20.94]; 0.0125
tatr = 3	2.56 [0.96,6.85]; 0.0605	3.13 [0.99,9.90]; 0.525	2.01 [0.73,5.49]; 0.1750	1.64 [0.59;4.53]; 0.3436	1.71 [0.61,4.79]; 0.3047	7.31 [1.39,38.37]; 0.0187
Overall p-value for *tatr*	0.0308	0.0160	0.1976	0.1937	0.2281	0.0889

[1]REF = reference group.

discarding cases. Analysis of this model versus the four groups with the interaction provided similar results to those results are presented here. The time to death-censored graft failure was significantly different between groups (Figure 4; Log-rank = 10.07 2 df; p = 0.0065). In multivariate proportional hazards regression stratified by clinical center and adjusted for creatinine at biopsy, the *iatr*-only group (2) was associated with a threefold increase in hazard over the no inflammation group (1) (Table 5). Although the hazard ratio decreased when adjusted for Banff *ci* or *ct* scores, the results remained statistically significant with hazard ratios in excess of 2.0 (Table 5). Additional adjustment of the models for the contribution of treatment for acute rejection was not significant (data not shown) and did not qualitatively alter these results.

Due to the large number of correlated covariates of interest, we performed model selection through stepwise regression of time to graft failure utilizing all Banff subscores, as well as *iatr*, *tatr*, *total i* score, C4d, creatinine at the time of biopsy, time to biopsy from transplant, and treated acute rejection. In this model, five variables were selected as significantly associated with graft outcome: creatinine at time of biopsy ($p < 0.0001$), *iatr* ($p < 0.0001$), central *ct* score ($p = 0.075$), central *mm* score ($p < 0.0001$) and peritubular capillaritis ($p = 0.0097$) while *total i* score, C4d presence and other inputted variables were not selected to fit this model. As shown in Table 6, *iatr* remained a strong predictor of graft failure, even after controlling for numerous variables that might affect graft outcome. This supports the notion that *iatr* is a strong and important predictor of allograft failure in late kidney biopsies for graft dysfunction.

Discussion

Identifying pathological correlates of *late* kidney allograft loss is critical for guiding potential therapies, designing clinical trials, and developing biomarkers that may be used to identify recipients at risk. In this regard, the DeKAF study, with central pathological analysis of biopsies obtained for rising creatinine or significant proteinuria, provides a unique chance for correlating Banff schema with outcomes, an opportunity that has not previously existed in such a large scale.

In this study, we examined the extent of inflammation in areas of fibrosis and tubulitis in areas of TA of the kidney allograft biopsied at late timepoints post transplantation.

Table 4: Two-way frequency of *iatr* and the presence of interstitial inflammation

	'*iatr*' Scores number and % of total N (337)				
Banff *i* score	0	1	2	3	Total
0	102 (29.6%)	75 (22.5%)	26 (7.8%)	7 (2.1%)	210
1	0 (0%)	25 (7.5%)	23 (6.9%)	4 (1.2%)	113
2	2 (0.6%)	9 (2.7%)	39 (11.7%)	2 (0.6%)	93
3	1 (0.3%)	4 (1.2%)	5 (1.5%)	13 (3.9%)	26
Total	105	113	93	26	337

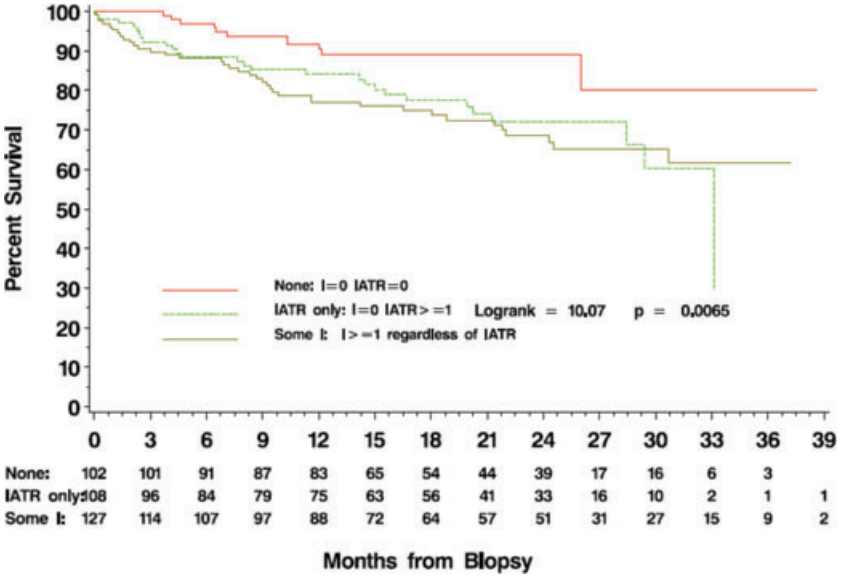

Figure 4. Time to death censored graft failure after allograft biopsy comparing biopsies with no inflammation (*i* = 0, *iatr* = 0; n = 102), inflammation only in areas of fibrosis (*i* = 0 and *iatr* ≥ 1; n = 108) and those with inflammation in both fibrotic and nonfibrotic areas (*i* ≥ 1, *iatr* ≥ 1; n = 124 and *i* > 0 and *iatr* = 0; n = 3) demonstrating the independent effect of *iatr* from Banff *i* score on allograft failure.

These are currently not included in the Banff schema (1,2), which focuses exclusively on inflammation in viable tissue, because inflammation in areas of atrophy and fibrosis was considered to be indicative of repair of previous injury. However, recent investigations into the role of cellular and antibody mediated inflammation in later allograft loss demonstrate a far reaching extent of response (4,9,13,14). Compartmentalizing the kidney into separate areas, those of atrophy and those without, while for pathological coding may seem sound, does not make biological sense, as inflammatory cell infiltration can have more far reaching effects in the microenvironment, through both secretory signals, as well as direct cell-cell contact (15).

While tubulitis in these atrophic areas had a significant relationship to graft failure, this relationship was not as strong as the presence of inflammation. We found *iatr* scores to be strongly associated with graft failure in biopsies of kidneys with new onset, late dysfunction. Moreover, this relationship remained independent of the extent of overall allograft fibrosis and atrophy, or of inflammation in areas of viable tubulo-interstitium ('i') and acute cellular rejection. Of those cases with *iatr* ≥ 1, only 62 (27%) had a primary or secondary diagnosis of rejection while remaining 168 cases did not have reported rejection. Thus acute cellular rejection occurred in the minority of biopsies classified with *iatr*. Similarly, only 19/230 (8.2%) of *iatr* biopsies had a primary or secondary diagnosis of antibody mediated rejection. Perhaps the strength of the association of the *iatr*-only group with outcome may be due to the presence of undiagnosed acute rejection, which, if left untreated, would naturally result in a worse outcome. This hypothesis is not directly testable in this data set due to detection bias in the assignment of treatment. Our findings suggest not only that *iatr* may be critical to the promulgation of ongoing injury, but also should be included in the Banff analyses

Table 5: Proportional hazards regression estimates of time to death-censored graft failure based on inflammation status: *iatr* only (*i* = 0, *iatr*>0), and *iatr* or *i* ≥ 0 relative to no inflammation (*i* = 0 *iatr* = 0) adjusted for serum creatinine at time of biopsy, and extent of inflammation (*i*), atrophy (*ct*), fibrosis (*ci*), C4d positivity ($C4d^+$) and Donor Specific Antibody positivity (DSA^+)

	Hazard ratio [95% Confidence Interval]; p-Value					
Group	Model 1 adjusted for serum creatinine	Model 2 adjusted for ci and serum creatinine	Model 3 adjusted for ct and serum creatinin	Model 4 adjusted for $C4d^+$, DSA^+ and serum creatinine	Model 5 adjusted for $C4d^+$, DSA^+, ci and serum creatinine	Model 6 adjusted for $C4d^+$, DSA^+, ct and serum creatinine
No inflammation [*i* = 0, *iatr* = 0]	REF[1]	REF	REF	REF	REF	REF
iatr only [*i* = 0, *iatr* > 0]	3.06 [1.51,6.19]; 0.0018	2.31 [1.10,4.83]; 0.0262	2.42 [1.16,5.08]; 0.0191	5.06 [1.71,14.97]; 0.0034	3.70 [1.21,11.34]; 0.0221	4.02 [1.28,12.60]; 0.0170
Inflammation [*i* > 0, *iatr* ≥ 0]	2.10 [1.09,4.06]; 0.0268	1.60 [0.81,3.17]; 0.1780	1.51 [0.76,3.01]; 0.2367	2.88 [1.02,8.07]; 0.0448	2.35 [0.84,6.63]; 0.1052	2.35 [0.82,6.70]; 0.1105

[1]REF = reference group.

Table 6: Stepwise regression analysis model estimates of time to death-censored graft failure of levels of iatr. The model selected *iatr*, and creatinine at time of biopsy, central *mm*, *ct*, and *ptc* scores as predictors of graft failure

iatr	Hazard ratio [95% Confidence Interval]; p-Value
0	REF[1]
1	2.27 [0.891, 5.77]; 0.0860
2	2.98 [1.07, 8.34]; 0.0371
3	4.75 [1.58, 14.27]; p = 0.0055

[1]REF = reference group.

of allograft biopsies. This could lead to new classifications and ultimately the testing of treatment for this finding to provide direct evidence of the impact of inflammation in the failing allograft.

Our data complements and adds to the studies of Mengel et al. who reported the association of inflammation in areas of fibrosis with decreased allograft survival. In their studies of late posttransplant biopsies for cause, 77 recipients had IF/TA. Of these, 46 biopsies had $\geq$ 50% of the fibrosis area showing infiltrates with 31 having <50% of fibrotic areas with infiltrates. Those with increased infiltrates in fibrotic areas had significantly decreased graft survival (p = 0.02) (10). In a subsequent study, Mengel et al. showed that the total inflammation score was a better predictor of post-biopsy graft survival than the Banff '*i*' score (9). In our multicenter study of late posttransplant biopsies for cause (n = 337), we show that inflammation in areas of atrophy (*iatr*) can be scored in a similar manner to inflammation in viable tissue (Banff *i* score). In a Cox model-controlled for transplant center, inflammation in viable tissue (Banff *i* score), the extent of IF ('ci') within the biopsy, C4d positivity, the presence of DSA and serum creatinine level at the time of biopsy, we found *iatr* to be associated with increased graft loss.

In summary, semi-quantitative analysis of inflammation in areas of IF provides a powerful measure of allograft injury and is a strong predictor of graft loss. Even after adjusting for IF and TA, and renal function at the time of biopsy, *iatr* remains a strong marker of graft failure. We suggest that Banff schema be updated to include more global assessments of inflammation within the biopsy to enhance the descriptive and predictive value of allograft biopsy when obtained in the setting of clinical concern.

Acknowledgments

We would like to thank our local pathologists (William Cook, Lynn Cornell, Gretchen Crary, Ian Gibson, Donna Lager, Ramesh Nair, Behzad Najafian and Kim Solez) who are playing a critical role in this study and Stephanie Daily and Wendy Bailey for their help in the preparation of the manuscript. This work was supported by NIH funding 5U01AI058013.

Disclosure

The authors of this manuscript have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

References

1. Solez K, Axelsen RA, Benediktsson H et al. International standardization of criteria for the histologic diagnosis of renal allograft rejection: The Banff working classification of kidney transplant pathology. Kidney Int 1993; 44: 411–422.
2. Racusen LC, Solez K, Colvin RB et al. The Banff 97 working classification of renal allograft pathology. Kidney Int 1999; 55: 713–723.
3. Cosio FG, Grande JP, Larson TS et al. Kidney allograft fibrosis and atrophy early after living donor transplantation. Am J Transplant 2005; 5: 1130–1136.
4. Cosio FG, Grande JP, Wadei H, Larson TS, Griffin MD, Stegall MD. Predicting subsequent decline in kidney allograft function from early surveillance biopsies. Am J Transplant 2005; 5: 2464–2472.
5. Moreso F, Ibernon M, Goma M et al. Subclinical rejection associated with chronic allograft nephropathy in protocol biopsies as a risk factor for late graft loss. Am J Transplant 2006; 6: 747–752.
6. Shishido S, Asanuma H, Nakai H et al. The impact of repeated subclinical acute rejection on the progression of chronic allograft nephropathy. J Am Soc Nephrol 2003; 14: 1046–1052.
7. Nankivell BJ, Borrows RJ, Fung CL, O'Connell PJ, Chapman JR, Allen RD. Delta analysis of posttransplantation tubulointerstitial damage. Transplantation 2004; 78: 434–441.
8. Solez K, Colvin RB, Racusen LC et al. Banff 07 classification of renal allograft pathology: Updates and future directions. Am J Transplant 2008; 8: 753–760.
9. Mengel M, Reeve J, Bunnag S 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 2009; 9: 1859–1867.
10. Mengel M, Reeve J, Bunnag S et al. Molecular correlates of scarring in kidney transplants: The emergence of mast cell transcripts. Am J Transplant 2009; 9: 169–178.
11. Gourishankar S, Leduc R, Connett J et al. Pathological and clinical characterization of the 'Troubled Transplant': Data from the DeKAF Study. Am J Transplant 2010; 10: 315–323.
12. Racusen LC, Colvin RB, Solez K et al. Antibody-mediated rejection criteria – an addition to the Banff 97 classification of renal allograft rejection. Am J Transplant 2003; 3: 708–714.
13. Einecke G, Sis B, Reeve J et al. Antibody-mediated microcirculation injury is the major cause of late kidney transplant failure. Am J Transplant 2009; 9: 2520–2531.
14. Hidalgo LG, Campbell PM, Sis B et al. De novo donor-specific antibody at the time of kidney transplant biopsy associates with microvascular pathology and late graft failure. Am J Transplant 2009; 9: 2532–2541.
15. Saadi S, Wrenshall LE, Platt JL. Regional manifestations and control of the immune system. FASEB J 2002; 16: 849–856.