Supplemental Information:

1) Supplemental Results:

2) Supplemental Concise Methods:

Supplemental Table 1. Individual Analysis of Predictors of Serum Creatinine Change from Baseline (left side) and Graft Failure (right side) within 24 Months

Effect	Serum Creatinine Change[1]		Graft Failure within 24 mo.[2]	
	Test Statistic	p-value	Test Statistic	p-value
Age	$F_{1,830} = 2.23$	0.136	$X^2_1 = 0.27$	0.603
Race	$F_{1,820} = 1.57$	0.210	$X^2_1 = 1.89$	0.169
Gender	$F_{1,830} = 1.75$	0.186	$X^2_1 = 0.09$	0.769
Graft Type	$F_{3,830} = 0.74$	0.530	$X^2_3 = 0.84$	0.840
Plasma PCR (index biopsy)[3]	$F_{1,559} = 6.41$	0.012	$X^2_1 = 0.00$	0.968
Week of PVN Diagnosis	$F_{1,815} = 0.04$	0.848	$X^2_1 = 0.02$	0.879
Polyomavirus Load Levels (PVL)	$F_{1,830} = 25.68$	<0.001	$X^2_1 = 4.64$	0.031
Viral Inclusion Bodies	$F_{1,815} = 11.20$	<0.001	$X^2_1 = 0.00$	0.967
Banff Scores				
i	$F_{1,812} = 3.62$	0.057	$X^2_1 = 0.44$	0.506
ti	$F_{1,830} = 4.09$	0.043	$X^2_1 = 4.31$	0.038
ci	$F_{1,830} = 8.14$	0.004	$X^2_1 = 5.64$	0.018
ct	$F_{1,830} = 3.32$	0.069	$X^2_1 = 2.80$	0.094
t	$F_{1,825} = 2.18$	0.140	$X^2_1 = 0.71$	0.400

[1]The basic analysis model used to test the association between each of these effects and change from baseline in serum creatinine (S-Cr) was a likelihood-based, mixed-effects repeated measures (MMRM) rank analysis, including fixed categorical effects for center and visit, with a continuous fixed covariate for the rank of baseline serum creatinine.

[2]A logistic regression model adjusting for study center was used to test the association separately for each effect on the occurrence of graft failure within 24 months.

[3]Rank Plasma PCR values were used due to extreme skewness in the data.

Supplemental Table 2. Forward Selection process (Step 2)[1] for Predictors of Serum Creatinine Change from Baseline (left) and Graft Failure (right) within 24 Months

Effect	Serum Creatinine Change		Graft Failure within 24 mo.	
	Test Statistic	p-value	Test Statistic	p-value
Age				
Race				
Gender				
Graft Type				
Plasma PCR (index biopsy)	$F_{1,559} = 1.75$	0.187		
Week of PVN Diagnosis				
Polyomavirus Load Levels (PVL)	Entered first		$X^2_1 = 2.71$	0.100
Viral Inclusion Bodies	$F_{1,815} = 1.51$	0.220		
Banff Scores				
i				
ti	$F_{1,830} = 0.08$	0.773	$X^2_1 = 1.02$	0.314
ci	$F_{1,830} = 3.47$	0.063	Entered first	
ct				
t				

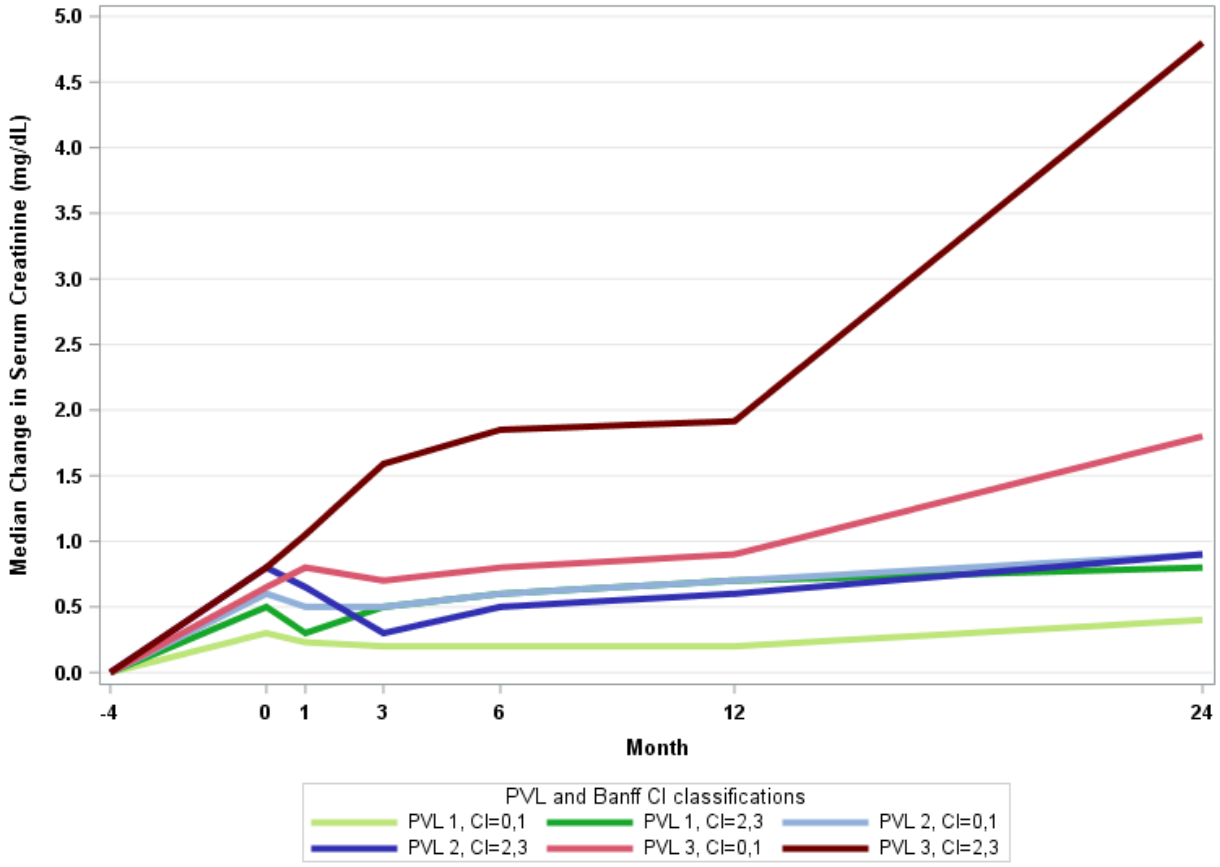
[1]The first step for both models was to include the effect that explained the most impact based on the results displayed in Supplemental Table 1 (denoted by the table entry "Entered first" above). The remainder of the table entries are the results of testing each of the significant effects from Supplemental Table 1 one at a time as the second predictor in the model. Blank cells above indicate that the effect was not significant in the first step.

Supplemental Table 3. Distribution of Banff ti-scores among PVN Disease Classes for Patients with Graft Failure at 24 months[1]

Banff ti-score	PVN Class I n graft failure / total N (%)[2]	PVN Class II n graft failure / total N (%)[2]	PVN Class III n graft failure / total N (%)[2]	All PVN Classes n graft failure / total N (%)[2]
0	2/15 (13)	3/10 (30)	0/0 (0)	5/25 (20)
1	3/18 (17)	8/38 (21)	0/1 (0)	11/57 (19)
2	1/7 (14)	14/37 (38)	4/5 (80)	19/49 (39)
3	1/4 (25)	9/24 (38)	7/16 (44)	17/44 (39)
All ti-scores	7/44 (16)	34/109 (31)	11/22 (50)	52/175 (30)

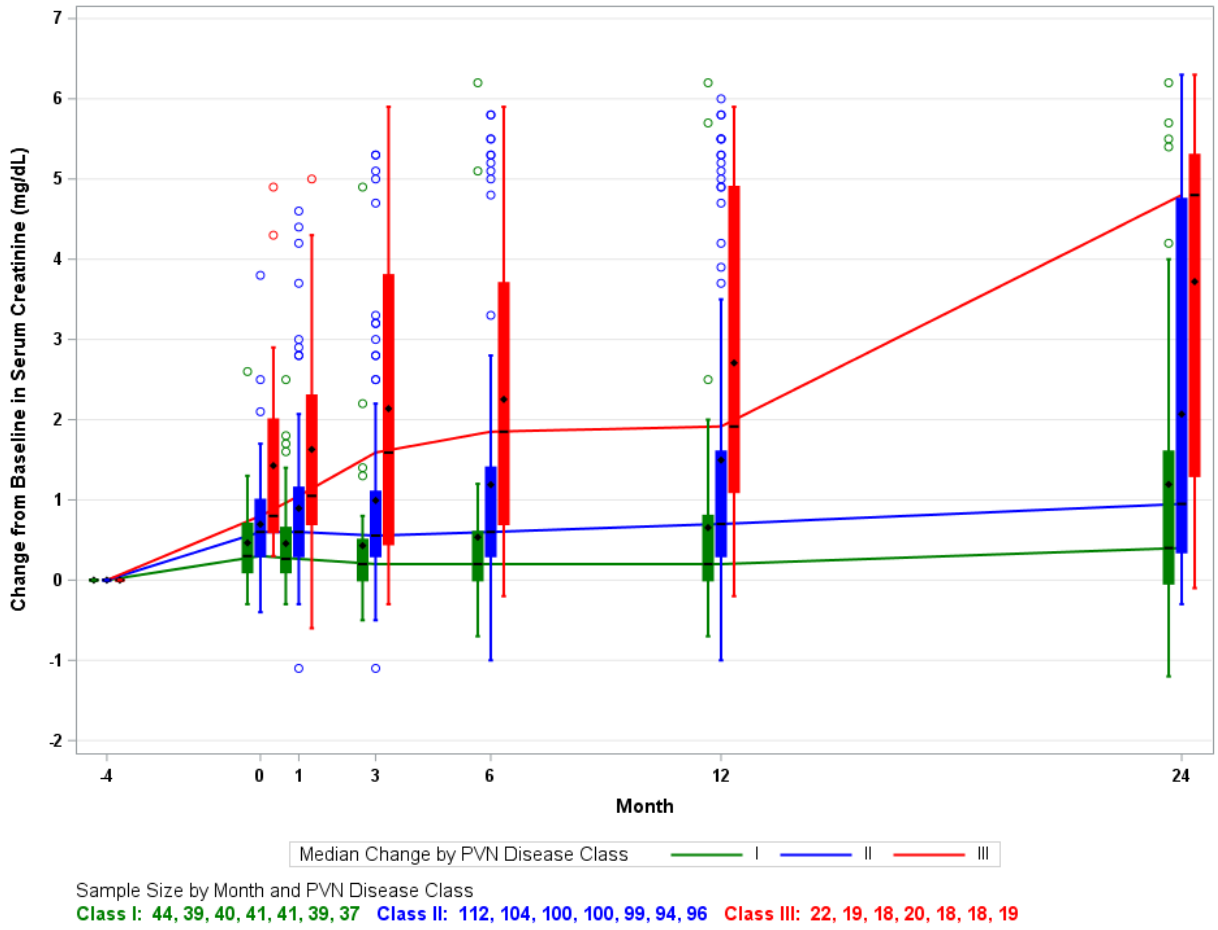
[1]There is no significant association between Banff ti-score and graft failure at 24 months when controlling for PVN disease classes (p=0.238). The incidence of graft failure at 24 months increases with PVN class, controlling for level of Banff ti-score (p=0.044).
[2]The percentages (%) are based on the number of patients who have graft failure within 24 months (n) out of the total number of patients who fall into the PVN-class/Banff ti category (N).

Supplemental Figure 1. Median Change in Serum Creatinine from Baseline by pvl and Banff ci Score in Index Biopsy



Median change in S-Cr from baseline (y-axis) is plotted by visit (x-axis) for all combinations of the two most significant determinants of overall allograft function and failure: pvl and Banff ci scores; pvl-scores (pvl-1: ≤1%, pvl-2: >1 - 10%, pvl-3: >10%) and Banff ci scores (≤1, ≥2), respectively.

Supplemental Figure 2. Distribution of Change of Serum Creatinine (mg/dL) from Baseline by PVN Class and Visit Month



Distributions of change from baseline in serum creatinine are presented in the form of box plots by visit month for PVN Classes I, II, and III. The bottom and top edges of the box indicate the intra-quartile range (IQR). The black diamond inside the box is the mean value; the line inside the box is the median value. Median values are connected between timepoints by a colored line representing the PVN Class. The "whiskers" that extend from each box indicate the range of values that fall between the IQR and 1.5IQR. Data points beyond 1.5IQR are considered outliers and are indicated by circles. Sample sizes for median serum creatinine change from baseline are given by PVN disease class and month.

Ad 2) Supplemental Statistical Methods

The lowest serum creatinine (S-Cr) measurement collected within 4 months prior to diagnosis is the “baseline level” and denoted as month -4 in all figures. Follow-up measurements were recorded at time of index biopsy (“peak”, denoted as month 0 in all figures) and subsequently at months 1, 3, 6, 12, and 24. Analysis of graft function over time, i.e. changes of S-Cr levels compared to baseline levels, was performed using a likelihood-based, mixed-effects repeated measures model (MMRM). The Kenward-Rogers correction to degrees of freedom was applied and an unstructured covariance matrix was assumed. The model included fixed categorical effects for center and visit (month 0 – 24), with a continuous fixed covariate for the baseline S-Cr reading. The association of demographic and histologic parameters and BK-viremia levels by PCR at time of index biopsy with allograft function (change in S-Cr levels) over 24 months follow-up compared to baseline S-Cr was determined with separate models (one for each parameter of interest). Individual significant predictors of allograft function over time were then analyzed with an MMRM using a forward approach, starting with the parameter explaining the most variance in function and adding additional predictors to test for further significant improvements to the overall model fit, based on two-sided tests with an $\alpha=0.050$. Due to the skewed nature of the S-Cr levels, a rank transformation was used in the MMRM for both the change from baseline reading and the baseline covariate. Note for presentation purposes and clarity, median change in S-Cr (unadjusted for center and baseline S-Cr) is displayed in the figures rather than the LSMean changes in the ranks. S-Cr values in the clinical setting of graft failure (or

values >7 mg/dl) were imputed with a value of 7 mg/dl at time of graft failure and at all pre-set time points thereafter.

The same set of predictors used above in the S-Cr models were also evaluated one at a time with respect to graft failure using logistic regression. This model was used to describe the linear relationship using maximum likelihood methods between each of the explanatory variables and the logit (or log of the odds) of graft failure versus no graft failure within 24 months of index biopsy, adjusting for study center. Individual significant predictors of graft failure were subsequently analyzed simultaneously with a logistic regression model using a forward approach, starting with the parameter explaining the most variance in graft failure. Additional predictors were added one at a time to test for further improvements to the overall model fit, until no explanatory variables remained which significantly predicted the log odds of graft failure within 24 months of index biopsy (α=0.050).

Finally parameters, which independently explained variation in either graft function or failure, were used to categorize patients into three PVN classes. An MMRM which included effects for baseline S-Cr reading, specific center, visit, PVN class, and PVN class-by-visit interaction was used to compare function between the PVN classes. Additional testing used to compare the PVN classes included Student's t-test, one-way ANOVA, Cochran-Mantel-Haenszel Chi-Square test using the midrank scores, or Kruskal–Wallis rank test, as needed. All statistical tests were two-sided with α=0.050. Statistical analyses were performed using SAS software (SAS Institute, Cary, NC, USA), Version 9.4 of the SAS System for Windows.