Microarray Analysis of Rejection in Human Kidney Transplants Using Pathogenesis-Based Transcript Sets

T. F. Mueller,1, †, G. Einecke, †, J. Reeve, B. Sis,1, M. Meng,1, G. S. Jhangri,1, S. Bunnag, J. Cruz, D. Wishart, C. Meng, G. Broderick, B. Kaplan, and P. F. Halloran,*

†Department of Medicine, Division of Nephrology & Immunology; ‡Department of Computing Sciences; †Department of Public Health Sciences, School of Public Health and †Department of Medicine, Division of Pulmonary Medicine, University of Alberta, Edmonton, Canada
*Department of Medicine & UIC Transplant Center, University of Illinois, Chicago, IL

Corresponding author: Philip F. Halloran, phil.halloran@ualberta.ca

Microarrays offer potential for objective diagnosis and insights into pathogenesis of allograft rejection. We used mouse transplants to annotate pathogenesis-based transcript sets (PBTs) that reflect major biological events in allograft rejection—cytotoxic T-cell infiltration, interferon-γ effects and parenchymal deterioration. We examined the relationship between PBT expression, histopathologic lesions and clinical diagnoses in 143 consecutive human kidney transplant biopsies for cause. PBTs correlated strongly with one another, indicating that transcriptome disturbances in renal transplants have a stereotyped internal structure. This disturbance was continuous, not dichotomous, across rejection and nonrejection. PBTs correlated with histopathologic lesions and were the highest in biopsies with clinically apparent rejection episodes. Surprisingly, antibody-mediated rejection had changes similar to T-cell-mediated rejection. Biopsies lacking PBT disturbances did not have rejection. PBTs suggested that some current Banff histopathology criteria are unreliable, particularly at the cut-off between borderline and rejection. Results were validated in 51 additional biopsies. Thus many transcriptome changes previously described in rejection are features of a large-scale disturbance characteristic of rejection but occurring at lower levels in many forms of injury. PBTs represent a quantitative measure of the inflammatory disturbances in organ transplants, and a new window on the mechanisms of these changes.

Key words: Banff, gene expression, histology, inflammation, kidney transplantation, rejection, transcriptome

†These authors contributed equally to this work.

Introduction

The diagnosis of allograft rejection remains an important issue in kidney transplantation. Rejection is associated with loss of function, proteinuria, atrophy, scarring and graft loss (1), and can be manifested as an episode of functional deterioration or as subtle deterioration. The two mechanisms of rejection recognized histologically in the Banff classification (2–4) are T-cell-mediated rejection, diagnosed by scoring interstitial inflammation, tubulitis and vasculitis, and antibody-mediated rejection, a hallmark of which is C4d deposition in peritubular capillaries (5). Despite international consensus, histopathologic grading of rejection is difficult to reproduce (6–10), particularly at the important interface separating T-cell-mediated rejection from borderline changes (8). Moreover, interstitial inflammation and tubulitis are not specific and are often found in stable transplants where their significance is unclear (11,12). The clinical relevance of the Banff classification with respect to treatment response and graft outcome has been validated in numerous studies (13–16). However, the histologic criteria empirically derived for T-cell-mediated rejection have never been independently validated as a diagnostic test, i.e. assessing sensitivity, specificity and diagnostic accuracy, because there is no true gold standard.

The emergence of microarray technology for genome-wide transcriptome analysis offers potential for objective and quantitative diagnosis and insights into pathogenesis. Transcript expression is altered in rejecting allografts (17,18) and in peripheral blood during allograft rejection (18,19), but the relationship between transcriptome changes, mechanisms and histopathologic lesions remains unclear and the findings are heterogeneous (17). This problem is due in part to the high error rate associated with single-gene based analyses using small data sets (20). More recent microarray-based analyses are addressing this limitation by grouping individual transcripts into sets that correspond to major biological processes (21–23).

To dissect and measure the transcriptome disturbances in rejection, we developed pathogenesis-based transcript sets (PBTs), defined in experimental mouse kidney
transplants and in cell cultures (24–26). The PBTs are similar in concept to an in vitro-derived wound-response transcript set applied in cancer studies (23). PBTs reflect mechanisms intrinsic to the rejection process—cytotoxic T-cell infiltration, interferon-γ (Ifng) effects and parenchymal deterioration. Thus summarizing large-scale transcriptome changes as PBTs provides a quantitative tool for assessment of allograft biopsies and permits the calculation of correlations with immunologic mechanisms and histopathology. The present study analyzed PBT expression in human kidney transplant biopsies for clinical indications (‘for cause’) and compared the results to histopathology and clinical findings.

Materials and Methods

Defining the PBTs

The PBTs evaluated in the present study were defined to represent the major biologic events in cellular graft rejection—cytotoxic T-cell infiltration, Ifng effects and epithelial deterioration—and were derived from experimental mouse kidney transplants and cultured cells. In mouse allografts, we observed increased expression of transcripts associated with T-cell infiltration and Ifng effects and loss of kidney transcripts.

Cytotoxic T lymphocyte-associated transcripts (CATs) were defined by high expression in cultured lymphocytes and low expression in normal kidneys. In addition, mouse CATs (CAT1) (n = 236) (25) had increased expression in day 5 mouse allografts (CBA into B6) compared to normal mouse (CBA) kidneys (n = 236). Human CATs (CAT2) (n = 382) are analogous to mCATs but were defined in cultured human effector T cells and were not filtered for high expression in rejecting kidneys. Details of the human CAT set will be published separately (Hidalgo LG et al. The transcriptome of human cytotoxic T cells: similarities and disparities among allostimulated CD4+ CTL, CD8+ CTL and NK cells. Manuscript in preparation). Ifng effects were summarized as Ifng-dependent rejection-induced transcripts (GRITs)—GRIT1 (n = 68) and GRIT2 (n = 326), defined in wild-type and Ifng-deficient mouse kidney allografts (26). Parenchymal injury was represented by two kidney transcript sets: a large set (KT1) (n = 1481), identified by high expression in normal mouse kidney and low expression in mouse inflammatory cells, and a more restricted set of solute carriers (KT2) (n = 64) (27). PBTs derived from mouse kidneys (25–27) were translated into human ortholog transcripts (www.affymetrix.com). A detailed description of the PBT algorithm can be found in supplementary Figure 1. Probe sets in each PBT are available on our homepage (http://transplants.med.ualberta.ca/).

Figure 1: Relationship between PBT scores and histopathologic lesions or clinical diagnosis. We calculated PBT scores for each biopsy (n = 143) and analyzed the relationship between PBT scores and the degree of interstitial infiltrate (i score) (Panel A), tubulitis (t score) (Panel B), or intimal arteritis (v score) (Panel C) in all biopsies for cause. For each of the three biological processes, we selected one PBT for representation in this figure (CAT1, GRIT1, KT1). Each panel represents the average PBT scores (log2 ± SD) for each of the histopathology scores 0–3. Panel D shows the relationship between PBT scores and the clinical diagnosis of rejection episode in those biopsies with tubulitis (t ≥ 1) (n = 67, excluding BK cases). We compared PBT scores in biopsies with clinically no T-cell-mediated rejection episode (n = 45) to those with a clinical episode (T-cell-mediated or mixed T-cell- and antibody-mediated rejection) (n = 22). Biopsies with a clinical rejection episode are separately shown for those where treatment was initiated postbiopsy (n = 19) or those in which anti-rejection therapy had been started before the biopsy (n = 3). Numbers represent the average score for each PBT in each diagnostic group (log2 values ± SD). The p-values were obtained by t-test. *p ≤ 0.05, **p ≤ 0.01.
Patient population and specimens

The study was approved by the University of Alberta Health Research Ethics Board (Issue # 5299). After receiving written informed consent, every consenting patient undergoing a transplant biopsy for cause (deterioration in function, proteinuria, stable impaired function) as standard of care between 01/2004 and 10/2006 was included. Clinical data was collected for each patient and entered into a Laboratory Information Management System. We included all consecutive biopsies for cause obtained during the specified time. As controls, normal kidney tissue was obtained from histopathologically unaffected areas of the cortex of native nephrectomies performed for renal carcinoma. The majority of patients (n = 85) had only one biopsy, but 24 patients had more than one biopsy during this time (17 patients with two biopsies, 4 with three, and 3 with four biopsies). The biopsies were taken at different time points (often months apart) and represent independent clinical indications for the biopsy. Histology scores and gene expression values were obtained for each biopsy independently, and the diagnosis can be entirely different in two biopsies from the same patient. Therefore, we assume that the biological processes reflected in the PBT scores from two biopsies for cause are independent. This was confirmed, retrospectively, in a separate analysis showing that repeat biopsies were not more similar to the previous biopsy than they were to random kidneys. Thus, biopsies from the same patient were not analyzed as repeated measurements of the same event.

To validate our findings, we used an additional 51 biopsies (consecutive biopsies for cause collected between 11/2006 and 03/2007), with demographics similar to the 143 biopsies in the original set, as an independent test set.

Biopsies were obtained under ultrasound guidance by spring-loaded needles (ASAP Automatic Biopsy, Microvasive, Watertown, MA). In addition to cores for conventional assessment, one 18-gauge biopsy core was collected for gene expression analysis and placed immediately in RNA Later, kept at 4 °C for 4–24 h, then stored at −20 °C. Following homogenization in 0.5 mL of Trizol reagent (Invitrogen, Carlsbad, CA), total RNA was extracted and purified using the RNeasy Micro Kit (Qiagen, Ontario, Canada) (average 4 μg/core). RNA (1–2 μg) was labeled using GeneChips HT One Cycle Target Labeling and Control Kit. Quality of labeled cRNA was assessed on an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) (RNA integrity number >7) before hybridization to HG U133 Plus 2.0 GeneChip (Affymetrix Santa Clara, CA). GeneChips were scanned using GeneArray Scanner (Affymetrix) and processed with GeneChip Operating Software Version 1.4.0 (Affymetrix). Detailed protocols are available in the Affymetrix Technical Manual (www.affymetrix.com).

Paraffin sections were prepared and graded according to Banff criteria by a renal pathologist (IBS) (3,5,28). C4d staining was performed on frozen sections using a monoclonal anti-C4d antibody (Quidel, San Diego, CA). C4d positivity was defined as a diffuse circumferential linear staining with an intensity of mild or greater in >50% of sampled peritubular capillaries, as established by the Banff schema (9). C4d staining was graded as diffuse, focal or negative if there was a linear staining in >50%, 1–50%, or no staining of peritubular capillaries, respectively. All samples had adequate cortical tissue for analysis by Banff criteria with the exception of six biopsies with less than two large arteries (four biopsies had only one artery, two biopsies had no arteries).

Diagnostic classifications

Histopathologic diagnoses included rejection (T-cell-mediated, antibody-mediated, mixed), borderline changes, BK nephropathy and others, e.g. calcineurin inhibitor toxicity. For patients with histopathologic rejection, we defined a clinical rejection ‘episode’ based on retrospective assessment of functional changes during the clinical course by two nephrologists, independent of the transcript results, based on compatible histopathology with clinically apparent functional changes: decrease in estimated glomerular filtration rate (eGFR) ≥ 25% from baseline (up to 4 months preceding biopsy to include cases with infrequent visits) and/or response to therapy (an increase in estimated GFR ≥ 25% within 1 month), in the absence of alternative explanations (e.g. obstruction, calcineurin inhibitor toxicity). GFR was estimated using Cockcroft Gault equation: [(140—recipient age) × recipient lean body weight × recipient gender] × [72 × recipient creatinine × 0.0113]. Biopsies showing BK virus by in situ hybridization and/or electron microscopy were designated BK nephritis (n = 6).

Data analysis

Data files were preprocessed using robust multi-chip averaging in Bioconductor version 1.9, R version 2.4. Gene expression is given as fold change versus controls. Gene expression within each PBT was summarized as the PBT score: the geometric mean of fold changes across all probe sets. The relationship between PBTs was assessed by Spearman rank correlation; the relationship between PBTs and histopathologic scores was assessed by ANOVA and multivariate analysis. Using linear discriminant analysis (LDA), separate PBT score classifiers were built on histopathology and on the retrospective diagnosis of episodes. In each case, two classes were defined: rejection (T-cell-mediated, antibody-mediated, mixed) and non-rejection (all other biopsies). Classifier accuracy (proportion of samples correctly classified), sensitivity, specificity and positive and negative predictive values (proportion of those samples predicted to be positive [negative] that are in fact positive [negative]) were estimated based on a resampling procedure in which 100 random 2:1 training: test set splits of the data were performed. Samples with classifier probabilities >0.25 were assigned calls of rejection. The final equation for calculating the probability of rejection in Figure 2 was based on the LDA using all 143 biopsies. There was no feature selection step—the CAT1, GRIT1 and KT1 scores for the samples were used in each validation loop. CAT2, GRIT2 and KT2 scores were not used because of their high correlations with CAT1, GRIT1 and KT1, respectively (see results below).

Microarray raw data files are available at our homepage at http://transplants.med.ualberta.ca.

Real-time RT-PCR

Expression of selected transcripts (CD8A, PRF1, GZMB, CXCL9, CXCL11 and FABP1) was confirmed by real-time RT-PCR using TaqMan Gene Expression Assays (Applied Biosystems, Foster City, CA). Detailed methods and ABI gene expression IDs are available as supplementary material.

Results

Patient demographics

All consenting patients were included in the study; there were no exclusions. We studied 143 consecutive renal transplant biopsies for cause obtained between 1 week and 20 years posttransplant (median 19 months) from 109 consenting patients (supplementary Table 1).

PBT scores correlate strongly with one another

We developed six PBTs representing three biologic processes in graft rejection—cytotoxic T-cell infiltration (CAT1, CAT2), Ifng effects (GRIT1, GRIT2) and parenchymal deterioration causing loss of kidney transcripts (KT1, KT2). The scores for CATs, GRITs and KTs correlated strongly with one another (Table 1): CAT1 with CAT2 (0.98); GRIT1 with
Figure 2: Relationship between PBT scores, histopathologic lesions, histopathologic and clinical diagnosis, and classifier predictions. (A) Biopsies for cause (n = 143) were sorted based on the CAT1 score (from lowest to highest). According to this order, scores for all PBTs (CAT1, CAT2, GRIT1, GRIT2, KT1, KT2) are illustrated for each individual biopsy for cause. The panel above the graph illustrates the relationship of the PBT scores to the presence of Acute Tubular Necrosis (ATN), the degree of interstitial infiltrate (i score), tubulitis (t score), intimal arteritis (v score), histopathology diagnosis, retrospective clinical-pathologic diagnosis and the probability of rejection (%), predicted from the classifiers. (B) We repeated the analysis in an independent set of 51 biopsies to validate the findings. Biopsies were sorted based on the CAT1 score, and the relationship between PBT scores, diagnosis and classifier predictions are shown.
Table 1: Correlations between PBT scores across 143 biopsies for cause

<table>
<thead>
<tr>
<th></th>
<th>CAT1 (n = 236)</th>
<th>CAT2 (n = 382)</th>
<th>GRIT1 (n = 68)</th>
<th>GRIT2 (n = 326)</th>
<th>KT1 (n = 1481)</th>
<th>KT2 (n = 64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT1</td>
<td>0.98</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>−0.70</td>
<td>−0.55</td>
</tr>
<tr>
<td>CAT2</td>
<td>0.96 (60)</td>
<td>0.92 (2)</td>
<td>0.92 (2)</td>
<td>0.95</td>
<td>−0.60</td>
<td>−0.47</td>
</tr>
<tr>
<td>GRIT1</td>
<td>0.92 (2)</td>
<td>0.93 (14)</td>
<td>0.95</td>
<td>−0.69</td>
<td>−0.55</td>
<td></td>
</tr>
<tr>
<td>GRIT2</td>
<td>0.92 (2)</td>
<td>0.93 (2)</td>
<td>0.95</td>
<td>−0.60</td>
<td>−0.55</td>
<td></td>
</tr>
<tr>
<td>KT1</td>
<td>−0.70</td>
<td>−0.74 (2)</td>
<td>−0.60</td>
<td>−0.69 (1)</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>KT2</td>
<td>−0.55</td>
<td>−0.61</td>
<td>−0.47</td>
<td>−0.55</td>
<td>0.86 (62)</td>
<td></td>
</tr>
</tbody>
</table>

*All values are significant at p < 10−9. Correlations below the diagonal show results after removing overlapping probesets (number of overlaps in parentheses). Correlations above the diagonal have not had overlaps removed. n = number of probesets in each PBT.

GRIT2 (0.95) and KT1 with KT2 (0.93). Moreover CATs correlated strongly with GRITs (0.93–0.94) and inversely with KT1 (−0.70, −0.74) and KT2 (−0.55, −0.61). The correlations remained high when all overlapping transcripts were eliminated (values below the diagonal in Table 1). The correlations indicate a strikingly coordinated structure of the inflammatory disturbances.

**PBT scores correlate with histopathologic lesions**

PBT scores correlated with the lesions of interstitial inflammation, tubulitis and vasculitis. Increasing lesion scores were associated with increasing CAT and GRIT scores and decreasing KT scores (Figure 1). However, transcripts did not distinguish i0 and i1 lesions, t1 and t2 lesions or v1 and v2 lesions. The lack of transcript differences between t1 and t2 was of particular interest because this is the threshold for histopathologic diagnosis of T-cell-mediated rejection. Table 2 reflects the relationship between transcript and lesion scores. Vasculitis scores correlated with CATs. Surprisingly C4d staining was associated with GRIT1 scores, indicating that antibody-mediated rejection is associated with Ifng production and effects. The KT loss correlated with i- and cv-scores.

**PBT scores correlate with histopathologic diagnosis, clinical episodes and treatment effects**

Biopsies with histopathologic diagnoses of rejection (T-cell-mediated, antibody-mediated or mixed) had elevated expression of CATs and GRITs, while biopsies without rejection (e.g. calcineurin inhibitor toxicity or interstitial fibrosis and tubular atrophy) showed little disturbance in PBTs (Table 3). PBTs were similar in antibody-mediated and T-cell-mediated rejection. CATs and GRITs were higher in biopsies from patients with clinical T-cell-mediated rejection episodes, compared to those lacking functional changes (Table 3). Antibody-mediated rejection with clinical episodes had greater transcript disturbances than those lacking episodes, although the differences were not significant. In cases with T-cell-mediated rejection, which received therapy before the biopsy, transcript changes were less than in untreated cases (Figure 1D).

**PBT scores reflect a stereotyped disturbance in the allograft**

Figure 2A illustrates transcript scores and the histopathologic lesions in all individual biopsies, sorted according to their CAT1 scores, illustrating how the CAT2, GRIT and KT scores were highly predictable from the CAT1 score. The degree of disturbance across all biopsies was continuous rather than dichotomous, with many other forms of renal injury having disturbances in PBTs similar to but at a lower level than the rejection cases. Biopsies with high tubulitis and interstitial inflammation scores and those classified as episodes had the greatest PBT disturbance. Biopsies with little PBT disturbance generally did not have rejection.

Table 2: Association between PBT scores and histopathology scores

<table>
<thead>
<tr>
<th>PBT</th>
<th>g&lt;sup&gt;a&lt;/sup&gt;</th>
<th>c&lt;sup&gt;a&lt;/sup&gt;</th>
<th>i&lt;sup&gt;b&lt;/sup&gt;</th>
<th>c&lt;sup&gt;i&lt;/sup&gt;</th>
<th>t&lt;sup&gt;i&lt;/sup&gt;</th>
<th>v&lt;sup&gt;a&lt;/sup&gt;</th>
<th>v&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ah&lt;sup&gt;a&lt;/sup&gt;</th>
<th>mm&lt;sup&gt;a&lt;/sup&gt;</th>
<th>C4d&lt;sup&gt;a&lt;/sup&gt;</th>
<th>R&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT2</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.005</td>
<td>0.05</td>
<td>0.003</td>
<td>g**, i**, t**, v*</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT1</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.006</td>
<td>0.06</td>
<td>0.005</td>
<td>g*, i**, t**, v*</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRIT2</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.01</td>
<td>&lt;0.001</td>
<td>g**, i**, t*, C4d*</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRIT1</td>
<td>&lt;0.001</td>
<td>0.07</td>
<td>&lt;0.001</td>
<td>0.08</td>
<td>0.05</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>i**, C4d*</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KT1</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
<td></td>
<td>i**, C4d*</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KT2</td>
<td>&lt;0.001</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td></td>
<td>i**, C4d*</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Due to small numbers with a score of 3; g, cg, ci, ct, cv, ah and mm lesions, samples with scores 2 and 3 were merged into one group.

<sup>b</sup>For the i score, samples were grouped 0/1, 2 and 3.

<sup>c</sup>For the t score, samples were grouped as 0, 1/2 and 3.

<sup>d</sup>Due to small numbers (no subjects had v = 2; only four subjects had v = 3), samples with scores 1–3 were merged into one group.

<sup>e</sup>C4d is either positive or negative.

<sup>f</sup>p < 0.05, **p < 0.01; empty cells = not significant with p > 0.1.
Table 3: Average PBT scores in diagnostic categories and comparison of each group to histologic or clinical T cell-mediated rejection (H-TCMR, C-TCMR)

| Diagnostic category (n) | CAT1 | | | | | | | | | | | | | | | | GRII1 | | | | | | | | | | KT1 | | | | | | | | | | p-Value¹ | p-Value² | p-Value¹ | p-Value² | p-Value¹ | p-Value² | p-Value¹ | p-Value² |
|------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Control Nephrectomies (8) | 0.00 ± 0.07 | < 0.001 | < 0.001 | 0.00 ± 0.10 | < 0.001 | < 0.001 | 0.00 ± 0.07 | < 0.001 | < 0.001 | 0.00 ± 0.03 | < 0.001 | < 0.001 | 0.00 ± 0.03 | < 0.001 | < 0.001 |
| T-cell-mediated rejection Borderline by histopathology (21) | 0.46 ± 0.10 | 0.15 | 0.007 | 0.67 ± 0.10 | 0.05 | < 0.001 | −0.15 ± 0.06 | 0.48 | 0.11 | 0.67 ± 0.10 | 0.05 | < 0.001 | −0.15 ± 0.06 | 0.48 | 0.11 |
| T-cell-mediated rejection by histopathology; H-TCMR (28) | 0.65 ± 0.09 | NA | 0.10 | 0.93 ± 0.09 | NA | 0.09 | −0.20 ± 0.04 | NA | 0.24 | 0.65 ± 0.09 | NA | 0.10 | 0.93 ± 0.09 | NA | 0.09 |
| Clinical rejection episode by histopathology (28) | 0.89 ± 0.11 | 0.10 | NA | 1.16 ± 0.10 | 0.09 | NA | −0.29 ± 0.06 | 0.24 | NA | 0.89 ± 0.11 | 0.10 | NA | 1.16 ± 0.10 | 0.09 | NA |
| T-cell-mediated rejection by histopathology; no clinical rejection episode (14) | 0.43 ± 0.07 | 0.06 | 0.002 | 0.70 ± 0.09 | 0.08 | 0.002 | −0.15 ± 0.05 | 0.37 | 0.07 | 0.43 ± 0.07 | 0.06 | 0.002 | 0.70 ± 0.09 | 0.08 | 0.002 |
| Antibody-mediated rejection by histopathology (8) | 0.60 ± 0.13 | 0.74 | 0.10 | 1.00 ± 0.13 | 0.65 | 0.36 | −0.21 ± 0.07 | 0.92 | 0.44 | 0.60 ± 0.13 | 0.74 | 0.10 | 1.00 ± 0.13 | 0.65 | 0.36 |
| Antibody-mediated rejection by histopathology; no clinical rejection episode (3) | 0.69 ± 0.20 | 0.87 | 0.41 | 1.05 ± 0.21 | 0.61 | 0.67 | −0.23 ± 0.12 | 0.88 | 0.67 | 0.69 ± 0.20 | 0.87 | 0.41 | 1.05 ± 0.21 | 0.61 | 0.67 |
| Mixed Antibody- & T-cell-mediated rejection by histopathology (8) | 0.74 ± 0.07 | 0.42 | 0.27 | 1.19 ± 0.04 | 0.01 | 0.69 | −0.20 ± 0.07 | 0.97 | 0.37 | 0.74 ± 0.07 | 0.42 | 0.27 | 1.19 ± 0.04 | 0.01 | 0.69 |
| Mixed Antibody- & T-cell-mediated rejection by histopathology; no clinical rejection episode (4) | 0.75 ± 0.05 | 0.35 | 0.28 | 1.24 ± 0.06 | 0.01 | 0.50 | −0.21 ± 0.11 | 0.97 | 0.58 | 0.75 ± 0.05 | 0.35 | 0.28 | 1.24 ± 0.06 | 0.01 | 0.50 |
| Mixed Antibody- & T-cell-mediated rejection by histopathology; no clinical rejection episode (4) | 0.73 ± 0.14 | 0.64 | 0.40 | 1.16 ± 0.04 | 0.02 | 0.98 | −0.19 ± 0.11 | 0.92 | 0.47 | 0.73 ± 0.14 | 0.64 | 0.40 | 1.16 ± 0.04 | 0.02 | 0.98 |
| Mixed Antibody- & T-cell-mediated rejection by histopathology; no clinical rejection episode (4) | 0.74 ± 0.07 | 0.42 | 0.27 | 1.19 ± 0.04 | 0.01 | 0.69 | −0.20 ± 0.07 | 0.97 | 0.37 | 0.74 ± 0.07 | 0.42 | 0.27 | 1.19 ± 0.04 | 0.01 | 0.69 |
| BK BK nephropathy (6) | 0.39 ± 0.13 | 0.14 | 0.01 | 0.70 ± 0.12 | 0.15 | 0.01 | −0.21 ± 0.08 | 0.95 | 0.45 | 0.39 ± 0.13 | 0.14 | 0.01 | 0.70 ± 0.12 | 0.15 | 0.01 |
| No rejection by histopathology, no BK (72) | 0.22 ± 0.03 | < 0.001 | < 0.001 | 0.42 ± 0.04 | < 0.001 | < 0.001 | −0.11 ± 0.02 | 0.04 | 0.007 | 0.22 ± 0.03 | < 0.001 | < 0.001 | 0.42 ± 0.04 | < 0.001 | < 0.001 |
| Other CNIT by histopathology (12) | 0.03 ± 0.04 | < 0.001 | < 0.001 | 0.16 ± 0.05 | < 0.001 | < 0.001 | −0.04 ± 0.03 | 0.002 | < 0.001 | 0.03 ± 0.04 | < 0.001 | < 0.001 | 0.16 ± 0.05 | < 0.001 | < 0.001 |
| T-cell-mediated rejection by histopathology; no clinical rejection episode (14) | 0.11 ± 0.08 | < 0.001 | < 0.001 | 0.28 ± 0.11 | < 0.001 | < 0.001 | −0.05 ± 0.06 | 0.06 | 0.01 | 0.11 ± 0.08 | < 0.001 | < 0.001 | 0.28 ± 0.11 | < 0.001 | < 0.001 |

¹p-Value for comparison of histopathology diagnosis of H-TCMR vs. PBTs.
²p-Value for comparison of clinical diagnosis of C-TCMR vs. PBTs.
³CNIT = calcineurin inhibitor toxicity.
⁴IFTANOS = interstitial fibrosis and tubular atrophy otherwise unspecified; NA = not applicable.
Biopsies with acute tubular necrosis but no rejection often had disturbances in PBTs.

**Using PBT scores to create a classifier for histopathologic diagnoses and clinical episodes**

PBT score classifiers built from the training set were applied to a test set to determine diagnostic accuracy in prediction of H and C diagnosis. We used PBT scores to construct classifiers to predict (i) histopathologic diagnosis of rejection (H-classifier) and (ii) clinically defined episodes (C-classifier). The classifier results are shown in Table 4. Validation statistics were calculated using (i) a resampling method where test set predictions based on training set classifiers were averaged over 100 training/test set splits of the data and (ii) leave-one-out cross-validation. The difference in results between the two approaches is small, indicating that the error estimates are robust to alternative validation methods. The probability of rejection derived from the H- and C-classifiers, built from the full dataset, is illustrated in Figure 2A.

**Validation of results**

To confirm the results of our analysis, we studied an independent validation set of biopsies for cause (Figure 2B). The results are expressed in the same form as Figure 2A and confirm our original findings: the PBTs are highly correlated with one another, with the lesions and diagnoses, and with the previously defined classifiers. In a second step, we repeated the validation analysis on the total set of 194 (143 biopsies from the original set plus 51 biopsies for cause from new samples). This analysis produced results similar to those in the original data set (Table 4).

**Cases with disagreement between PBT scores and histopathology**

Biopsies with low CAT scores seldom had rejection on histopathology, and most cases with high CAT scores had rejection by histopathology (Figure 2). We studied the cases in which the PBTs differed from the histopathologic diagnosis. In 14 cases with a histopathologic diagnosis of rejection (1 antibody-mediated, 13 T-cell-mediated) but with log2 CAT1 score <0.32, six had been treated before biopsy. Five were in an unusual group with v lesions with tubulointerstitial changes below Banff Ia T-cell-mediated rejection.

Five biopsies with log2 CAT1 >0.54 were read as no histopathologic rejection, four in situations where current histopathologic criteria are problematic (two with ‘borderline changes’ and two with severe atrophy and fibrosis, which is not readable by current criteria). One case had severe interstitial nephritis of possible bacterial origin.

Thus most disparities between PBTs and histopathology arise from either problematic aspects of histopathologic diagnosis (isolated vasculitis, borderline changes and severe fibrosis and atrophy) or from effects of treatment before the biopsy, which should be taken into consideration as a confounding factor.

**Confirmation by RT-PCR:**

Expression of selected transcripts (CD8A, GZMB, CXCL9, CXCL11 and FABP1) was confirmed by real time RT-PCR. Pearson correlation values were 0.86 for CD8A, 0.93 for CXCL9, 0.92 for CXCL11, 0.86 for GzmB and 0.44 for FABP1 compared to the microarray values.

**Discussion**

We used PBTs defined from mouse transplants and cell cultures (25–27) to translate microarray results into measurements of biological processes in human renal transplant biopsies. Thus the selection of genes was not driven or ‘trained’ by bioinformatic analysis of the human samples. PBT disturbances correlated with one another across the entire spectrum of diseases, as well as with histopathologic lesions, histopathologic diagnoses and clinical rejection episodes. Biopsies with antibody-mediated and T-cell-mediated rejection manifested similar disturbances, and PBTs were most disturbed when patients had clinical as well as histopathologic criteria for rejection episodes. Biopsies with minimal PBT disturbances had a very low incidence of rejection. The findings were confirmed in an independent validation set of 51 consecutive biopsies for cause. Thus PBTs emerge as a robust new quantitative tool for assessing rejection and inflammation. Moreover,
the PBTs reveal that previous biomarkers of rejection based on cytotoxic T cells or Ifng-inducible genes are qualitatively specific but are seen at lower levels in other types of renal inflammation.

PBTs provide a reproducible quantitative score without the inherent variability of histopathology, and will be a valuable addition to clinical assessment. Histopathology has severe problems with reproducibility: agreement of two pathologists on lesion scoring is 10–50% and on diagnosis is 45–70% on the same core (6,7,10). The microarray assessment is 98% reproducible on one sample, and the PBTs are 80–95% reproducible in multiple cores from one kidney (Supplementary Figure 3). We have analyzed the relative performance of microarrays and RTPCR in detail (manuscript in preparation). RTPCR has a greater dynamic range, and the array will thus miss some individual genes. However, the lack of sensitivity for individual transcripts is offset by the used sets of genes. The lower PBT scores in biopsies from treated rejection suggest that PBTs reflect treatment more quickly than does histopathology. Nevertheless, transcriptome analysis of biopsy cores cannot reflect events in small compartments such as glomeruli and arteries or deposition of proteins such as C4d. Interstitial fibrosis constitutes mature fibrous tissue, which is unlikely to be represented in the transcriptome. Similarly, the degree of tubular atrophy and interstitial fibrosis is probably not reflected in the transcriptome, as indicated by the lack of correlations between the KT score and ci and ct scores. We suspect that this is due to low transcriptional activity in these areas. The amount of total RNA hybridized to the array remains constant—if atrophic tubules have no transcripts, all transcripts are derived from nonatrophic tubules. Thus atrophic areas likely do not contribute to the transcriptome, neither in a positive nor in a negative way. Thus histopathology and gene expression analysis are complementary diagnostic approaches, and future diagnostic systems will incorporate both. As research tools, histopathology and gene expression are mutually instructive, as each can be used to guide understanding of the other.

While the PBT results independently validate histopathologic diagnoses, they also flag areas where current diagnostic criteria may be flawed. The PBTs were not different in lesions scored t1 versus t2, a noteworthy result given that this distinction is the threshold for diagnosis of T-cell-mediated rejection. In this respect, the PBTs confirm the conclusions of pathologists with regard to the low reproducibility of the t1–t2 distinction (7,10). This is not an isolated problem: assessment of quantitative changes in lesions as thresholds for treatment is problematic generally in histopathology scoring systems. In addition, relatively low PBT scores were observed in some biopsies with intimal arteritis with minimal tubulointerstitial changes that did not meet the Banff criteria for T-cell mediated rejection (TCMR) based on i and t lesions (i<2, t<2). In general, arterial inflammation in allografts is triggering aggressive therapeutic interventions. But the clinical significance of isolated arteritis with tubulointerstitial changes that do not meet the Banff criteria for TCMR based on i and t lesions has been challenged by other groups indicating that isolated-v cases without significant tubulitis and interstitial inflammation were correlating with a better response to treatment (16). Thus the present microarray results can be used as an independent standard to identify the areas where the current grading system needs to be reevaluated, and in many cases reinforces well-known concerns about the potential weak points in the histologic grading.

The high correlations among the PBTs indicate that distressed kidneys express varying degrees of a continuous, coordinate disturbance, in which thousands of genes—including many previously described rejection biomarkers—change their expression in concert. The discovery of the continuity of these measurements across many different diagnoses reflects the rigorous study design: unlike previous studies that ‘cherry-picked’ biopsies to represent disease classes, these consecutive unelected biopsies displayed a spectrum of PBT scores rather than distinct categories. The PBT profile was not altered in early versus late biopsies and seems to be independent of time posttransplant. (To some extent this is like histopathology, which also has its borderline cases.) The CATs and GRITs and KTs are disturbed coordinately in many types of renal inflammation, not only in T-cell-mediated rejection. Thus the PBT disturbance is quantitatively diagnostic for rejection and strongly correlated with lesions, but not qualitatively specific. This must affect our interpretation of mechanisms. Probably very few cognate events trigger a massive inflammatory disturbance trapping memory and effector T cells. Thus the disturbance we detect by the PBTs is actually a consequence of the cognate events that can be triggered by either TCMR or ABMR or at lower levels by other processes. We initially assumed that CATs and GRITs reflect antigen-stimulated effector T cells entering the tissue during T-cell-mediated rejection and producing Ifng. However, CATs and Ifng production are also features of effector memory T cells (29), which also enter inflamed sites nonspecifically. Kidney injury induces Ifng production, accounting for the GRITs in inflamed sites (26). Thus many changes previously described as specific for rejection—such as perforin (30), granzyme B (31), Fas ligand (30)—are actually features of the inflammatory compartment that is created in T-cell-mediated rejection but also in antibody-mediated rejection, and occurs at lower levels in many forms of kidney injury and inflammatory states. This observation was missed in some previous studies because their study designs deliberately excluded equivocal cases, creating the impression of dichotomous categories and qualitative differences.

In the same vein, the similarity between antibody-mediated and T-cell-mediated rejection, despite their differences in histopathology, probably arises because both produce inflammatory compartments with the same large-scale
disturbances, even though the actual mechanism of injury differs. Both antibody and antigen-specific effector T cells are probably generated in the secondary lymphoid organs, and recognize antigen in different sites in the transplant, creating distinct lesions but activating a similar inflammatory response. Each profoundly disturbs the tissue, creating an inflammatory compartment that recruits the inflammatory infiltrate and induces Ifng production, and alters renal genes as a nonspecific consequence of the antigen-specific injury mechanisms. Antibody may also recruit NK cells by virtue of Fc and complement receptors, and creates an inflammatory compartment that may trap effector memory T cells. Both effector memory T cells and NK cells express many CATs and can release Ifng. GRIT1 induction in antibody-mediated rejection may reflect activation of NK cells and may explain why, in multivariate analysis, expression of GRIT1 genes correlates with C4d staining. Thus the similarity of T-cell-mediated and antibody-mediated rejection lies in the fact that most of the large-scale transcriptome disturbance is the downstream consequence of rejection injury. In contrast, we are currently developing new gene sets that differentiate antibody-mediated rejection from T-cell-mediated rejection (32,33) (manuscript in preparation). Thus the disturbance described here (which we have dubbed ‘the disturbance in the force’) includes most of the transcriptome changes previously described in rejection and is quantitatively diagnostic for rejection, but is in fact a consequence of rejection. The more specific features of individual mechanisms will be revealed by their deviation from the stereotyped disturbance.

This first application of experimentally derived gene sets in human inflammatory disease illustrates how measurements derived from model systems can be successfully applied to human clinical states to provide both diagnostic and mechanistic insight. The PBTs defined and validated in mouse transplants that develop the same pathologic lesions as humans, as well as cell culture experiments, were used to analyze a patient population without training the microarray data on this population. This has the advantage that the selection of genes was based on the biology, not trained on the cases used to assess the genes. We focus on the major biological processes in our analysis that we know occur in human transplants: CTL infiltration, IFNG effects and epithelial deterioration. We acknowledge that there are many components of the rejection and injury process that we have not captured in this analysis. The fact that mice and human differ in aspects of these processes needs to be kept in mind, since new details will eventually emerge in all-human analyses. All of these issues will be the focus of new analyses. Our results confirm the striking similarity of the rejection processes in mouse and human, when the mouse model shares the same lesions as the human. By reflecting the intensity of inflammation and the parenchymal response to injury, PBTs estimate the probability of rejection with high-negative predictive value. But the specificity may be higher than current estimates, because many of the outliers may reflect inaccuracy of histopathology, rather than the errors in transcriptome assessment. Thus PBTs, and the current definition of the disturbance, represent a new window on disease pathogenesis and on the structure of inflammation in diverse disease states, with objective quantification that may be useful in clinical practice, clinical trials and research.

Acknowledgments

Special thanks and appreciation to Dr. Deborah James for her assistance with the manuscript; to Dr. Zija Jacaj for help with collection of the clinical data; and to Vido Ramassar, Anna Hutton and Sujatha Guttikonda for technical support. This research has been supported by funding and/or resources from Genome Canada, Genome Alberta, the University of Alberta, the University of Alberta Hospital Foundation, Alberta Advanced Education and Technology, Roche Molecular Systems, Hoffmann-La Roche Canada Ltd., the Alberta Ministry of Advanced Education and Technology, the Roche Organ Transplant Research Foundation, the Kidney Foundation of Canada and Astellas Canada. Dr. Halloran also holds a Canada Research Chair in Transplant Immunology and the Muttart Chair in Clinical Immunology. Dr. Mengel received a research stipend from the Dr. Werner Jackstädt Kidney Foundation.

References

Microarray Analysis of Rejection

The following supplementary material is available for this article:

Supplemental Methods

Table S1: Patient demographics.

Figure S1: Algorithms to derive pathogenesis-based transcript sets. (A) Cytotoxic T-cell-associated transcripts (CATs) were defined by high expression in lymphocytes (mixed lymphocyte cultures and cloned cultured cytotoxic T cells) and low expression in normal kidneys (25). In addition, mouse CATs (CAT1) had increased expression in day 5 mouse allografts (CBA into B6) compared to normal mouse (CBA) kidneys (n = 236). Human CATs (CAT2) are analogous to CAT1 but were primarily defined based on human data (high expression in human lymphocytes and low expression in normal kidneys) but were not filtered for high expression in rejecting kidneys (n = 382). (B) Ifng-effects in the allograft are represented by Ifng-dependent rejection-induced transcripts (GRITs)—GRIT1 (n = 68) and GRIT2 (n = 326), identified based on wild-type and Ifng-deficient mouse allografts (26). (C) Changes in the renal parenchyma are represented by two sets of renal genes that were identified by high expression in normal mouse kidney and low expression in mouse inflammatory cells: renal transcripts (KT1) (n = 1481) and the more restricted set of solute carriers (KT2) (n = 64) (27).

Figure S2: Reproducibility of gene expression in replicate cores. Both transcriptome analysis and histopathology are subjected to the sampling error inherent in 18-gauge needle biopsy cores. Two pathologists agree on the score of a specific lesion grading in only 10–50% and final diagnosis at best 45–60% when assessing the same core (6,7,10). The microarray assessment is 98% reproducible when repeated on the same sample. To assess the reproducibility of microarray analysis, we compared expression of the
transcripts in the PBT sets in two cores from the same kidney.

This material is available as part of the online article from: http://www.blackwellpublishing.com/doi/abs/10.1111/j.1600-6143.2007.02005.x

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.