American Journal of Transplantation 2006; 6: 150–160
Blackwell Munksgaard

doi: 10.1111/j.1600-6143.2005.01175.x

Noninvasive Discrimination of Rejection in Cardiac Allograft Recipients Using Gene Expression Profiling

M. C. Deng*, H. J. Eisen*, M. R. Mehra, M. Billingham, C. C. Marboe, G. Berry, J. Kobashigawa, F. L. Johnson, R. C. Starling, S. Murali, D. F. Pauly, H. Baron, J. G. Wohlgemuth, R. N. Woodward, T. M. Klingler, D. Walther, P. G. Lal, S. Rosenberg and S. Hunt for the CARGO Investigators

Drexel University College of Medicine, Philadelphia, Pennsylvania, USA (HJE), Columbia University, New York, NY, USA (MCD, CM, HB), University of Maryland, Baltimore, Maryland, USA (MM, FLJ), Stanford University School of Medicine, Stanford, California, USA (MB, GB, SH), University of California at Los Angeles, Los Angeles, California, USA (JK), Cleveland Clinic Foundation, Cleveland, Ohio, USA (RCS), University of Pittsburgh, Pennsylvania, USA (SM), University of Florida, Gainesville, Florida, USA (DP), XDx, Inc., South San Francisco, California, USA (JGW, RNW, TMK, SR, DW, PGL)
M. C. Deng and H. J. Eisen contributed equally to this work. Corresponding authors: H. J. Eisen, Howard.Eisen@DrexelMed.edu; Mario C. Deng, md785@columbia.edu*

Rejection diagnosis by endomyocardial biopsy (EMB) is invasive, expensive and variable. We investigated gene expression profiling of peripheral blood mononuclear cells (PBMC) to discriminate ISHLT grade 0 rejection (quiescence) from moderate/severe rejection (ISHLT $\geq$3A). Patients were followed prospectively with blood sampling at post-transplant visits. Biopsies were graded by ISHLT criteria locally and by three independent pathologists blinded to clinical data. Known alloimmune pathways and leukocyte microarrays identified 252 candidate genes for which real-time PCR assays were developed. An 11 gene real-time PCR test was derived from a training set (n = 145 samples, 107 patients) using linear discriminant analysis (LDA), converted into a score (0–40), and validated prospectively in an independent set (n = 63 samples, 63 patients). The test distinguished biopsy-defined moderate/severe rejection from quiescence (p = 0.0018) in the validation set, and had agreement of 84% (95% CI 66% C94%) with grade ISHLT $\geq$3A rejection. Patients $>$1 year post-transplant with scores below 30 (approximately 68% of the study population) are very unlikely to have grade $\geq$3A rejection (NPV = 99.6%). Gene expression testing can detect absence of moderate/severe rejection, thus avoiding biopsy in certain clinical settings. Additional clinical experience is needed to establish the role of molecular testing for clinical event prediction and immunosuppression management.

Key words: Allogeneic, biological markers, gene expression profiling, graft rejection, heart transplantation, immune response genes, immunologic, immunologic monitoring, transplantation.

Received 22 April 2005, revised and accepted for publication 7 October 2005

Introduction

The goal of care after cardiac transplantation is to prevent allograft rejection while minimizing immunosuppressive side effects (1,2). The gold standard of rejection surveillance in cardiac transplantation is endomyocardial biopsy (EMB). However, EMB is invasive, expensive, subject to sampling error, inter-observer variability (3–5), and causes morbidity (0.5–1.5%). Although noninvasive alternatives to EMB are clearly needed, methods such as echocardiography, ultrasonic myocardial back-scatter, radionuclide imaging, magnetic resonance imaging, intra-myocardial electrograms and multiparametric immune monitoring have been difficult to validate and implement (6–20).

As recirculating peripheral blood mononuclear cells (PBMC) may reflect earlier host responses than those at local sites, measurement of PBMC gene expression might provide useful diagnostic information and reduce the need for EMB in patients who are asymptomatic. Recent studies using microarray analysis (21) or real-time PCR analysis of cytokine genes (22) have suggested that gene expression measurements in PBMC may be correlated with cardiac allograft rejection. However, these single center studies are limited by the absence of methodology to recognize the imperfect 'gold standard' nature of EMB, which creates significant challenges for diagnostic development and validation study design and analysis (23,24). In addition, the absence of multicenter independent validation sets in both studies suggests the need for more extensive investigation.

Based on the assumption that a gene expression signature of immune activation and leukocyte trafficking would be detectable in recipient PBMC and reflect the rejection status of the donor allograft, we tested the hypothesis that

a gene expression test could discriminate ISHLT grade 0 rejection (quiescence) from moderate/severe (ISHLT grade ≥3A) rejection (nonquiescence).

Methods

Study design

After approval by local Institutional Reviews Boards at eight centers, all patients undergoing heart transplantation and providing informed consent were eligible for the Cardiac Allograft Rejection Gene Expression Observational (CARGO) study beginning in September 2001. Enrolled patients were followed at each clinical encounter with data collection including EMB, hemodynamics and/or echocardiography, immunosuppression, laboratory data and complications, which were captured in electronic clinical report forms. EMB slides were obtained from centers for interpretation by a panel of pathologists blinded to the clinical data.

The study was conducted in three phases (Figure 1A): (1) *candidate gene discovery* using a combination of focused genomic and knowledge-base approaches; (2) *diagnostic development* using PCR assays and rigorous statistical methods and (3) *validation* in a prospective and blinded study. Samples were selected and divided into a training set, used for candidate gene discovery and diagnostic development, and a set used for validation of the gene expression signature described below.

Data from an additional set of representative samples not used in any of the three phases were evaluated after the validation studies to estimate the negative predictive value (NPV) and positive predictive value (PPV) in the CARGO population.

Biopsy Samples

Biopsies performed by standard techniques were graded by local pathologists. A subset of biopsy samples, including all local grades 1B, 2, 3A and 3B and a representative set of grades 0 and 1A samples, were also graded by three independent ('central') pathologists blinded to clinical information. After an evaluation of the concordance of these biopsy grades by the four pathologists, criteria for selecting acute cellular *rejection* and *quiescent* samples were defined prior to developing and validating the classifier.

Blood Samples

PBMC were isolated from eight mL of venous blood using density gradient centrifugation (CPT, Becton-Dickinson). Samples were frozen in lysis buffer (RLT, Qiagen) within 2 h of phlebotomy. Total RNA was isolated from each sample (RNeasy, Qiagen).

The effects of processing time on gene assays were tested in PBMC isolated from six venous blood samples from each of nine donors. Samples were treated identically except the interval between blood draw and first centrifugation step was varied from 1 to 8 h. Any gene assays showing significant systematic variations across this time period were eliminated from the development process.

Candidate Gene Discovery

Microarray expression profiling

A custom microarray was designed using RNA sequences expressed in stimulated and resting human leukocytes (PCR Select, Clontech) and from publicly available sequence databases. A total of 7370 genes were represented by 50-mer oligonucleotides (Sigma) on a spotted custom microarray

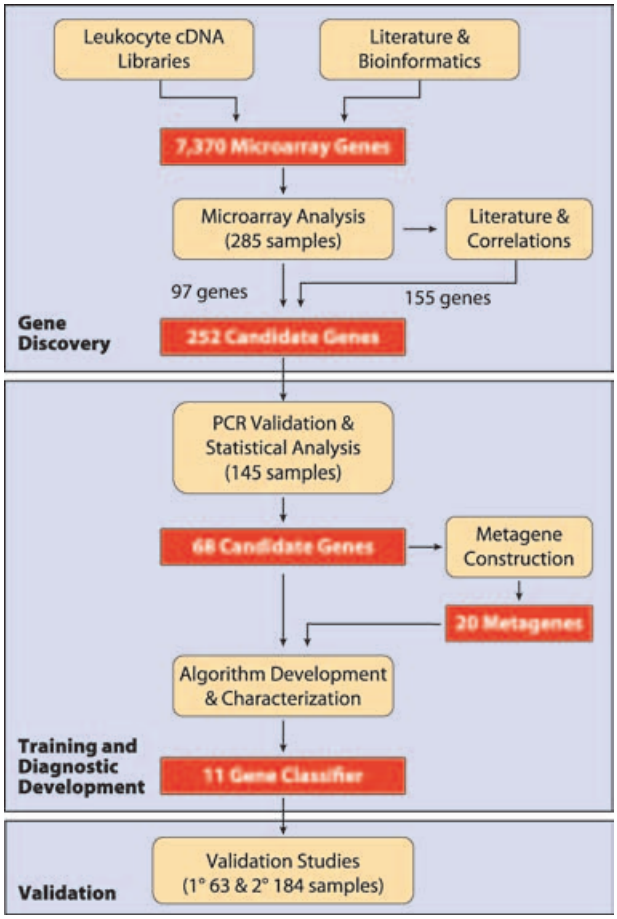

Figure 1: (A) Strategy used for the gene discovery, diagnostic development and validation. Initial discovery efforts using genomic (microarray) and knowledge-base (literature and sequence analyses) method produced 252 candidate genes for further quantitative PCR assay development. These 252 assays were applied to 145 samples to generate the dataset for training of the classifier. Statistical learning methods were used with these real-time PCR measurements to refine the gene set to 68 genes. Gene expression correlations were captured as composite metagenes. Automated statistical methods were used to build a classifier, and bootstrap and cross-validation methods to estimate classifier performance. The linear discriminant classifier developed was evaluated on an independent, blinded validation set from independent patients (primary validation set) and independent samples (secondary validation set).

(Telechem). To increase the power and quality of results, a large number of clinical samples (285) were used. Microarray data are available at GEO (25) with accession number GSE2445. The experimental methods are described in detail in the Supplement Section.

Knowledge-base gene discovery

Our leukocyte-focused genomic microarray approach was complemented with (1) a review of the literature on pathways involved in immune activation, recruitment and mobilization, in general, and solid organ transplant rejection, in particular; and (2) genes related to genes suggested to be significant by microarrays (by pathways and families).

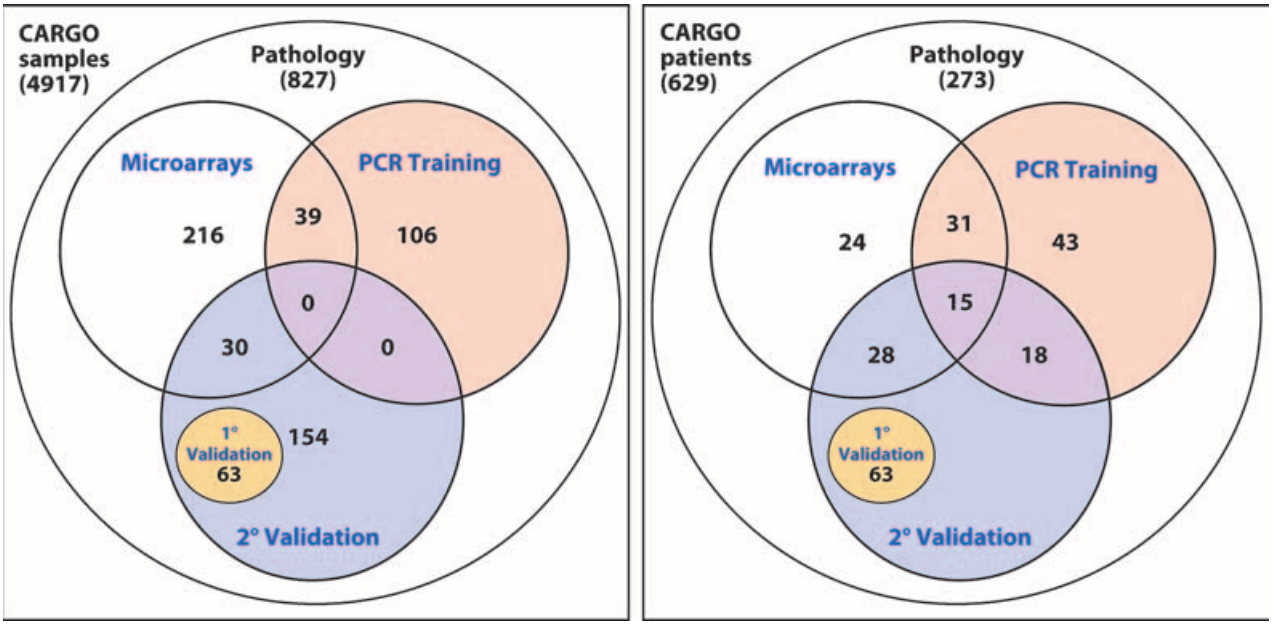

Figure 1: (B) Venn diagrams of samples and patients used for CARGO studies. A total of 827 samples were examined by centralized pathology. A set of samples and patients independent from both the microarray and PCR training studies are reported in the primary validation study, whereas an independent set of samples from the PCR training was used in the secondary validation study.

Diagnostic Development

Real-time PCR (RT-PCR)

PCR primers and probes were designed using PRIMER3 (version 0.9, Whitehead Research Institute). Assays were designed on the full-length mRNA after masking to avoid problematic sequences. Assays were qualified for inclusion in the training set by specificity, linear dynamic range and efficiency using both human PBMC cDNA and synthetic oligonucleotide templates. For each gene, triplicate 10 μL real-time PCR reactions were performed on the ABI 7900HT system using FAM-TAMRA probes and standard Taqman protocols (Applied Biosystems) on cDNA from 0.5 ng total RNA.

Normalization and control genes

Normalization genes were empirically selected using PCR data from the training samples. Genes which did not discriminate between rejection and quiescence samples with small standard deviations across all samples were considered as normalization genes. Six such genes spanning different expression levels were chosen.

Three additional assays were included as controls: two to detect genomic DNA contamination by the difference between a transcribed and nontranscribed region of the Gus-B gene and the third, a spiked-in control template for an Arabidopsis gene to determine if the PCR reaction was successful.

Discriminant equation development

RT-PCR data on 252 genes for the training set of 36 rejection and 109 quiescent CARGO samples were generated to derive a panel of candidate genes for classifier development and to validate microarray results. Gene expression results were analyzed with Student's *t*-test, median ratios, hierarchical clustering by TreeView and an expert assessment of biological relevance. Metagenes, defined as transcripts behaving in a concordant manner (26), were constructed by averaging gene expression levels that were correlated across training samples with correlation coefficients of at least 0.7. Genes significantly distinguishing rejection from quiescence in the PCR training set by *t*-test ($p \leq 0.01$), by median ratio differences of <0.75 or >1.25 or by correlation with significant genes were used for metagene construction and classifier development.

The methods for analyzing gene expression data included principal components analysis, linear discriminant analysis (LDA, StatSoft, Inc.), logistic regression (SAS Institute, Inc.), prediction analysis of microarrays (PAM) (27), voting, classification and regression trees (TreeNet, Salford Systems), Random Forests, nearest shrunken centroids and k-nearest neighbors. We sought to develop a classifier that quantitatively distinguished current moderate/severe acute cellular rejection (ISHLT grade ≥3A) from quiescence (ISHLT grade 0) using gene and metagene expression levels as the variables.

The final classifier was developed using LDA as implemented in the 'discriminant function analysis' module of Statistica (StatSoft, Inc.). LDA constructs a linear classifier by automatically selecting genes and/or metagenes that, in combination, optimally separate rejection and quiescent samples in the training set. The robustness of selected genes and the appropriate number of genes in the classifier were both evaluated by cross-validation.

Validation

Design

An independent cohort of CARGO patients was selected to validate the effectiveness of the LDA classifier defined in the diagnostic development phase using a prospective and blinded study protocol. The primary objective of the validation study was to test the pre-specified hypothesis that the diagnostic score distinguishes quiescence, defined as ISHLT grade 0, from moderate/severe biopsy-proven acute rejection, defined as ISHLT grade ≥3A, both grades determined from local and centralized cardiopathological examination. This was assessed using a 2-tailed Student's

t-test for comparing score distributions for rejection and quiescent samples. Secondary and exploratory objectives included documentation of diagnostic performance across thresholds and description of correlations to clinical variables.

Results for the validation study are reported for unique samples from patients not used for training (primary validation study), as well as for a larger set of samples not used for training (secondary validation study). These latter samples may provide improved power but may be biased to the extent that a longitudinal set of samples from an individual patient are not completely independent with respect to gene expression.

Prevalent Population Studies

A representative set of samples, across all local biopsy grades and ≥1 year post-transplant were evaluated to assess the discriminant equation performance on a stable patient population. From these samples, PPV (fraction of samples with scores at or above the threshold expected to have concurrent biopsy grade ≥3A) and NPV (fraction of samples with scores below the threshold expected to be free from biopsy grade ≥3A) were estimated at multiple test thresholds. Given the risk associated with undetected acute cellular rejection, and the clinical use of EMB, we sought a threshold that maximized the NPV at the expense of the PPV.

Quantitative CMV assays

Plasma was tested for quantitative CMV viral load using the COBAS protocol (Roche). These samples were selected from the CARGO study and represented known or suspected CMV infection and matched controls.

Results

Patients and samples used in these studies were selected from the CARGO database, with donor and recipient characteristics similar to those reported by the United Network for Organ Sharing (UNOS) for 2003 (28) (Table 1). The relationships between the samples and patients used in the three phases are shown in Figure 1B.

Gene discovery

In the gene discovery phase, 285 rejection and quiescent samples from 98 patients were hybridized to the leukocyte microarrays covering 7370 genes. Ninety-seven genes were selected as candidates for PCR assay development from these microarray studies based on false detection rates from SAM <20% (29), p-values in nonparametric analysis <0.05 or clustering with genes involved in rejection. This gene set was expanded to include related genes identified by correlated expression or related functions, as well as genes from the literature involved in transplant rejection, yielding an additional 155 gene candidates.

Diagnostic development

In the diagnostic development phase, 252 real-time PCR assays were developed to assess and confirm the discriminatory ability of the candidate genes from the gene discovery phase. These PCR assays were performed on 145 samples including 36 rejections (from 28 patients) and 109 quiescent samples (from 86 patients). Centralized pathology reading was used to identify these samples, where at least two of four pathologists were required to classify a sample as grade ≥3A for rejection, and three of four pathologists were required to classify a sample as grade 0 for quiescence. These criteria were set prospectively based upon centralized reading of over 800 CARGO samples and were used in the diagnostic development and validation PCR studies (30).

Analysis of this set of PCR data (see PCR-heatmap Figure 2A) yielded 68 genes that distinguished rejection samples from quiescent samples by *t*-test ($p < 0.01$), median ratio of (>1.25 or <0.75), or by correlation to discriminatory genes (Table 2). By hierarchical clustering (31) (Figure 2B), the predominant genes showing increased expression with rejection were T-cell/NK and $CD8^+$ T-cell activation markers (perforin, granulysin) and erythropoiesis markers (ALAS2, WDR40A, MIR). Six genes (CXCR4, hIAN7, HBG, CXCR3, ADM and TNFSF6) were eliminated due to significant variation in gene expression with sample processing time (32) yielding 62 genes for discriminatory signature development.

To take advantage of the correlations observed in gene expression (Figure 2B), 20 metagenes (26) were created by averaging correlated gene expression levels of the 62 genes. Using the training data set of 145 samples and these 82 variables (62 genes and 20 metagenes), a linear discriminant equation was derived by sequentially fitting the gene expression data to maximize agreement with the biopsy-based samples classification. The final equation, yielding a score between 0 and 40, combines the expression levels of four individual genes and three metagenes, constituting 11 genes in total (five from microarray and six from literature), which best distinguished rejection from quiescence (Table 2, Figure 2B). Additional terms did not further improve performance above approximately 75% correct classification. More complex statistical methods than LDA did not yield better performance and are less amenable to rigorous cross-validation as seen in other analyses (33).

Validation

In the validation phase, the discriminant equation performance was first estimated using the bootstrap method on the entire training set of samples. As shown in Table 3, agreement with biopsy ≥3A and biopsy grade 0, was estimated at 80% and 59%, respectively, at a single, predefined threshold of 20 (scores ≥20 indicate rejection).

To rigorously determine the test performance, an independent primary validation set of 63 unique samples (31 rejection, 32 quiescent samples) from 63 patients was tested in a prospective and blinded manner. The classifier distinguished moderate/severe rejection from quiescence (*t*-test, $p = 0.0018$). At the prospectively defined score threshold of 20, 84% (95% CI 66–94%) of

Table 1: Clinical characteristics of study patients and samples

			Microarray discovery			Diagnostic development & training			PCR 1° validation (unique patients)			PCR 2° validation (unique samples)		
	UNOS 2003	CARGO (N = 629 patients, 4917 samples)	Rejection (N = 28 patients, 38 samples)	No rejection (N = 94 patients, 247 samples)	p-value	Rejection (N = 28 patients, 36 samples)	No rejection (N = 86 patients, 109 samples)	p-value	Rejection (N = 31 patients, 31 samples)	No rejection (N = 32 patients, 32 samples)	p-value	Rejection (N = 50 patients, 62 samples)	No rejection (N = 83 patients, 122 samples)	p-value
Recipient age														
Under 18	14.0%	6.4%	0.0%	0.0%		0.0%	1.8%		6.4%	12.5%		3.2%	1.6%	
18–34	9.4%	9.9%	15.8%	13.0%		13.9%	5.5%		12.9%	9.4%		14.5%	13.9%	
35–49	21.1%	17.7%	23.7%	12.6%	NS	13.9%	21.1%	NS	3.2%	18.7%	0.006	17.7%	18.0%	NS
50–64	47.2%	53.1%	57.9%	65.6%		69.4%	56.9%		67.7%	31.2%		53.2%	54.9%	
65+	8.5%	12.9%	2.6%	8.9%		2.8%	14.7%		9.7%	28.1%		11.3%	11.5%	
Recipient race														
White	71.1%	72.3%	73.7%	74.1%		72.2%	78.9%		70.9%	65.6%		67.7%	66.4%	
Black	16.0%	17.3%	21.1%	15.4%		19.4%	10.1%		16.1%	15.6%		24.2%	18.9%	
Hispanic	8.8%	6.2%	5.3%	7.7%	NS	8.3%	6.4%	NS	6.4%	3.1%	NS	6.5%	9.0%	NS
Asian	1.9%	1.2%	0.0%	0.0%		0.0%	1.8%		3.2%	3.1%		0.0%	2.5%	
Other	2.1%	3.1%	0.0%	2.8%		0.0%	2.8%		0.0%	3.1%		1.6%	3.3%	
Recipient sex														
Male	73.6%	74.6%	92.1%	77.3%	NS	86.1%	73.4%	NS	74.2%	78.1%	NS	80.6%	81.1%	NS
Female	26.4%	25.4%	7.9%	22.7%		13.9%	26.6%		22.6%	12.5%		19.4%	18.9%	
Donor age														
Under 18	21.6%	17.9%	18.4%	11.4%		17.1%	17.5%		16.1%	21.8%		20.0%	17.1%	
18–34	44.2%	46.5%	47.4%	49.4%		51.4%	46.6%		45.2%	46.8%		38.2%	51.3%	
35–49	25.7%	24.9%	26.3%	28.2%	NS	22.9%	24.3%	NS	6.45%	15.6%	NS	32.7%	19.7%	NS
50–64	8.2%	10.5%	7.9%	11.0%		8.6%	11.7%		16.9%	9.3%		9.1%	11.1%	
65+	0.2%	0.2%	0.0%	0.0%		0.0%	0.0%		0.0%	3.1%		0.0%	0.9%	
Donor race														
White	69.6%	71.4%	73.7%	66.9%		75.0%	62.4%		86.6%	67.8%		83.9%	66.4%	
Black	12.0%	14.2%	5.3%	14.6%		5.6%	12.8%		10%	14.3%		8.1%	13.9%	
Hispanic	15.7%	11.4%	21.1%	16.3%	NS	19.4%	18.3%	NS	3.3%	14.3%	NS	8.1%	13.1%	NS
Asian	1.6%	0.7%	0.0%	2.1%		0.0%	0.9%		0.0%	0.0%		0.0%	0.8%	
Other	1.2%	2.3%	0.0%	0.0%		0.0%	5.5%		0.0%	3.5%		0.0%	5.7%	
Donor sex														
Male	68.4%	63.7%	71.1%	66.1%	NS	61.1%	68.5%	NS	53.3%	75%	NS	58.1%	74.2%	NS
Female	31.6%	36.3%	28.9%	33.9%		38.9%	31.5%		46.7%	25%		41.9%	25.8%	
Primary diagnosis														
Coronary artery disease	42.1%	23.8%	15.8%	21.9%		22.2%	32.1%		29.1%	34.4%		16.1%	30.3%	
Cardiomyopathy	47.0%	70.0%	81.6%	69.2%		77.8%	58.7%		61.3%	53.1%		79.0%	64.8%	
Congenital heart disease	8.5%	2.3%	0.0%	0.4%	NS	0.0%	1.8%	NS	0.0%	3.1%	NS	0.0%	2.5%	0.025
Retransplant	3.3%	0.5%	2.6%	5.3%		0.0%	1.8%		0.0%	0.0%		3.2%	0.8%	
Valvular disease	1.9%	2.4%	0.0%	1.2%		0.0%	1.8%		0.0%	0.0%		0.0%	0.8%	
Other	0.5%	1.0%	0.0%	2.0%		0.0%	3.7%		6.4%	0.0%		1.6%	0.8%	
Immunosuppression*														
Cyclosporine	64.9%	50.4%	71.1%	71.7%		52.8%	44.0%		58.1%	34.4%		72.6%	53.3%	
FK-506	42.9%	36.6%	26.3%	25.5%		47.2%	54.1%		38.7%	50%		25.8%	38.5%	
Mycophenolate	80.5%	72.0%	81.6%	87.4%		72.2%	78.0%		74.2%	75%		80.6%	83.6%	
Rapamycin	7.5%	9.8%	5.3%	2.4%	NS	22.2%	14.7%	NS	12.9%	6.3%	0.048	12.9%	8.2%	NS
Azathioprine	14.7%	1.4%	0.0%	0.4%		0.0%	1.8%		0.0%	3.1%		0.0%	1.6%	
Corticosteroids	91.1%	82.2%	97.4%	94.3%		94.4%	91.7%		77.4%	75%		88.7%	82.8%	
Zenapax	2.5%	9.4%	2.6%	4.5%		0.0%	7.3%		3.2%	3.1%		6.5%	4.1%	
Days post-Tx														
Average days post-Tx	NA	241	83	62		254	206		149	147		205	265	

NS = Not significant ($p \geq 0.05$).

*From UNOS 2001 data. This percentage represents the number of transplants in which a particular drug was used for maintenance at any point in the year after transplant divided by the number of transplants in 2001, and only accounts for patients with immunosuppressive information.

Comparison of clinical parameters of patients and samples used in the microarray, diagnostic development and validation studies.

CARGO = Cardiac Allograft Rejection Gene expression Observation study.

UNOS = United Network for Organ Sharing.

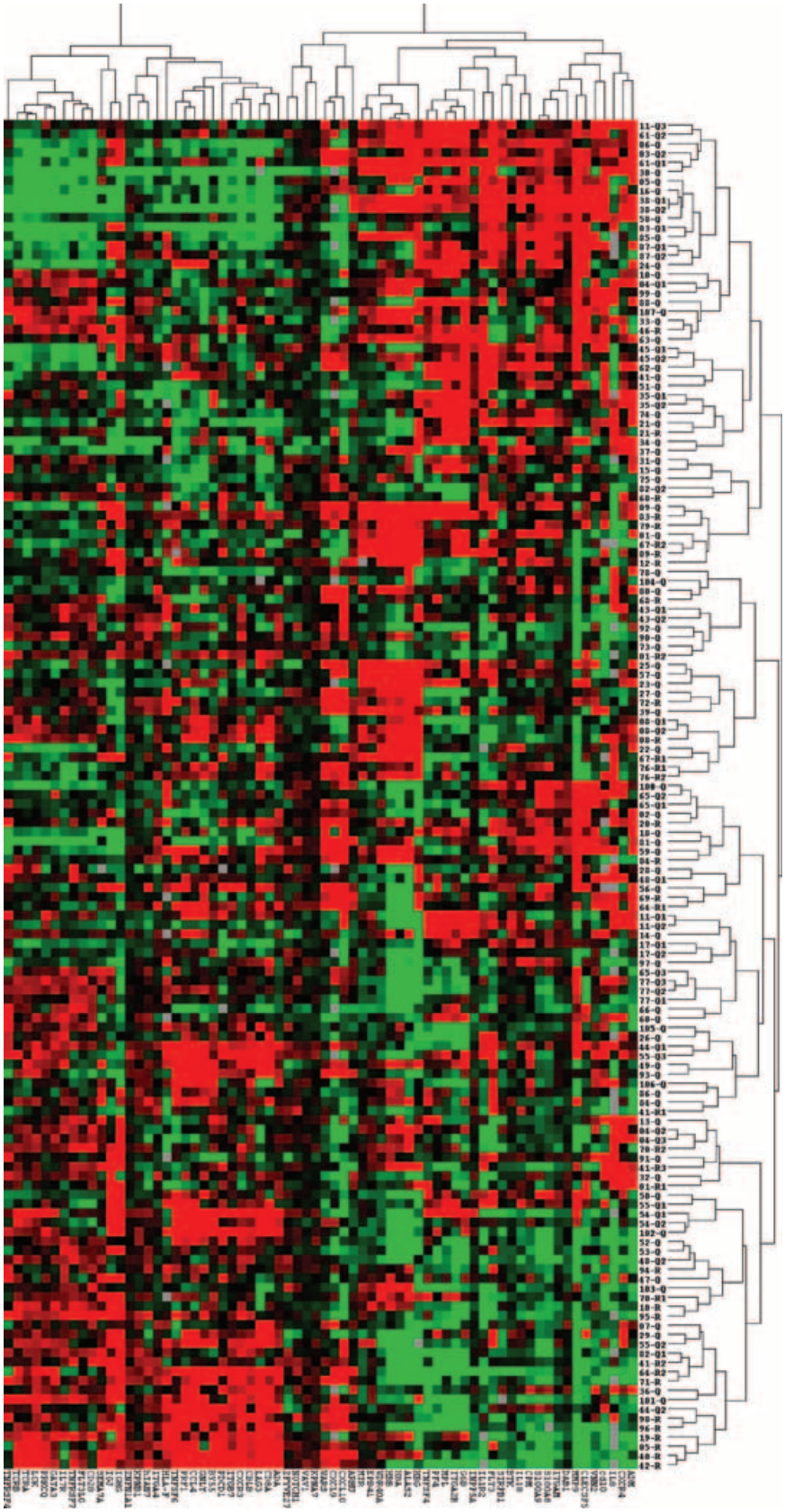

Figure 2: (A) Heat map of PCR expression measurements for 68 genes listed in Table 2 for 145 training samples. A heat map (red = up-regulated; green = down-regulated) for genes that significantly (*t*-test $p < 0.01$) differ between rejection and quiescence or have median ratios >1.25 or <0.75 or are correlated with significant genes is shown. Complete linkage clustering was used with Pearson correlation as the distance metric: genes (labeled by gene IDs) are clustered vertically according to correlated expression across samples; samples (labeled by patient no.— rejection (R) or quiescent (Q) and an ordinal if multiple samples from a single patients) are clustered horizontally according to correlated expression across genes.

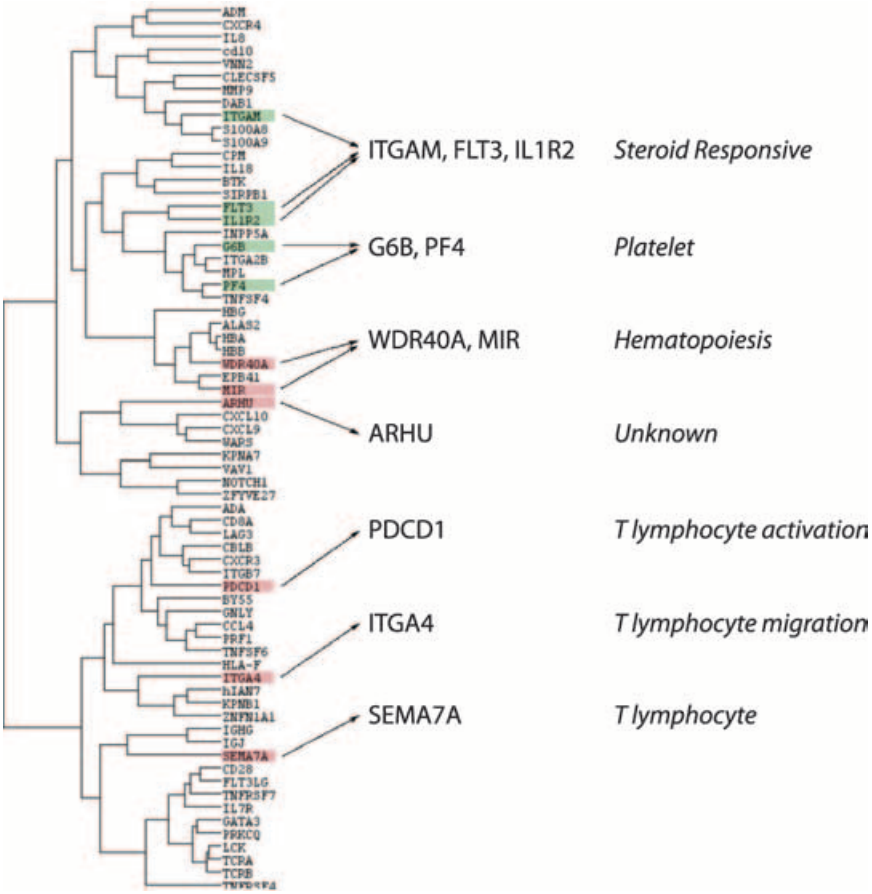

Figure 2: (B) Dendrogram of genes listed in Table 2, showing algorithm composition and cell type/function annotation. Gene correlation tree from heat map in Figure 2B expanded to show up- and down-regulated genes included in discriminant equation. The equation consists of a constant, 3 metagenes (of 3, 2 and 2 genes, respectively) and 4 single gene terms. Terms are annotated with functional, pathway or cell-type information where the supporting biology is known.

rejection and 38% (95% CI 22–56%) of quiescence samples were classified correctly. Receiver operating characteristics (ROC) analysis, shown in Figure 3A, yields an area under the curve of 0.72 ± 0.06. Similar results were obtained for the secondary validation set of 184 samples (62 rejection, 122 quiescent) from 124 patients that includes the 63 primary validation patient samples and additional samples from patients who contributed samples to the gene discovery or diagnostic development phases of this work (Figure 1B). With the same threshold of 20, 76% (95% CI 63–85% of rejections and 41% (95% CI 32–50%) of quiescent samples were classified correctly ($p = 0.0001$).

Analysis of the validation studies showed that time post-transplant was the single most important score-correlated variable. This time-dependence was responsible for the overall low specificity relative to biopsy using the single score threshold of 20. Scores increased with time post-transplant in association with the weaning of maintenance steroid doses, which generally occurs in the first year. Therefore, we investigated performance relative to biopsy in the ≥6 months and ≥1 year post-transplant periods (Table 3). In the ≥6 months period with a threshold of 28, 71% of rejection and 79% of quiescent secondary validation samples were classified correctly. For ≥1 year post-transplant, a threshold of 30 results in 80% of rejection and 78% of quiescent samples classified correctly. Similar improvements in performance in the primary validation study are observed, although the number of samples is small (Table 3). The areas under the ROC curves, shown in Figure 3B for the ≥6 month and ≥1 year secondary validation samples are 0.80 ± 0.14 and 0.86 ± 0.09, respectively.

Prevalent population studies

The validation study samples were highly enriched in rejection samples in order to more accurately estimate agreement with biopsy for this important but relatively rare class. In order to determine algorithm performance on the distribution of patients expected to be encountered in clinical practice, we tested 281 CARGO samples from 166 patients ≥1 year post-transplant, consisting of 160 (56.9%) grade 0, 68 (24.1%) grade 1A, 23 (8.1%) grade 1B, 21 (7.4%) grade 2 and 9 (3.2%) grade ≥3A, similar to that of the entire CARGO database. The grade 1B scores were significantly higher than grade 0 scores ($p = 0.0004$),

Table 2: Genes that discriminate between quiescence and rejection

Gene	GenBank	Discovery Source	PCR training *t*-test	PCR training Ratio
Programmed Cell death 1	PDCD1	Literature	1.6E–05	1.46
Semaphorin 7A	SEMA7A	Array	6.3E–05	1.29
Interleukin-1 receptor-soluble form	IL1R2	Array	6.4E–05	0.48
Importin alpha7	KPNA6	Array	1.1E–04	1.13
Chemokine (C-X-C motif) receptor 3	CXCR3	Literature	1.1E–04	1.32
Ikaros	ZNFN1A1	Literature	1.5E–04	1.15
Integrin beta 7	ITGB7	Literature	2.7E–04	1.21
Integrin alpha-M	ITGAM	Literature	3.6E–04	0.85
Chemokine (C-X-C motif) receptor 4	CXCR4	Array	4.6E–04	0.60
Matrix metalloproteinase 9	MMP9	Literature	6.1E–04	0.30
Vanin-2	VNN2	Array	6.4E–04	0.70
FLT3 ligand	FLT3LG	Literature	7.0E–04	1.27
CD160 NK cell receptor	BY55	Array	8.6E–04	1.21
Integrin alpha4	ITGA4	Literature	0.0011	1.18
Lymphocyte specific kinase	LCK	Literature	1.3E–03	1.27
T-cell receptor beta	TCRB	Literature	1.4E–03	1.30
Adenosine deaminase	ADA	Literature	0.0015	1.24
Adrenomedullin	ADM	Array	0.0016	0.70
Fms-like tyrosine kinase 3	FLT3	Literature	0.0020	0.60
Inositol polyphosphate-5-phosphatase	INPP5A	Array	0.0022	0.89
Lymphocyte activation gene 3	LAG3	Literature	0.0026	1.30
Fas Ligand	TNFSF6	Literature	0.0027	1.59
Signal regulatory protein beta-1	SIRPB1	Literature	0.0029	0.76
Carboxypeptidase M	CPM	Literature	0.0030	0.79
DAP12 associating lectin 1	CLECSF5	Literature	0.0030	0.72
Platelet factor 4	PF4	Literature	0.0032	0.74
Immune associated nucleotide receptor 7	hIAN7	Array	0.0032	1.26
Calgranulin A	S100A8	Array	0.0048	0.69
Guanine nucleotide exchange factor	VAV1	Literature	0.0057	1.08
Thrombopoietin receptor	MPL	Literature	0.0061	0.84
G6b Inhibitory receptor	G6b	Literature	0.0068	0.67
Ras homolog gene family, member U	ARHU	Array	0.0068	1.20
Notch homolog 1	NOTCH1	Literature	0.0073	1.11
Cas-Br-M (murine) ecotropic retroviral transforming sequence	CBLB	Literature	0.0081	1.15
T-cell transcription factor	GATA3	Literature	0.0095	1.15
T-cell R alpha	TCRA	Literature	0.0096	1.11
Calgranulin B	S100A9	Array*	0.0098	0.80
IL18	IL18	Literature	0.01	0.84
CD8A antigen	CD8A	Literature	0.01	1.13
Bruton's tyrosine kinase	BTK	Array	0.01	0.93
Granulysin	GNLY	Literature	0.02	1.29
CD28 antigen	CD28	Literature	0.02	1.13
Immunoglobulin J polypeptide	IGJ	Array*	0.02	1.49
erythrocyte membrane protein band 4.1	EPB41	Array*	0.02	1.18
Zn finger containing protein	ZFYVE27	Array	0.02	1.10
Cellular mediator of immune response	MIR	Array	0.03	1.14
Disabled homolog 1	DAB1	Array	0.03	0.93
Hemoglobin gamma	HBG	Array*	0.03	1.35
Integrin alpha2b	ITGA2B	Literature	0.03	0.64
CD10 antigen	CD10	Literature	0.04	0.68
Tumor necrosis factor receptor superfamily, member 7	TNFRSF7	Literature	0.04	1.24
PKC theta	PRKCQ	Literature	0.04	1.11
IgG heavy chain	IgHG	Array*	0.04	1.89
tumor necrosis factor (ligand) superfamily, member 4	TNFSF4	Literature	0.05	0.86
MIP-1-beta	CCL4	Array	0.07	1.27
Perforin	PRF1	Literature	0.08	1.32
OX40 receptor	TNFRSF4	Literature	0.08	1.32
Karyopherin beta 1	KPNB1	Array	0.10	1.01
WD40 motif bone marrow protein	WDR40A	Array	0.11	1.16
Tryptophanyl-tRNA synthetase	WARS	Array	0.14	1.27
Interleukin 8	IL8	Literature	0.20	1.62
Chemokine (C-X-C motif) ligand 10	CXCL10	Literature	0.22	1.35
Aminolevulinate, delta-, synthase 2	ALAS2	Array*	0.26	1.43
Chemokine (C-X-C motif) ligand 9	CXCL9	Literature	0.34	1.58
Hemoglobin alpha	HBA	Array*	0.36	1.36
Interleukin 7 receptor	IL7R	Literature	0.37	1.04
Major histocompatibility complex, class I, F	HLA-F	Array	0.42	1.27
Hemoglobin beta	HBB	Array*	0.62	1.32

*Correlated to significant array gene.

grade 1A scores ($p = 0.001$) and grade 2 scores ($p = 0.01$), but similar to grade ≥3A scores. The classifier scores and the prevalence of each biopsy grade in the CARGO database were used to estimate PPV and NPV. At a threshold of 30, the PPV is 6.8%, the NPV is 99.6%, and 68% of the tests are estimated to be below this value.

Discussion

This study tested the hypothesis that PBMC gene expression analysis using real time PCR for multiple genes and pathways detects the absence of moderate/severe acute cellular cardiac allograft rejection (ISHLT grade ≥3A) thus potentially reducing the frequency of graft biopsy in certain clinical settings. The multicenter validation study of independent patients showed that an 11 gene test distinguishes these states ($p = 0.0018$) and that patients ≥1 year

Table 3: Classifier performance in training, primary and secondary validation studies

			Biopsy grade 3A rejection				Biopsy grade 0 quiescence			
Sample set	Months post-Tx	Threshold	#Patients	#Samples	#Agree	%Agree	#Patients	#Samples	#Agree	%Agree
Training	All	20	29	36	29*	80.0%*	99	109	64*	59.0%*
1° Validation	All	20	31	31	26	83.9%	32	32	12	37.5%
2° Validation	All	20	50	62	47	75.8%	83	122	51	41.8%
1° Validation	>6	28	12	12	10	83.3%	14	14	10	71.4%
2° Validation	>6	28	19	21	15	71.4%	38	47	37	78.7%
1° Validation	>12	30	6	6	6	100%	7	7	4	57.1%
2° Validation	>12	30	10	10	8	80.0%	15	18	14	77.8%

*Bootstrap estimates.

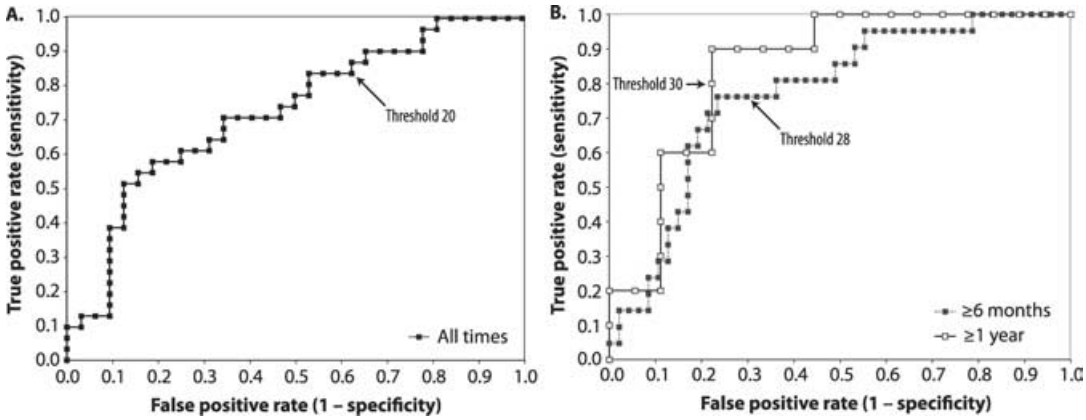

Figure 3: (A) Receiver operator characteristic (ROC) curve for primary validation cohort of 63 unique patient samples (31 rejection, 32 quiescent). Area under the ROC curve is 0.72 ± 0.06. **(B) ROC curves for secondary validation set of 184 patient samples (62 rejection, 122 quiescent) for the ≥6 months and ≥1 year periods.** Areas under the ROC curves are: 0.80 ± 0.114 for ≥6 months and 0.86 ± 0.09 for ≥1 year.

post-transplant with low molecular scores have a very low risk of current moderate/severe rejection (NPV >99%).

The specific genes identified as discriminating between quiescence (ISHLT grade 0) and moderate/severe rejection (ISHLT ≥ 3A) in PBMC-derived RNA encompass a wide variety of mechanisms and cell types (Table 2). Genes identified in T-cell mediated rejection in other organs have focused primarily on intragraft expression, with a more limited set of studies on peripheral fluids, including blood and urine. We found that both perforin and FasL/TNFSF6 as well as granulysin, were up-regulated in our study (see Table 2). All of these have been identified in renal allograft rejection studies by a variety of authors. In the study, which most closely resembles ours in terms of methodology and focus, PBMC gene expression was examined by RT-PCR of cytokine and cytokine receptor genes in heart allograft rejection (22). Of the four genes identified as most highly discriminatory in their study ($p < 0.01$), we also identified three as significant (CXCR3, FasL/TNFSF6 and perforin, Table 2), and we did not measure the other gene (COX2). Despite their significance, these genes were not included in the final LDA, because either they were not as robust as others or due to effects of sample processing on their expression. FasL/TNFSF6 and perforin cluster with PDCD1, the most discriminatory gene we identified (Table 2). PDCD1 has been identified as an important gene in an animal model of cardiac transplant rejection (34). CXCR3 and FasL/TNFSF6 were eliminated from consideration in the LDA, because expression of these genes is systematically dependent on sample processing time.

Although the primary endpoint of the study was achieved, important technical and clinical limitations of this study will have to be addressed to further evaluate the clinical role of this approach.

The derivation of the discriminant equation was critically dependent upon pathological classifications by multiple independent readers. The study design assumed a gold standard clinical endpoint of biopsy-based detection of rejection. However, the CARGO study demonstrated that this gold standard was limited by considerable inter-observer variability (30). Further work will be needed to assess how this variability is reflected in gene expression studies and clinical outcomes.

In the gene discovery phase, two approaches were taken: a focused leukocyte microarray and a knowledge-base or literature review, similar to that used to derive a validated

PCR-based test for breast cancer recurrence (35). Our goal was to find a set of genes that could be reproducibly measured by RT-PCR in a PBMC RNA preparation. The microarray approach used was limited to genes expressed in leukocytes potentially ignoring important genome-wide interactions. The knowledge-base approach, focusing on known genes, was ignoring new biology, which might be apparent in a nonhypothesis-driven approach. Although these complementary approaches led to the identification of genes that can distinguish the different rejection states, whole genome arrays may yield additional, different or better gene candidates. Moreover, additional studies may help to determine the basis for the differences observed between the microarray and PCR methods in significance of specific genes. These may be due to (i) lower sensitivity of microarrays leading to the elimination of genes which show discrimination in PCR, (ii) enhanced reproducibility of RT-PCR allowing measurement of small differences in gene expression (changes as small as 10–20%), likely undetectable by microarrays (usually eliminating genes that show <1.5- to 2-fold changes), (iii) use of a variety of classification definitions in the microarray analysis, which was focused to a single definition for the diagnostic development and validation phases.

Our microarray study results have to be interpreted in the context of recent insights from cancer biology-related array-based prediction studies (33). This reanalysis of seven studies showed that the list of genes identified as predictors of prognosis was highly unstable, and molecular signatures strongly depended on the selection of patients in the training sets. It highlights the primary challenge in using microarray results, where one is patient-limited and gene-rich: whether genes and signatures are truly significant or whether they are products of random variation (i.e. over-fitted to noise). We attempted to address two sources of variation—experimental (variability due to measurement technology) and biological (variability due to individual genetic and environmental differences). Experimental variation in microarrays was addressed by testing all candidate genes from the microarray discovery experiments with RT-PCR. Of the 97 candidate genes derived from microarray analyses, 27 were subsequently validated with RT-PCR (Table 2) and were used for further development. Biological variation was addressed in the RT-PCR data from the training set with statistical methods closely related to and more extensive than the suggested multiple random validation strategy (33), including multiple cross-validation and balanced bootstrapping techniques employed to calculate confidence intervals for performance measurements (details described in Supplement). Despite these results, it is still possible that some components of the LDA classifier are not robust; however, in using a two-technology, multiple-phase, cross-validated approach to validate discriminatory genes and the LDA signature, we have gone beyond most validation approaches previously described.

This study represents the first multicenter validation of a gene expression test in cardiac transplantation to identify patients at low risk for moderate/severe rejection. However, there are several limitations on the clinical implications of this work. Episodes of mild rejection on biopsy cannot be ruled out with this test. In addition, although the test captures patients with moderate/severe rejection, it has a low positive predictive value relative to biopsy. A nonquiescent score requires full workup including EMB to differentiate ongoing or impending cellular rejection, antibody-mediated rejection or chronic rejection/vasculopathy. Although CMV infection does not appear to confound the molecular signature (36,37), the impact of other infections on the test are not known. This study also did not address the predictive capacity of molecular testing for future rejection and clinical events, which could enable improved management in immunosuppressive therapy.

These study findings and limitations suggest several promising areas for future research including (A) more extensive gene discovery using whole-genome-based array approaches, (B) mechanistic hypothesis-driven research into individual genes and pathways identified and their intragraft role in acute rejection, (C) evaluation of the approach and specific genes and pathways in the setting of other solid organ transplantation. In summary, these results show that gene expression testing of blood cells can detect the absence of moderate/severe rejection, thus avoiding biopsy in certain clinical settings, but additional clinical experience is necessary to conclusively establish the predictive capacity of molecular testing for clinical events and its utility for monitoring immunosuppression.

Supplemental Material

The following supplemental material is available for this article online:

Appendix S1. Supplemental methods.
This material is available as part of the online article from http://www.blackwell-synergy.com

Acknowledgments

This study was supported in part by a grant from XDx, Inc. The authors wish to thank Drs Eric Topol (Cleveland Clinical Foundation), Jeff Bluestone (University of California, San Francisco), Ralph Snyderman (Duke University) and Fred Cohen (University of California, San Francisco) for their insightful reviews of the manuscript, Dr Thomas Quertermous (Stanford) for advice on scientific issues, and Drs Rob Tibshirani and Mike Walker (Stanford) for invaluable support and advice on statistical design and analysis and Dr Adriana Zeevi (Pittsburgh) for contributions to the study and review of the manuscript. The study was made possible through the hard work of the many CARGO investigators and researchers listed in the supplemental material.

References

1. Eisen HJ, Tuzcu EM, Dorent R et al. Everolimus for the prevention of allograft rejection and vasculopathy in cardiac-transplant recipients. N Engl J Med 2003; 349: 847–858.
2. Kobashigawa J, Miller L, Renlund D et al. A randomized active-controlled trial of mycophenolate mofetil in heart transplant recipients. Mycophenolate Mofetil Investigators. Transplantation 1998; 66: 507–515.
3. Nielsen H, Sorensen FB, Nielsen B, Bagger JP, Thayssen P, Baandrup U. Reproducibility of the acute rejection diagnosis in human cardiac allografts. The Stanford Classification and the International Grading System. J Heart Lung Transplant 1993; 12: 239–243.
4. Winters GL, McManus B. Consistencies and controversies in the application of the International Society for Heart and Lung Transplantation working formulation for heart transplant biopsy specimens. Rapamycin Cardiac Rejection Treatment Trial Pathologists. J Heart Lung Transplant 1996; 15: 728–735.
5. Winters GL, Marboe CC, Billingham M. The International Society for Heart and Lung Transplantation grading system for heart transplant biopsy specimen: clarification and commentary. J Heart Lung Transplant 1998; 17: 754–760.
6. Angermann CE, Nassau K, Stempfle HU et al. Recognition of acute cardiac allograft rejection from serial integrated backscatter analyses in human orthotopic heart transplant recipients. Comparison with conventional echocardiography. Circulation 1997; 95: 140–150.
7. Bourge R, Eisen H, Hershberger R et al. Noninvasive rejection monitoring of cardiac transplants using high resolution intramyocardial electrograms: initial US multicenter experience. PACE 1998; 21: 2338–44.
8. Deng MC, Erren M, Roeder N et al. T-cell and monocyte subsets, inflammatory molecules, rejection and hemodynamics early after cardiac transplantation. Transplantation 1998; 65: 1255–1261.
9. Eisen HJ, Eisenberg S, Saffitz JE, Bolman RM, Sobel BE, Bergmann S. Noninvasive diagnosis of cardiac transplant rejection using labeled lymphocytes. Circulation 1987; 75: 868.
10. Frist W, Yasuda T, Segall G et al. Noninvasive detection of human cardiac transplant rejection with indium-111 antimyosin (Fab) imaging. Circulation 1987; 76: 81–85.
11. Hammer C, Klanke D, Lersch C et al. Cyoimmunologic monitoring (CIM) for differentiation between cardiac rejection and viral, bacterial, or fungal infection: its specificity and sensitivity. Transplant Proc 1989; 21: 3631–3633.
12. Hummel M, Dandel M, Knollmann F et al. Long-term surveillance of heart-transplanted patients: noninvasive monitoring of acute rejection episodes and transplant vasculopathy. Transplant Proc 2001; 33: 3539–3542.
13. Kimball P, Radovancevic B, Isom T, Frazier B, Kerman R. Cytokine panel predicts early rejection and therapeutic response after cardiac transplantation. Transplant Proc 1995; 27: 1286–1287.
14. Knosalla C, Grauhan O, Muller J et al. Intramyocardial electrogram recordings (IMEG) for diagnosis of cellular and humoral mediated cardiac allograft rejection. Ann Thorac Cardiovasc Surg 2000; 6: 89–94.
15. Masuyama T, Valantine HA, Gibbons R, Schnittger I, Popp R. Serial measurement of integrated ultrasonic backscatter in human cardiac allografts forthe recognition of acute rejection. Circulation 1990; 81: 829–839.
16. Melter M, Exeni A, Reinders ME et al. Expression of the chemokine receptor CXCR3 and its ligand IP-10 during human cardiac allograft rejection. Circulation 2001; 104: 2558–2564.
17. Park JW, Warnecke H, Deng M, Schueler S, Heinrich KW, Hetzer R. Early diastolic left ventricular function as a marker of acute cardiac rejection: a prospective serial echocardiographic study. Int J Cardiol 1992; 37: 351–359.
18. Pfitzmann R, Muller J, Grauhan O, Hetzer R. Intramyocardial impedance measurements for diagnosis of acute cardiac allograft rejection. Ann Thorac Surg 2000; 70: 527–532.
19. Smart FW, Young JB, Weilbaecher D, Kleiman NS, Wendt RE 3rd, Johnston D. Magnetic resonance imaging for assessment of tissue rejection after heterotopic heart transplantation. J Heart Lung Transplant 1993; 12: 403–410.
20. Warnecke H, Muller J, Cohnert T et al. Clinical heart transplantation without routine endomyocardial biopsy. J Heart Lung Transplant 1992; 11: 1093–1102.
21. Horwitz PA, Tsai EJ, Putt ME et al. Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. Circulation 2004; 110: 3815–3821.
22. Schoels M, Dengler TJ, Richter R, Meuer SC, Giese T. Detection of cardiac allograft rejection by real-time PCR analysis of circulating mononuclear cells. Clin Transplant 2004; 18: 513–517.
23. Walter SD, Irwig L, Glasziou P. Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol 1999; 52: 943–951.
24. Zhou X, Obuchowski N, McClish D. Statistical Methods in Diagnostic Medicine. 2002, New York: John Wiley & Sons.
25. GEO: http://www.ncbi.nlm.nih.gov/geo/
26. Brunet JP, Tamayo P, Golub TR, Mesirov J. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004; 101: 4164–4169.
27. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002; 99: 6567–6572.
28. www.unos.org
29. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A, 2001; 98: 5116–5121.
30. Marboe CC, Billingham M, Eisen H et al. Nodular endocardial infiltrates (Quilty Lesions) cause significant variability in the diagnosis of ISHLT rejection grades 2 and 3A in endomyocardial biopsies from cardiac allograft recipients. J Heart Lung Transplant 2005; 24: S219–S226.
31. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998; 95: 14863–14868.
32. Baechler EC, Batliwalla FM, Karypis G et al. Expression levels for many genes in human peripheral blood cells are highly sensitive to ex vivo incubation. Genes Immunology 2004; 5: 347–353.
33. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005; 365: 488–492.
34. Ozkaynak E, Wang L, Goodearl A et al. Programmed death-1 targeting can promote allograft survival. J Immunol 2002; 169: 6546–6553.
35. Paik S, Shak S, Tang G et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351: 2817–2826.
36. Deng MC, Mehra M, Hunt S et al. Leukocyte gene expression signature of CMV viremia in cardiac allograft recipients is distinct from that seen during allograft rejection. Am J Transplant, 2004. 4e: 455.
37. Jenner RG, Young R. Insights into host responses against pathogens from transcriptional profiling. Nat Rev Microbiol 2005; 3: 281–294.