Published in final edited form as:
Circulation. 2011 March 22; 123(11): 1174–1184. doi:10.1161/CIRCULATIONAHA.110.002857.

Transcriptomic Biomarkers for the Accurate Diagnosis of Myocarditis

Bettina Heidecker, MD[*], **Michelle M. Kittleson, MD, PhD**[†‡], **Edward K. Kasper, MD**[†], **Ilan S. Wittstein, MD**[†], **Hunter C. Champion, MD, PhD**[†], **Stuart D. Russell, MD**[†], **Ralph H. Hruban, MD**[†], **E. Rene Rodriguez, MD**[‡], **Kenneth L. Baughman, MD**[††,§], and **Joshua M. Hare, MD**[*]

[*]University of Miami, Miller School of Medicine, Baltimore [†] The Johns Hopkins Hospital, Baltimore [‡]Cleveland Clinic Foundation, Cleveland [†‡]University of California, Los Angeles [††]Brigham and Women's Hospital, Boston

Abstract

Background—Lymphocytic myocarditis is a clinically important condition that is difficult to diagnose and distinguish. We hypothesized that the transcriptome obtained from an endomyocardial biopsy (EMB) would yield clinically relevant and accurate molecular signatures.

Methods and results—Microarray analysis was performed on samples from patients with histologically proven lymphocytic myocarditis (n=16) and idiopathic dilated cardiomyopathy (IDCM, n=32) to develop accurate diagnostic transcriptome-based biomarkers (TBB) using multiple classification algorithms. We identified 9,878 genes differentially expressed in lymphocytic myocarditis vs. IDCM (FC>1.2, FDR<5%), from which a TBB containing 62 genes was identified, which distinguished myocarditis with 100% sensitivity (95% CI: 46-100%) and 100% specificity (95% CI: 66-100%) and which was generalizable to a broad range of secondary cardiomyopathies associated with inflammation (n=27), ischemic cardiomyopathy (n=8) and the normal heart (n=11). Multiple classification algorithms and quantitative realtime RT-PCR analysis further reduced this subset to a highly robust molecular signature of 13 genes, which still performed with 100% accuracy.

Conclusions—Together these findings demonstrate that transcriptomic biomarkers from a single EMB can improve the clinical detection of patients with inflammatory diseases of the heart. This approach advances the clinical management and treatment of cardiac disorders with highly variable outcome.

Keywords

Gene expression; heart failure; myocarditis; transcriptome; biomarker

Corresponding author: Joshua M. Hare, MD Interdisciplinary Stem Cell Institute University of Miami Miller School of Medicine Biomedical Research Building / Room 824 P.O. Box 016960 (R-125) Miami, FL 33101 USA Tel: (305) 243-5579 Fax: (305) 243-5584 jhare@med.miami.edu.
[§]Deceased

For detailed methods see online supplement.

DISCLOSURES
None.

INTRODUCTION

The myocardites are inflammatory diseases of the heart that have variable clinical presentations and are caused by a range of underlying inflammatory variants[1, 2]. Of new onset heart failure (HF), 10-30% may be caused by cardiac inflammation, and viral infection[3, 4] systemic or local inflammatory diseases, or genetic predisposition represent inciting factors[5-7]. Myocarditis can be difficult to diagnose requiring multiple endomyocardial biopsies (EMBs)[8-11]. Even with multiple biopsies, consensus among pathologists has been difficult to attain[12]. Inaccurate or uncertain diagnosis is of major concern, since emerging therapies specifically targeting inflammatory or viral heart disease, have the potential to reverse the disease process[11, 13-15]. In a previous decision analysis investigating the value of EMBs to improve clinical outcome with specific therapy, histological inaccuracy was a major limiting factor for treatment efficacy[11].

Current attempts to improve diagnostic accuracy include screening for viral RNA in EMBs[16, 17], serum anti-heart autoantibodies[16], and use of magnetic resonance imaging (MRI)[18, 19]. Transcriptomics has emerged as highly valuable tool for complex pathologic diagnosis. Examples include delineation of childhood tumors[20], determination of organ rejection[21, 22], and delineation between ischemic and non-ischemic heart disease[23]. Based upon recent findings indicating that a single EMB contains sufficient RNA to perform a microarray without amplification[24, 25], we sought to test the hypothesis that the transcriptome could be used to create biomarkers that add diagnostic accuracy to clinical, pathological and imaging modalities currently used to diagnose myocarditis.

METHODS

Study population

We performed transcriptomic analysis of EMBs in matched cohorts of patients with IDCM (n=32) and myocarditis (n=16) selected from a biorepository containing samples from patients with new onset HF (n=350). Similarity of baseline conditions was tested with student's t test and fisher exact test. There was no difference between the two groups. Four to six biopsy specimens were obtained from each patient and examined by an experienced cardiac pathologist. Myocarditis was defined according to Dallas criteria[26, 27] while IDCM was a diagnosis of exclusion. If the diagnosis was equivocal based on standard histology, special stains were performed, such as immunofluorescence for IgG, IgM, IgA, C1q, C3d, C4d, fibrinogen, stains for AFB, fungi, elastosis, glycogen or iron accumulation.

One biopsy sample from each patient, obtained independently from the histological samples, was flash frozen and stored in liquid nitrogen for microarray analysis. A total of 115 biopsy samples were included for microarray analysis in this study, of which 81 samples were newly processed, and 34 samples from a previous study were included for validation [23]. Forty-eight samples were selected for our first transcriptomic study, including samples from patients with myocarditis (n=16) [26, 27] and IDCM (n=32) selected in a case-control fashion. In addition, samples from 6 patients with myocarditis and divergent baseline criteria were used for independent validation of the TBB. Furthermore, we tested the ability of the biomarker to detect active myocardial inflammation in patients with secondary cardiomyopathies associated with myocarditis (n=27). This group included patients with stress induced cardiomyopathy (Takotsubo) (n=4), sarcoidosis (n=9), peripartum cardiomyopathy (n=6), arrhythmogenic right ventricular dysplasia (ARVD, n=3), giant cell myocarditis (n=3) and systemic lupus erythematosus (SLE, n=2). Finally, we tested the transcriptomic biomarker for myocarditis in samples from a previous study[23], which included samples from patients with normal hearts (n=11), ischemic cardiomyopathy (n=8) and IDCM (n=15), and analyzed them with a prototype microarray, the Affymetrix U133A

Gene Chip. By using this approach, we evaluated generalizability of the molecular signature to various heart conditions, tested its performance in hearts free of disease, and evaluated its intraplatform reproducibility.

Transcriptomic analysis

Total RNA was extracted and hybridized as previously described[24, 25]. Microarray data was normalized with Robust Multiarray Average[28] and analyzed with Significance Analysis of Microarrays (SAM)[29] to identify differentially expressed genes in patients with myocarditis (n=16) vs IDCM (n=32). The resulting gene list was further processed with Meta Core pathway analysis from GeneGo Inc. In order to determine the minimum number of differentially expressed genes required for detection of patients with myocarditis, we used PAM[20]. The nearest shrunken centroid classifier was developed from a train set (n=33), consisting of 2/3 of data, and applied to an independent test set (n=15) containing 1/3 of data[20].

After developing the TBB with a case-control design, we tested its performance in unmatched samples (n=6) with higher ejection fractions (65± 4.7%) to evaluate generalizability.

In order to test, if previously established classification algorithms can further reduce the number of genes necessary for accurate prediction, we applied MiPP, a novel classification software package [22]. We subsequently applied the following classification rules, implemented in the MiPP package: SVM-rbf, SVM-lin, qda, lda and combination of lda, qda and svm-rbf. Models were developed based upon 5-fold cross validation in a train set (2/3 of data) and subsequent validation in an independent test set (1/3 of data).

In order to evaluate, if distinct models are generated from additional random splits, we performed 50 random divisions to develop individual classification models, which were then validated in 200 independent splits. In addition, we performed PCA to illustrate how well patients with myocarditis can be separated from patients with IDCM based on the original 62 genes molecular signature, and to test if genes that we identified by MiPP analysis to be the most robust classifiers, would also be discovered to be important when PCA was applied. PCA depicts highly robust classifiers with vectors having their endpoints far from the center.

Validation of microarrays with quantitative realtime RT-PCR

Validation with realtime RT-PCR was performed in a randomly selected subset of patients (IDCM: n=10, myocarditis: n=10), with triplicates replication. First-strand cDNA was synthesized from 100ng total RNA and amplified with MessageAmp II Amplification Kit. Importantly, this amplification step was only performed on validation samples, after the original biomarker was developed from pure total RNA that did not undergo any amplification, in order to eliminate any possibility of amplification bias that may impact the resulting molecular signature. TaqMan probes were designed for a subset of 13 candidate genes from microarray analysis: CD14, FCER1G, TLR1, TLR2, TLR7, ITGB2, SIGLEC 1, ADCY7, MEGF9, PTPLAD1, SWAP70, MSI1, and LCE1E, as well as the housekeeping gene 18S RNA. Finally, the results from RT-PCR were illustrated as a heatmap created with unsupervised hierarchical clustering based on Euclidean distance.

RESULTS

Table 1 depicts baseline clinical variables of patients of the selected case-control population with IDCM and Dallas criteria[26, 27] defined lymphocytic myocarditis.

Phenotype specific differences in gene expression

To identify differential gene expression between patients with IDCM (n=32) vs lymphocytic myocarditis (n=16), we used oligonucleotide microarrays to analyze RNA obtained from EMBs from affected patients at first presentation with new onset HF. We identified 9,878 differentially expressed genes (q<5%, fold change (FC >1.2) in patients with IDCM compared to myocarditis (figure 1). Transcripts with FC>2 (141 over-expressed and 16 down-regulated transcripts) are provided as supplemental tables 1 and 2. Pathway analysis with GeneGo Metacore revealed overexpression of 8 networks in myocarditis vs IDCM (supplemental table 3).

Molecular signature to distinguish myocarditis from non-inflammatory cardiomyopathy

We applied prediction analysis of microarrays (PAM) in a training set containing 2/3 of data (IDCM: n=22; myocarditis: n=11) and evaluated its accuracy in an independent test set, containing 1/3 of data (IDCM: n=10; myocarditis: n=5). The developed transcriptomic diagnostic biomarker consisted of a minimal set of 62 transcripts (table 2). When the molecular signature was tested in matched independent samples (n=15), it performed with 100% accuracy (sensitivity: 100%, 95 CI: 46-100%; specificity: 100%, 95 CI: 66-100%; positive predictive value, PPV: 100%, 95 CI: 46-100%; negative predictive value, NPV: 100%, 95 CI: 66-100%; figure 2). All samples were predicted correctly, independent of degree of inflammation – borderline or active myocarditis.

We next tested the transcriptomic biomarker in an additional set of independent samples derived from patients with myocarditis (n=6), who presented with higher ejection fractions (65±4.7%), compared to the case-control samples. In this group, the molecular signature still identified 83% of patients with myocarditis correctly (sensitivity: 91%, 95 CI: 57-100%; specificity: 100%, 95 CI: 66-100%; PPV: 100%, 95 CI: 66-100%; NPV: 91%, 95 CI: 57-100%, data not shown).

Performance of predictive algorithm in secondary cardiomyopathy/myocarditis

To evaluate generalizability in an additional relevant population, we applied the transcriptomic biomarker to biopsies from patients with secondary cardiomyopathies associated with myocarditis (stress induced cardiomyopathy n=4, sarcoidosis n=9, peripartum cardiomyopathy n=6, ARVD n=3, giant cell myocarditis n=3 and SLE n=2). In this setting, the biomarker distinguished myocarditis with a similar accuracy to that of idiopathic myocarditis (sensitivity: 100%, 95 CI: 46-100%; specificity: 95%, 95 CI: 75-100%; PPV: 83%, 95 CI: 36-99%; NPV: 100%, 95 CI: 80-100%, figure 3). Among this set of secondary cardiomyopathies, five biopsies were found to contain significant inflammatory changes based on immunohistochemistry, of which one of them was from a patient with stress induced cardiomyopathy (sample #109), one from a patient with SLE (sample #76) and three from patients with giant cell myocarditis. Indeed, all samples were correctly identified as inflammatory cardiomyopathy, while in the remaining samples, the molecular signature successfully ruled out inflammatory disease with very high accuracy. There was only one patient with sarcoidosis (sample #113) that got misclassified.

In addition, we evaluated the biomarker performance in patients from a previous data set (n=34)[23] containing samples with ischemic cardiomyopathy (n=8), IDCM (n=15) and normal heart (n=11): all samples were correctly classified.

Additional novel classification strategies

In order to obtain a parsimonious molecular signature we first applied multiple established classification algorithms using the misclassification-penalized posteriors classification (MiPP) package in R that includes lineal discriminant analysis (lda), quadratic discriminant

analysis (qda), supervector machine with radial basis function (svm-rbf), and supervector machine with lineal function as kernel (svm-lin). When applied to the 62 gene signature, these algorithms identified a highly diagnostic set of three transcripts (mean error of 0.167 in independent validation sets, n=18). Table 3 contains the mean error for each established set of genes developed by individual rules or combination of rules.

We continued our analysis by testing if a different random split of data would reveal distinct models. Splitting of data into train (2/3) and test set (1/3) and selecting a model for a given split were repeated 50 times. KRT78, MSI1, POU4F1, LCE1 and the EST 1556507_at resulted as top classifiers (mean error 0.086 after validation in 200 independent splits, table 4). As an additional measure for performance of a given gene model, we evaluated mean sMiPP, a parameter that approximates 1 with increasing accuracy. When the top 5 gene models (table 4) were validated in 200 independent random splits, mean sMiPP ranged from 0.776 - 0.791 (table 4). Since those models were built from 50 initial random splits, it is likely that identical gene clusters are identified in subsequent splits, as it occurred in our analysis (table 4: split #17 and split #45). Principal components analysis (PCA) is a valuable tool to illustrate importance of individual genes for classification of their corresponding phenotype. In agreement with results from our MiPP analysis, the transcripts 1556507_at, KRT78, LCE1E, MSI1 and POU4F1 were identified as highly important, with vectors having their endpoints distant from the center (figure 4.a). Additional highly robust transcripts were ITGB2, HERC6, ADCY7, NEK3, MEGF9, as well as the ESTs 1558605_at and 1565662_at (data not shown).

In addition, PCA clustered patients with similar expression patterns as one principal component (PC). As shown in figure 4.b, samples from patients with myocarditis noticeably separated from patients with IDCM.

Validation with quantitative realtime RT-PCR

To obtain technical validation of the results from microarray analysis, we performed realtime RT-PCR on a subset of 13 genes (table 5). Genes were selected from the resulting gene lists of our bioinformatic approach, based on biological plausibility and robustness as classifiers for lymphocytic myocarditis. Biological plausibility was defined according to pathway analysis, which identified those genes as being significantly involved in inflammation and remodeling.

Fold change (FC) of most genes measured by quantitative realtime RT-PCR strongly correlated with data obtained from microarray analysis, except for MSI1, where realtime RT-PCR data revealed much stronger downregulation in patients with myocarditis vs lymphocytic cardiomyopathy compared to microarray data. Genes with highest FC as per RT-PCR were CD14 (FC = +6.8), FCER1G (FC = +5), TLR1 (FC = +4.2), TLR2 (FC = +5.9), SIGLEC1 (FC = +4.3) and ADCY7 (+4.2) (table 5). However, among the 5 candidate genes from MiPP analysis, KRT78 and POU4F1 could not be confirmed with realtime RT-PCR. Since KRT78 appeared highly robust as classifier based on microarray results, we used two different primer pairs to detect either the 3’ or the 5’ end of the gene sequence. However, none of them was able to detect KRT78 in any of the samples. When we used total RNA from immortalized keratinocytes as positive control, we received a signal from each primer pair. In order to exclude possibility of cross-hybridization that may have occurred on the microarray assay, we performed batch search in the NCBI database *(http://blast.ncbi.nlm.nih.gov/Blast.cgi*) of the target sequence that was used on the Affymetrix chip. However, there was no significant sequence homology with any gene other than KRT78. Despite this minimal incoherence between microarray analysis and the more specific realtime RT-PCR, we minimized the diagnostic biomarker to a very small set of 13 genes that performed highly robust with both methods (100% sensitivity, 100% specificity,

figure 5). Finally we confirmed overrepresentation of HLA-DQ1+ patients in myocarditis (60%), while only 20% of patients with IDCM were positive for DQ1 (data not shown) by realtime RT-PCR.

When applied to a subset of myocarditis patients with higher ejection fraction, the 13 gene signature performed with a sensitivity of 75% (95CI: 36-96%), specificity of 100% (95CI: 52-100%), PPV of 100% (95CI: 52-100%) and NPV of 75% (95CI: 36-96%).

DISCUSSION

Distinction of inflammatory as compared to non-inflammatory cardiomyopathies by standard histology represents a major diagnostic challenge[9, 27, 30]. Moreover, delineating between different inflammatory cardiomyopathies with highly variable clinical courses is an even more challenging task[3, 31]. Given the emerging value of transcriptomics to add greatly to the accuracy of complex diagnoses[23, 32, 32, 33], we sought to apply this technology to the problem of diagnostic inaccuracy in myocarditis, and here we report our success with this approach.

Inflammatory disorders of the heart are notoriously difficult to diagnose due to the patchy nature of the inflammation[11]. In addition, a wide variety of underlying inflammatory conditions, with highly variable clinical outcomes, can affect the heart[2]. Here we employed the transcriptome obtained from a single EMB to develop a biomarker that enhances diagnostic accuracy for lymphocytic myocarditis. Our findings are in agreement with previous transcriptomic approaches in heart disease[23-25, 33, 34]. Specifically Ruppert et al[35] reported a set of 42 genes different between inflammatory vs noninflammatory cardiomyopathy. Their findings suggested that the transcriptome of various subtypes of cardiomyopathy differs significantly from each other and that these differences may be used as a diagnostic biomarker, as shown successfully here. Consistent with the data from Ruppert and colleagues[35], we found significant activation of the toll like receptor signaling pathway in inflammatory cardiomyopathy. In particular genes such as TLR 1, 2 and 7, as well as CD 14 were overexpressed in patients with myocarditis vs IDCM[36]. Further in agreement with their findings, we found more overexpressed than downregulated genes in inflammatory vs noninflammatory cardiomyopathy. Entirely novel in our study was the identification of the smallest set of genes required to identify inflammatory cardiomyopathy from a single endomyocardial biopsy and validation of the developed molecular signature in multiple independent sets of samples consisting of various types of cardiomyopathy as well as normal heart. We have previously used TBBs to distinguish between idiopathic and ischemic cardiomyopathy[23] and to predict long term prognosis in new onset dilated cardiomyopathy[24]. Margulies and colleagues discovered a biomarker that predicts recovery from HF[37], and Deng and co-workers developed a molecular signature which detects early cardiac transplant rejection[34] that has now entered the clinic[21]. Our discoveries reported here are clinically relevant as high diagnostic sensitivity in cardiomyopathy facilitates the appropriate use of new myocarditis specific therapies[2, 3, 12-15, 38-42]. Early and accurate diagnosis in this condition is essential so as to avoid excessive myocardial damage resulting from failure to apply therapies. New candidate therapies for myocarditis include anti-inflammatory cytokines[42], anti-viral agents, and immunoabsorption[2, 3, 12-15, 38-42]. In this regard, IFN B therapy has been safely applied in humans, leading to increased LV function and elimination of viral infection[13]. Immunoglobin administration[41] in acute myocarditis as well as application of Ca-channel blockers[42], are potential approaches with promising preliminary data that entail further evaluation. While immunosuppressive therapy in inflammatory cardiomyopathy is highly controversial[12, 14, 15, 15, 40, 43], there is growing consensus that early identification and treatment of myocarditis is crucial for positive outcome.

Our diagnostic biomarker also performed accurately in patients with secondary cardiomyopathies associated with inflammation. For example, patients with SLE, sarcoidosis, or peripartum cardiomyopathy have significant incidences of myocarditis, which has clinical importance in these conditions. The TBB had a similar degree of accuracy in this population. In patients with GCM, a very aggressive form of myocarditis, the TBB accurately detected 3 of 3 patients.

Accurate diagnosis is also critical for prognostic assessment, since clinical outcome in inflammatory cardiomyopathies correlates with disease etiology[9, 10]. Based on previous findings from others[20, 22, 34], as well as from our group[23, 24], we argue that TBBs add valuable information to a comprehensive diagnostic evaluation of new onset HF. TBBs obtained from peripheral blood or tissue samples have emerged as highly successful in neoplastic[20], cardiovascular[23, 24, 34, 44], and other disease processes[22].

In order to achieve an accurate biomarker we employed a broad range of bioinformatic approaches[20, 22-25, 29, 34, 37, 44-46]. These included SAM, PAM, MiPP, unsupervised hierarchical clustering and PCA. Using SAM, we discovered a large number of differentially expressed genes in lymphocytic myocarditis vs IDCM. Importantly and predictably, differentially expressed genes involved multiple biological networks with inflammatory components. Using these differentially expressed genes, we identified a subset that functioned as highly accurate biomarker using nearest shrunken centroids.

To find the smallest set of genes for classification, we used SVM-rbf, SVM-lin, QDA, LDA and combination of LDA, QDA and SVM-rbf in MiPP. Overall, all rules applied in MiPP consistently revealed 5 classifiers, which were further confirmed using PCA. Interestingly, two of those five “robust” predictive genes were not found to be present when quantitative realtime RT-PCR was used for validation. Finally we developed a highly parsimonious biomarker using MSI1 and LSI1 in combination with a subset of biologically relevant genes selected from the PAM-derived 62 gene TBB, and from SAM analysis and evaluated this signature using realtime RT-PCR; the 13 gene signature performed with perfect accuracy in the independent test set of our case-control study. The observation that mean FCs obtained from realtime RT-PCR were not entirely identical with the results from SAM analysis underlines the strength of molecular signature analysis for the development of biomarkers, a classification strategy that emphasizes differentially expressed gene expression patterns rather than individual genes. Since the expression level of an individual gene may vary across a population that shares the same phenotype, the overexpression or downregulation of an entire cluster of genes is more specific for a disease.

Based on these findings, we conclude that both the transcriptomic biomarker derived from PAM analysis, as well as the parsimonious molecular signature that resulted from multiple classification algorithms and testing for biological plausibility, performed highly accurately and should be a clinically valuable tool for the detection of myocarditis. While the more comprehensive biomarker of 62 genes performed with slightly higher accuracy, the 13 genes molecular signature is more practical for clinical application.

Since our original dataset, in which we developed the TBB, was matched in a case-control fashion, we further evaluated if the molecular signature is generalizable, or if it is possibly overfit to this particular study design[33, 47]. It has been shown in the past that confounding factors such as gender, age and therapy can affect gene expression[25, 33, 47-49]. When the TBB was applied in an additional validation set containing samples from patients with an average EF that was twice as high as the average EF of the original data set (65 vs 30%), the biomarker performed with almost perfect accuracy. Furthermore, the transcriptomic biomarker was broadly applicable to various cardiomyopathies, as well as normal heart and

performed highly accurate in data that was derived using a prototype microarray, confirming intraplatform reproducibility.

Both molecular signatures require testing in a clinical trial, to evaluate the diagnostic value of those biomarkers in comparison to a combination of current diagnostic tools, such as MRI, EKG, cardiac enzymes, viral screening and auto-heart antibodies. Most likely, its addition to current diagnostic standards will dramatically increase sensitivity for myocarditis. The ability to detect inflammatory components, such as involvement of the complement cascade or genes involved in cell adhesion such as ITGB2 by microarray analysis may explain why this technology is able to identify myocarditis with much greater sensitivity at an earlier stage than standard histology, a method that requires presence of inflammatory cells.

While the main goal of this study was to develop a highly accurate biomarker to distinguish lymphocytic myocarditis from IDCM, our results also provide insight into disease pathophysiology at the molecular level. Among overexpressed genes in myocarditis was CD8, involved in inflammation and binding and reported to play a fundamental role in myocarditis[30]. Interestingly, a pathway involving the TSH receptor was overexpressed in patients with myocarditis, implicating potential pathophysiologic overlap with inflammatory thyroid disease, a finding clinically established for giant cell myocarditis (Graves')[50]. There was overrepresentation of patients, positive for the HLA-DQ1B locus in myocarditis vs IDCM, suggesting possible susceptibility for lymphocytic myocarditis in this group.

Many transcripts, involving structural proteins and muscle development (late cornified envelope 1 E, collagen type I), were downregulated in myocarditis, possibly explaining structural defects and consequent dilatation in patients with this type of disease.

Study limitations

While collection of samples and clinical data over a 10 year period is a major strength of this study, a consequent limitation is the diagnosis of our patients according to the Dallas criteria[26, 27], which were standard when the study was initiated, but have been suggested to have limited sensitivity. In the meantime, several investigators suggested screening for serum anti-heart antibodies[16] and viral RNA[31] in EMBs. Notwithstanding this technical drawback, all patients received comprehensive testing in a highly specialized institution. We anticipate that in the future the transcriptomic approach coupled with determination of viral persistence and/or utilization of highly specific imaging techniques might enhance diagnostic accuracy and be used for further diagnostic refinement so as to distinguish between viral and non-viral causes of myocarditis. Ongoing work is under way to evaluate, if the presented transcriptomic biomarker will also be able to detect samples from patients with myocarditis, in whom comprehensive diagnostic testing was required to detect disease, while diagnosis of myocarditis would have been missed by Dallas criteria.

Another limitation of this study that warrants mention is that the number of samples with secondary cardiomyopathy was small, due to the known low incidence of these types of myocardial diseases. Consequently, negative and positive predictive values were estimated based on small sample size.

In short, we discovered a TBB, derived from a single EMB, which identified samples with lymphocytic myocarditis with very high accuracy. Our findings are highly relevant for a clinical application, since this novel diagnostic tool exceeds sensitivity and specificity of any technology that has been applied previously. The molecular signature was highly robust and replicated multiple times by a broad set of established classification algorithms. Validation in three independent data sets revealed high diagnostic accuracy and genes within the

transcriptomic biomarker suggest biological plausibility. Altogether, using this approach dramatically increases diagnostic accuracy of a single EMB, which may be of critical importance to the development and allocation of emerging specific therapies for inflammatory conditions of the heart.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge the efforts of Gina Edness, RN, Elayne Breton, RN, and the staff of the Johns Hopkins Hospital cardiac catheterization laboratory for support in the collection of patient samples. We thank Francisco Martinez Murillo, PhD, Linda Dorsch, B.S., and Ira Maine, PhD. from the Johns Hopkins Microarray Core Facility for consultation and their assistance with sample processing.

The authors dedicate this work to Dr. Kenneth Lee Baughman.

FUNDING SOURCES

This work was supported by NIH grant U54-HL081028 (Specialized Center for Cell Based Therapy) and R01s HL084275, AG025017, HL065455, and HL094849 to JHM and a fellowship grant of the Myocarditis Foundation.

Reference List

1. Lieberman EB, Hutchins GM, Herskowitz A, Rose NR, Baughman KL. Clinicopathologic description of myocarditis. J Am Coll Cardiol. 1991; 18:1617–26. [PubMed: 1960305]
2. Cooper LT Jr. Myocarditis. N Engl J Med. 2009; 360:1526–38. [PubMed: 19357408]
3. Kuhl U, Pauschinger M, Seeberg B, Lassner D, Noutsias M, Poller W, Schultheiss HP. Viral persistence in the myocardium is associated with progressive cardiac dysfunction. Circulation. 2005; 112:1965–70. [PubMed: 16172268]
4. Maekawa Y, Ouzounian M, Opavsky MA, Liu PP. Connecting the missing link between dilated cardiomyopathy and viral myocarditis: virus, cytoskeleton, and innate immunity. Circulation. 2007; 115:5–8. [PubMed: 17200452]
5. Caforio AL, Keeling PJ, Zachara E, Mestroni L, Camerini F, Mann JM, Bottazzo GF, McKenna WJ. Evidence from family studies for autoimmunity in dilated cardiomyopathy. Lancet. 1994; 344:773–7. [PubMed: 7916071]
6. Hannenhalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, Morrisey EE, Margulies KB, Cappola TP. Transcriptional genomics associates FOX transcription factors with human heart failure. Circulation. 2006; 114:1269–76. [PubMed: 16952980]
7. Pulerwitz TC, Cappola TP, Felker GM, Hare JM, Baughman KL, Kasper EK. Mortality in primary and secondary myocarditis. Am Heart J. 2004; 147:746–50. [PubMed: 15077094]
8. Cooper LT, Baughman KL, Feldman AM, Frustaci A, Jessup M, Kuhl U, Levine GN, Narula J, Starling RC, Towbin J, Virmani R. The role of endomyocardial biopsy in the management of cardiovascular disease: a scientific statement from the American Heart Association, the American College of Cardiology, and the European Society of Cardiology. Circulation. 2007; 116:2216–33. [PubMed: 17959655]
9. Felker GM, Thompson RE, Hare JM, Hruban RH, Clemetson DE, Howard DL, Baughman KL, Kasper EK. Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. N Engl J Med. 2000; 342:1077–84. [PubMed: 10760308]
10. McCarthy RE III, Boehmer JP, Hruban RH, Hutchins GM, Kasper EK, Hare JM, Baughman KL. Long-term outcome of fulminant myocarditis as compared with acute (nonfulminant) myocarditis. N Engl J Med. 2000; 342:690–5. [PubMed: 10706898]
11. Hrobon P, Kuntz KM, Hare JM. Should endomyocardial biopsy be performed for detection of myocarditis? A decision analytic approach. J Heart Lung Transplant. 1998; 17:479–86. [PubMed: 9628566]

12. Mason JW, O'Connell JB, Herskowitz A, Rose NR, McManus BM, Billingham ME, Moon TE. A clinical trial of immunosuppressive therapy for myocarditis. The Myocarditis Treatment Trial Investigators. N Engl J Med. 1995; 333:269–75. [PubMed: 7596370]

13. Kuhl U, Pauschinger M, Schwimmbeck PL, Seeberg B, Lober C, Noutsias M, Poller W, Schultheiss HP. Interferon-beta treatment eliminates cardiotropic viruses and improves left ventricular function in patients with myocardial persistence of viral genomes and left ventricular dysfunction. Circulation. 2003; 107:2793–8. [PubMed: 12771005]

14. Mann DL. Targeted anticytokine therapy and the failing heart. Am J Cardiol. 2005; 95:9C–16C.

15. Frustaci A, Russo MA, Chimenti C. Randomized study on the efficacy of immunosuppressive therapy in patients with virus-negative inflammatory cardiomyopathy: the TIMIC study. Eur Heart J. 2009; 30:1995–2002. [PubMed: 19556262]

16. Caforio AL, Calabrese F, Angelini A, Tona F, Vinci A, Bottaro S, Ramondo A, Carturan E, Iliceto S, Thiene G, Daliento L. A prospective study of biopsy-proven myocarditis: prognostic relevance of clinical and aetiopathogenetic features at diagnosis. Eur Heart J. 2007; 28:1326–33. [PubMed: 17493945]

17. Pauschinger M, Doerner A, Kuehl U, Schwimmbeck PL, Poller W, Kandolf R, Schultheiss HP. Enteroviral RNA replication in the myocardium of patients with left ventricular dysfunction and clinically suspected myocarditis. Circulation. 1999; 99:889–95. [PubMed: 10027811]

18. Laissy JP, Hyafil F, Feldman LJ, Juliard JM, Schouman-Claeys E, Steg PG, Faraggi M. Differentiating acute myocardial infarction from myocarditis: diagnostic value of early- and delayed-perfusion cardiac MR imaging. Radiology. 2005; 237:75–82. [PubMed: 16126925]

19. Shonk JR, Vogel-Claussen J, Halushka MK, Lima JA, Bluemke DA. Giant cell myocarditis depicted by cardiac magnetic resonance imaging. J Comput Assist Tomogr. 2005; 29:742–4. [PubMed: 16272843]

20. 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; 99:6567–72. [PubMed: 12011421]

21. Pham MX, Teuteberg JJ, Kfoury AG, Kfoury AG, Starling RC, Deng MC, Cappola TP, Kao A, Anderson AS, Cotts WG, Ewald GA, Baran DA, Bogaev RC, Elashoff B, Baron H, Yee J, Valantine HA. Gene-Expression Profiling for Rejection Surveillance after Cardiac Transplantation. N Engl J Med. 2010; 362:1890–900. [PubMed: 20413602]

22. Martinez-Llordella M, Lozano JJ, Puig-Pey I, Orlando G, Tisone G, Lerut J, Benitez C, Pons JA, Parrilla P, Ramirez P, Brugera M, Rimola A, Sanchez-Fueyo A. Using transcriptional profiling to develop a diagnostic test of operational tolerance in liver transplant recipients. J Clin Invest. 2008; 118:2845–57. [PubMed: 18654667]

23. Kittleson MM, Ye SQ, Irizarry RA, Minhas KM, Edness G, Conte JV, Parmigiani G, Miller LW, Chen Y, Hall JL, Garcia JG, Hare JM. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. Circulation. 2004; 110:3444–51. [PubMed: 15557369]

24. Heidecker B, Kasper EK, Wittstein IS, Champion HC, Breton E, Russell SD, Kittleson MM, Baughman KL, Hare JM. Transcriptomic biomarkers for individual risk assessment in new-onset heart failure. Circulation. 2008; 118:238–46. [PubMed: 18591436]

25. Heidecker B, Lamirault G, Kasper EK, Wittstein IS, Champion HC, Breton E, Russell SD, Hall J, Kittleson MM, Baughman KL, Hare JM. The gene expression profile of patients with new-onset heart failure reveals important gender-specific differences. Eur Heart J. 2009; 31:1188–1196. [PubMed: 20031959]

26. Aretz HT. Myocarditis: the Dallas criteria. Hum Pathol. 1987; 18:619–24. [PubMed: 3297992]

27. Baughman KL. Diagnosis of myocarditis: death of Dallas criteria. Circulation. 2006; 113:593–5. [PubMed: 16449736]

28. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]

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; 98:5116–21. [PubMed: 11309499]

30. Hare, JM. The dilated, restrictive and infiltrative cardiomyopathies.. In: Zipes, DP.; Libby, P.; Bonow, R.; Braunwald, E., editors. Braunwald's Heart Disease. Elsevier; 2007.

31. Kuhl U, Pauschinger M, Noutsias M, Seeberg B, Bock T, Lassner D, Poller W, Kandolf R, Schultheiss HP. High prevalence of viral genomes and multiple viral infections in the myocardium of adults with “idiopathic” left ventricular dysfunction. Circulation. 2005; 111:887–93. [PubMed: 15699250]

32. Cooper LT Jr. Onuma OK, Sagar S, Oberg AL, Mahoney DW, Asmann YW, Liu P. Genomic and proteomic analysis of myocarditis and dilated cardiomyopathy. Heart Fail Clin. 2010; 6:75–85. [PubMed: 19945063]

33. Kittleson MM, Irizarry RA, Heidecker B, Hare JM. Transcriptomics: Translation of Global Expression Analysis to Genomic Medicine. Handbook of Genomic Medicine. 2008:1–11.

34. Deng MC, Eisen HJ, Mehra MR, Billingham M, Marboe CC, Berry G, Kobashigawa J, Johnson FL, Starling RC, Murali S, Pauly DF, Baron H, Wohlgemuth JG, Woodward RN, Klingler TM, Walther D, Lal PG, Rosenberg S, Hunt S. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. Am J Transplant. 2006; 6:150–60. [PubMed: 16433769]

35. Ruppert V, Meyer T, Pankuweit S, Moeller E, Funck RC, Grimm W, Maisch B. Gene expression profiling from endomyocardial biopsy tissue allows distinction between subentities of dilated cardiomyopathy. J Thorac Cardiovasc Surg. 2008; 136:360–9. [PubMed: 18692642]

36. Fallach R, Shainberg A, Avlas O, Fainblut M, Chepurko Y, Porat E, Hochhauser E. Cardiomyocyte Toll-like receptor 4 is involved in heart dysfunction following septic shock or myocardial ischemia. J Mol Cell Cardiol. 2010; 48:1236–44. [PubMed: 20211628]

37. Margulies KB, Matiwala S, Cornejo C, Olsen H, Craven WA, Bednarik D. Mixed messages: transcription patterns in failing and recovering human myocardium. Circ Res. 2005; 96:592–9. [PubMed: 15718504]

38. Cooper LT Jr. Berry GJ, Shabetai R. Idiopathic giant-cell myocarditis--natural history and treatment. Multicenter Giant Cell Myocarditis Study Group Investigators. N Engl J Med. 1997; 336:1860–6. [PubMed: 9197214]

39. Cooper LT Jr. Shabetai R. Immunosuppressive therapy for myocarditis. N Engl J Med. 1995; 333:1713–4. [PubMed: 7477237]

40. Maisch B, Hufnagel G, Kolsch S, Funck R, Richter A, Rupp H, Herzum M, Pankuweit S. Treatment of inflammatory dilated cardiomyopathy and (peri)myocarditis with immunosuppression and i.v. immunoglobulins. Herz. 2004; 29:624–36. [PubMed: 15912438]

41. Robinson J, Hartling L, Vandermeer B, Crumley E, Klassen TP. Intravenous immunoglobulin for presumed viral myocarditis in children and adults. Cochrane Database Syst Rev. 2005; (1):CD004370. [PubMed: 15674945]

42. Yuan Z, Kishimoto C, Shioji K. Beneficial effects of low-dose benidipine in acute autoimmune myocarditis: suppressive effects on inflammatory cytokines and inducible nitric oxide synthase. Circ J. 2003; 67:545–50. [PubMed: 12808275]

43. Parrillo JE, Cunnion RE, Epstein SE, Parker MM, Suffredini AF, Brenner M, Schaer GL, Palmeri ST, Cannon RO. A prospective, randomized, controlled trial of prednisone for dilated cardiomyopathy. N Engl J Med. 1989; 321:1061–8. [PubMed: 2677721]

44. Morgun A, Shulzhenko N, Perez-Diez, Diniz RV, Sanson GF, Almeida DR, Matzinger P, Gerbase-DeLima M. Molecular profiling improves diagnoses of rejection and infection in transplanted organs. Circ Res. 2006; 98:e74–e83. [PubMed: 16794196]

45. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000; 406:747–52. [PubMed: 10963602]

46. Bueno-de-Mesquita JM, van Harten WH, Retel VP, van't Veer LJ, van Dam FS, Karsenberg K, Douma KF, van Tinteren H, Peterse JL, Wesseling J, Wu TS, Atsma D, Rutgers EJ, Brink G, Floore AN, Glas AM, Roumen RM, Bellot FE, van Krimpen C, Rodenhuis S, van de Vijver MJ, Linn SC. Use of 70-gene signature to predict prognosis of patients with node-negative breast

cancer: a prospective community-based feasibility study (RASTER). Lancet Oncol. 2007; 8:1079–87. [PubMed: 18042430]

47. Heidecker B, Hare JM. Cardiovascular Genetic Medicine: Genomic Assessment of Prognosis and Diagnosis in Patients with Cardiomyopathy and Heart Failure. Journal of Cardiovascular Translational Research. 2008; 1:225–231. [PubMed: 20559924]
48. Heidecker B, Hare JM. The use of transcriptomic biomarkers for personalized medicine. Heart Fail Rev. 2007; 12:1–11. [PubMed: 17393305]
49. Isensee J, Witt H, Pregla R, Hetzer R, Regitz-Zagrosek V, Noppinger PR. Sexually dimorphic gene expression in the heart of mice and men. J Mol Med. 2008; 86:61–74. [PubMed: 17646949]
50. Limas CJ, Iakovis P, Anyfantakis A, Kroupis C, Cokkinos DV. Familial clustering of autoimmune diseases in patients with dilated cardiomyopathy. Am J Cardiol. 2004; 93:1189–91. [PubMed: 15110223]

CLINICAL PERSPECTIVE

New diagnostic tools based on gene signatures derived from the entire complement of messenger RNAs in a cell or tissue have become established in the clinical management of certain disorders, particularly cancer. The comprehensiveness of this approach contributes to its accuracy. Myocarditis is a disorder that causes a substantial proportion of patients presenting with new-onset heart failure and left-ventricular dysfunction. Typically diagnosed by endomyocardial biopsy and evaluated with histologic criteria called the Dallas criteria, clinical management is hampered by a low sensitivity and specificity as well as the need for multiple cardiac biopsies. The present study suggests that the application of a transcriptomic based biomarker can substantially improve the diagnostic accuracy of heart biopsy for myocarditis. Using endomyocardial biopsy tissue obtained at the time of clinical presentation, we developed a molecular signature comprising 62 genes that predicted highly accurately the presence of myocarditis in a population of 48 patients. Importantly this required evaluation of tissue from a single endomyocardial biopsy sample, and therefore is clinically practical. The present results could provide treating physicians with important and accurate diagnostic information about individual patients and could provide tools for personalized treatment or monitoring. Given emerging treatment strategies for viral and inflammatory myocarditis, accurate diagnostic tools are of increased importance.

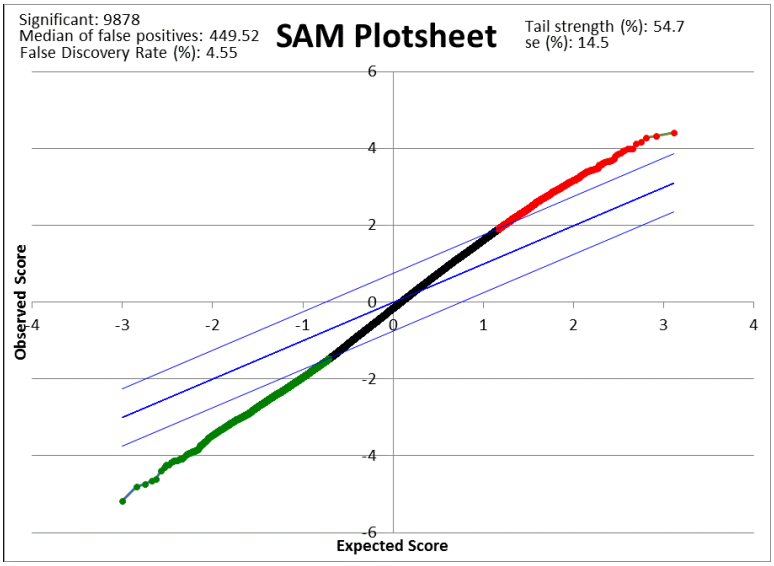

Figure 1. Significance Analysis of Microarrays Plot of differentially expressed genes in lymphcytic myocarditis vs idiopathic dilated cardiomyopathy

There were 9,878 genes differentially expressed in myocardits (n=16) vs IDCM (n=32; q<5%; fold change>1.2), of which 2,313 were overexpressed (depicted in red) and 7,565 were downregulated (depicted in green).

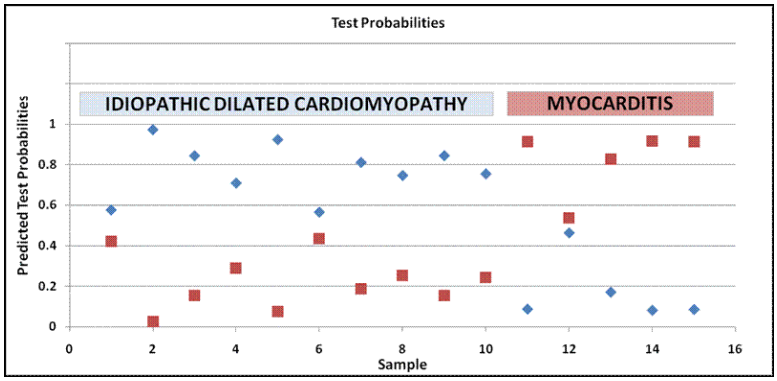

Figure 2. Validation of a 62-gene molecular signature in an independent test set (idiopathic dilated cardiomyopathy: n=10, myocarditis: n=5) using Prediction Analysis of Microarrays (PAM)

The y-ordinate illustrates the predicted test probability values obtained from PAM analysis; x-ordinate lists the number of samples. While samples were assigned to different classes with varying probability values, the classification accuracy of the transcriptomic biomarker was 100%.

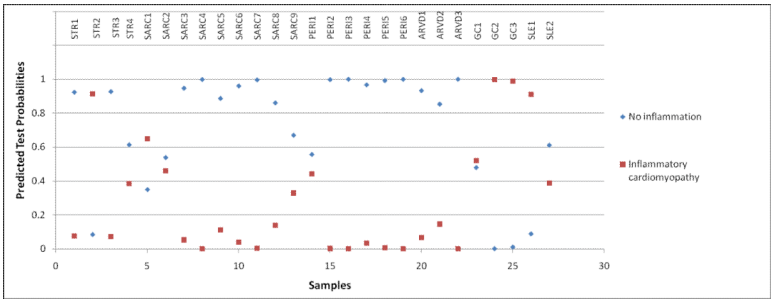

Figure 3. Prediction Analysis of Microarrays (PAM) applying the developed molecular signature for inflammatory cardiomyopathy in patients with secondary cardiomyopathy (n=27)

The transcriptomic biomarker performed with 100% sensitivity and 95% specificity in identifying inflammation in patients with stress induced cardiomyopathy (STR, n=4), sarcoidosis (SARC, n=9), peripartum cardiomyopathy (PERI, n=6), arrhythmogenic right ventricular dysplasia (ARVD, n=3), giant cell myocarditis (GC, n=3) and systemic lupus erythematosus (SLE, n=2). One patient with STR (sample #109) and another one with SLE (sample #76) were identified as inflammatory cardiomyopathy. Indeed, when results from immunohistochemistry were revised, those 2 samples contained significant lymphocytic infiltrates. One sample from the group with sarcoidosis (sample #113) was misclassified as inflammatory cardiomyopathy, while the report from histopathology revealed no signs of inflammation. All samples from patients with giant cell myocarditis were correctly identified.

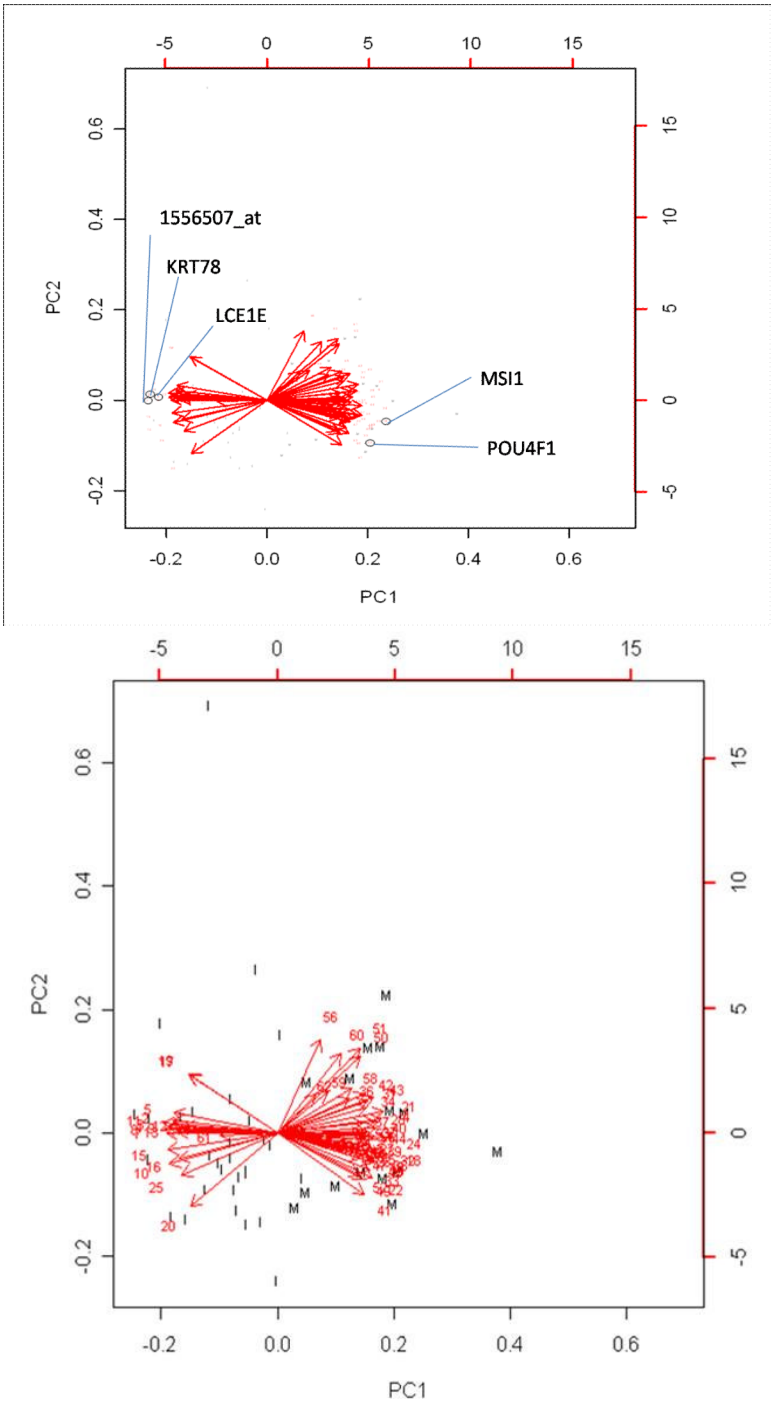

Figure 4. Principal Components Analysis (PCA) of patients with myocarditis vs idiopathic dilated cardiomyopathy (IDCM)

To illustrate significance of each of the 62 genes for phenotypic categorization, we performed PCA with correlation matrix in samples from patients with myocarditis (n=16) or IDCM (n=32) with genes as variables. Genes are labeled with serial numbers and expression levels of each individual gene are illustrated as Eigen vector towards the class, in which they are overexpressed. Vectors close to the center with close to vertical direction depict genes that were less robust, while genes that were highly specific for a phenotype were illustrated

as vectors with endpoint distant from the center directing towards the corresponding clustered set of samples of a specific phenotype.

a.) Encircled are genes that were repeatedly identified to be the most robust markers of myocarditis, when various algorithms of Misclassified-Penalized Posterior classification were applied. Output from PCA places those genes both far from the center as well as distant from the vertical line, confirming that these are highly robust classifiers for myocarditis.

b.) Clustered samples from patients with myocarditis are labeled “M”, while IDCM samples are labeled “I”. All samples from myocarditis, except two, were noticeably grouped together, suggesting that a small set of 62 genes enables clear distinction between patients with inflammatory heart disease and IDCM. Importantly, those two samples were also misclassified in our heatmap analysis, while Prediction Analysis of Microarrays identified both of them correctly.

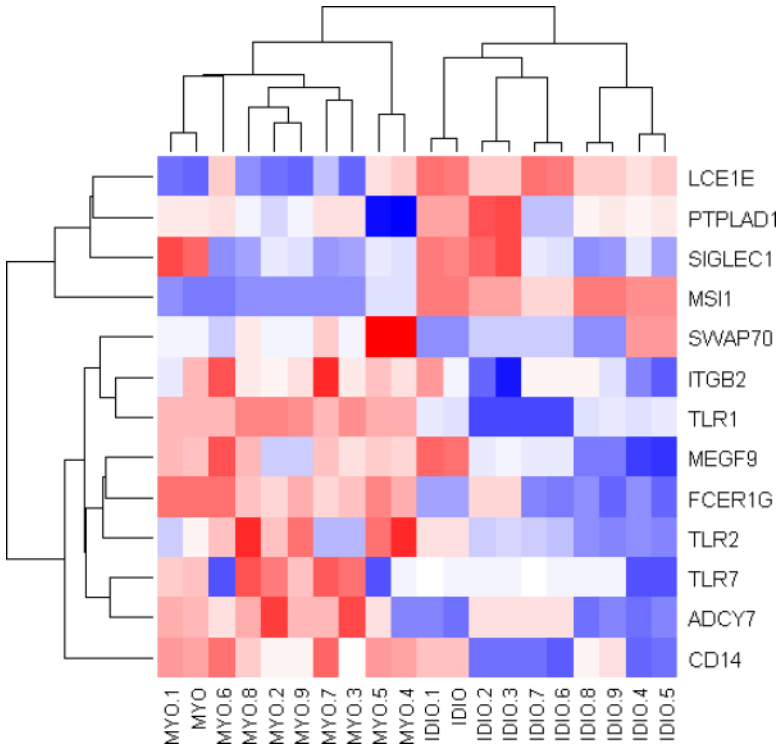

Figure 5. Distinction of patients with idiopathic dilated cardiomyopathy vs lymphocytic myocarditis based on results from quantitative realtime RT-PCR

This heatmap was created with an unsupervised clustering approach based on Euclidean distance in R, using the detected gene expression levels from quantitative realtime RT-PCR as confirmatory test. Columns represent samples and rows represent genes labeled with their corresponding gene symbol. Application of the developed 13 genes molecular signature through realtime RT-PCR correctly identified all samples.

Table 1

Baseline conditions of patients with idiopathic dilated cardiomyopathy and lymphocytic myocarditis

	Idiopathic dilated cardiomyopathy (n=32)	**Myocarditis (n=16)**
Age	48 (±3)	45 (±6)
Male, n (%)	11 (38)	11 (69)
NYHA, n (%)		
I	9 (28)	4 (25)
II	10 (31)	3 (19)
III	13 (59)	8 (50)
IV	3 (9)	1 (6)
LV EF, %	26 ± 2	33 ± 4
LVIDD, cm	5 ± 0.3	5 ± 0.2
PAP, mmHg		
Systolic	38 ± 3	37 ± 3
Diastolic	18 ± 2	15 ± 2
PCWP, mmHg	15 ± 2	12 ± 2
Systolic BP, mmHg	128 ± 5	119 ± 5
Diastolic BP, mmHg	76 ± 2	70 ± 4
Medications, n (%)		
B-Antagonist	20 (62)	9 (56)
ACE inhibitor	20 (62)	14 (88)
Aldosterone antagonist	4 (13)	1 (6)
Diuretic	14 (64)	13 (81)
Intravenous inotropic therapy	NA	NA

± refers to standard error of the mean

Table 2

Transcriptomic diagnostic biomarker for detection of patients with myocarditis: 62 genes

Probe Set ID	Gene Symbol	Gene Title	GO biological process term
1552302_at	FLJ77644, TMEM106A	similar to transmembrane protein 106A, transmembrane protein 106A	NA
1552310_at	C15orf40	chromosome 15 open reading frame 40	NA
1553212_at	KRT78	keratin 78	NA
1555349_a_at	ITGB2	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)	apoptosis, inflammatory response, leukocyte adhesion
1555878_at	RPS24	Ribosomal protein S24	translation
1556033_at	NA	NA	NA
1556507_at	NA	NA	NA
1558605_at	NA	NA	NA
1559224_at	LCE1E	late cornified envelope 1E	keratinization
1562785_at	HERC6	Hect domain and RLD 6	protein modification process
1565662_at	NA	NA	maintenance of gastrointestinal epithelium
1565830_at	NA	NA	NA
202375_at	SEC24D	SEC24 related gene family, member D (S. cerevisiae)	transport, intracellular protein transport
202445_s_at	NOTCH2	Notch homolog 2 (Drosophila)	cell fate determination
203741_s_at	ADCY7	adenylate cyclase 7	cAMP biosynthetic process, signal transduction
204222_s_at	GLIPR1	GLI pathogenesis-related 1	NA
206052_s_at	SLBP	stem-loop binding protein	mRNA processing, histone mRNA 3'-end processing
206333_at	MSI1	musashi homolog 1 (Drosophila)	nervous system development
206770_s_at	SLC35A3	solute carrier family 35 (UDP-N-acetylglucosamine (UDP-GlcNAc) transporter), member A3	UDP-N-acetylglucosamine metabolic process, transport,
209307_at	SWAP70	SWAP-70 protein	somatic cell DNA recombination, isotype switching
211089_s_at	NEK3	NIMA (never in mitosis gene a)-related kinase 3	protein amino acid phosphorylation, mitosis
211341_at	LOC100131317, POU4F1	similar to hCG1781072, POU class 4 homeobox 1	transcription, regulation of transcription, DNA-dependent, regulation of transcription from RNA polymerase II promoter
212511_at	PICALM	phosphatidylinositol binding clathrin assembly protein	protein complex assembly, endocytosis, receptor-mediated endocytosis
212830_at	MEGF9	multiple EGF-like-domains 9	NA
212999_x_at	hCG_1998957, HLA-DQB1 /2 , HLA-DRB1/2 /3 / 4 /5	major histocompatibility complex, class II, DR beta 1/2/3/4/5; similar to major histocompatibility complex, class II, DQ beta 1	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II
213501_at	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl	generation of precursor metabolites and energy, lipid metabolic process
213831_at	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II
217054_at	NA	NA	NA
217182_at	MUC5AC	mucin 5AC, oligomeric mucus/gel-forming	cell adhesion, digestion, fibril organization and biogenesis
217322_x_at	NA	NA	NA

Probe Set ID	Gene Symbol	Gene Title	GO biological process term
217777_s_at	PTPLAD1	protein tyrosine phosphatase-like A domain containing 1	I-kappaB kinase/NF-kappaB cascade
218803_at	CHFR	checkpoint with forkhead and ring finger domains	protein polyubiquitination, mitotic cell cycle, ubiquitin-dependent protein catabolic process
219425_at	SULT4A1	sulfotransferase family 4A, member 1	lipid metabolic process, steroid metabolic process
221663_x_at	HRH3	histamine receptor H3	signal transduction, G-protein coupled receptor protein signaling pathway, neurotransmitter secretion
223077_at	TMOD3	tropomodulin 3 (ubiquitous)	NA
224327_s_at	DGAT2	diacylglycerol O-acyltransferase homolog 2 (mouse)	glycerol metabolic process, lipid metabolic process, lipid biosynthetic process, triacylglycerol biosynthetic process
224996_at	NA	NA	NA
225579_at	PQLC3	PQ loop repeat containing 3	NA
226240_at	MGC21874	transcriptional adaptor 2 (ADA2 homolog, yeast)-beta	transcription, regulation of transcription, DNA-dependent
227280_s_at	CCNYL1	Cyclin Y-like 1	NA
227618_at	NA	NA	NA
227983_at	RILPL2	Rab interacting lysosomal protein-like 2	NA
228980_at	RFFL	ring finger and FYVE-like domain containing 1	intracellular protein transport, apoptosis
229191_at	TBCD	tubulin folding cofactor D	protein folding, beta-tubulin folding
230836_at	ST8SIA4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	protein modification process, protein amino acid glycosylation, nervous system development
231599_x_at	DPF1	D4, zinc and double PHD fingers family 1	transcription, regulation of transcription, DNA-dependent, induction of apoptosis
234495_at	KLK15	kallikrein-related peptidase 15	proteolysis
234986_at	NA	NA	NA
234987_at	NA	NA	NA
236232_at	STX4	Syntaxin 4	transport, neurotransmitter transport, intracellular protein transport
236404_at	NA	NA	NA
236698_at	NA	NA	NA
238327_at	LOC440836	similar to MGC52679 protein	cell growth
238445_x_at	MGAT5B	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase, isozyme B	NA
239463_at	NA	NA	NA
242383_at	NA	NA	NA
242563_at	NA	NA	NA
243819_at	NA	NA	NA
244841_at	SEC24A	SEC24 related gene family, member A (S. cerevisiae)	transport, intracellular protein transport, ER to Golgi vesicle-mediated transport
32069_at	N4BP1	NEDD4 binding protein 1	NA
44673_at	SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin	inflammatory response, cell adhesion
53720_at	C19orf66	chromosome 19 open reading frame 66	NA

Table 3

Most predictive gene signatures identified by MiPP in a dataset of patients with myocarditis (n=16) vs idiopathic dilated cardiomyopathy in training (n=32)

Validation was performed in independent test sets (n=18)

Gene signatures	Selection method	Prediction rule	Class comparison	Mean ER in training set	Mean ER in validation set
MSI1, 1556507_at	MiPP	SVM-rbf	2	0	0.167
KRT78	MiPP	SVM-lin	2	0.033	0.167
KRT78, 1556507_at	MiPP	QDA	2	0	0.167
KRT78, 1556507_at	MiPP	LDA	2	0	0.167
1556507_at	MiPP	LDA, QDA,SVM-rbf	2	0	0.167

Circulation. Author manuscript; available in PMC 2012 July 30.

Table 4

Models obtained from 50 random splits into train and test set

Genes obtained from 50 random splits were further validated in 200 independent random splits. Illustrated are the results from the top 5 gene clusters with the lowest mean error (ER). Mean sMipp is an additional parameter for performance and converges towards 1, as accuracy of the model increases.

Split	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	mean ER	mean sMiPP	5% ER	50% ER	95% ER
17	KRT78	1556507_at	NA	NA	NA	NA	0.078	0.789	0.188	0.063	0
45	KRT78	1556507_at	NA	NA	NA	NA	0.078	0.789	0.188	0.063	0
44	MSI1	POU4F1	1556507_at	NA	NA	NA	0.09	0.776	0.188	0.063	0
43	MSI1	POU4F1	1556507_at	LCE1E	NA	NA	0.091	0.789	0.188	0.063	0
41	LCE1E	POU4F1	MSI1	NA	NA	NA	0.092	0.791	0.188	0.063	0

Circulation. Author manuscript; available in PMC 2012 July 30.

Table 5

Realtime RT-PCR data of patients with lymphocytic myocarditis (n=10) vs idiopathic dilated cardiomyopathy (n=10)

Probe Set	Gene Symbol	Fold Change by SAM	Fold Change by qPCR	P<0.05 by SAM	P<0.05 by qPCR
201721_s_at	CD14	+5.9	+6.8	Y	Y
1554899_s_at	FCER1G	+5.3	+5	Y	Y
210146_x_at	TLR1	+4.5	+4.2	Y	Y
204923_at	TLR2	+3.9	+5.9	Y	Y
1555349_a_at	ITGB2	+3.1	+1.95	Y	Y
44673_at	SIGLEC1	+2.3	+4.3	Y	Y
219938_s_at	TLR7	+2.3	+2.8	Y	Y
203741_s_at	ADCY7	+2	+4.2	Y	Y
212830_at	MEGF9	+1.5	+2.3	Y	Y
217777_s_at	PTPLAD1	+1.5	+1.7	Y	Y
209307_at	SWAP70	+1.4	+2.1	Y	Y
206333_at	MSI1	-1.8	-8.4	Y	Y
1559224_at	LCE1E	-2.3	-2.6	Y	Y