Transcriptomic Biomarkers for Individual Risk Assessment in New-Onset Heart Failure

Bettina Heidecker, MD; Edward K. Kasper, MD; Ilan S. Wittstein, MD; Hunter C. Champion, MD, PhD; Elayne Breton, RN; Stuart D. Russell, MD; Michelle M. Kittleson, MD, PhD; Kenneth L. Baughman, MD; Joshua M. Hare, MD

Background—Prediction of prognosis remains a major unmet need in new-onset heart failure (HF). Although several clinical tests are in use, none accurately distinguish between patients with poor versus excellent survival. We hypothesized that a transcriptomic signature, generated from a single endomyocardial biopsy, could serve as a novel prognostic biomarker in HF.

Methods and Results—Endomyocardial biopsy samples and clinical data were collected from all patients presenting with new-onset HF from 1997 to 2006. Among a total of 350 endomyocardial biopsy samples, 180 were identified as idiopathic dilated cardiomyopathy. Patients with phenotypic extremes in survival were selected: good prognosis (event-free survival for at least 5 years; n=25) and poor prognosis (events [death, requirement for left ventricular assist device, or cardiac transplant] within the first 2 years of presentation with HF symptoms; n=18). We used human U133 Plus 2.0 microarrays (Affymetrix) and analyzed the data with significance analysis of microarrays and prediction analysis of microarrays. We identified 46 overexpressed genes in patients with good versus poor prognosis, of which 45 genes were selected by prediction analysis of microarrays for prediction of prognosis in a train set (n=29) with subsequent validation in test sets (n=14 each). The biomarker performed with 74% sensitivity (95% CI 69% to 79%) and 90% specificity (95% CI 87% to 93%) after 50 random partitions.

Conclusions—These findings suggest the potential of transcriptomic biomarkers to predict prognosis in patients with new-onset HF from a single endomyocardial biopsy sample. In addition, our findings offer potential novel therapeutic targets for HF and cardiomyopathy. (Circulation. 2008;118:238-246.)

Key Words: biopsy ■ genes ■ heart failure ■ prognosis ■ transcriptome

The clinical course of patients with newly diagnosed heart failure varies drastically, with some patients recovering and returning to completely normal levels of ejection fraction,1,2 whereas others develop severe symptoms of cardiac decompensation that require insertion of a left ventricular assist device (LVAD) or a heart transplant.3,4 Accurate risk assessment and prediction of prognosis at first presentation are crucial for appropriate allocation of therapy,5,6 monitoring,1,7 and patient management.1,7 Prediction tools based on standard criteria have had limited accuracy.8–11 In the present study, we present a transcriptome-based biomarker (TBB),5,7 which has been derived from a single endomyocardial biopsy sample (EMB) and which predicts the long-term clinical outcome of patients with idiopathic dilated cardiomyopathy (IDCM) with very high accuracy.
repository of 180 IDCM samples, biopsy samples were selected in a case-control fashion based on the phenotypic extremes in survival of the cohort. A group with good prognosis (n=25) was defined as having event-free survival for at least 5 years after initial presentation with heart failure symptoms; a group with poor prognosis (n=18) experienced an event within the first 2 years. Events included death (n=14), requirement for an LVAD (n=2), or cardiac transplantation (n=2).

### Processing of Biopsy Samples

EMBs were immediately flash-frozen in liquid nitrogen for storage in a biorepository. All steps of RNA isolation and processing were performed according to MIAME guidelines (Minimum Information About a Microarray Experiment). Tissue samples (average diameter ~2 mm) were homogenized with the MM 301 Mixer Mill (Retsch, Inc, Newtown, Pa; catalog No. 85120). Trizol reagent together with the Micro-to-Midi Total RNA Purification System (Invitrogen, Carlsbad, Calif; catalog No. 12183-018) was used for extraction of total RNA (success rate 97% of samples). Concentration and integrity of total RNA were measured with the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc, Santa Clara, Calif). All RNA samples exhibited intact 28S and 18S ribosomal RNA on denaturing agarose gel electrophoresis, and the 260/280-nm absorbance readings fell within the acceptable range of 1.8 to 2.1. An average of 568±92 ng (mean±SEM) of total RNA was isolated and preprocessed with the Ovation Biotin RNA Amplification and Labeling System (NuGEN Technologies, Inc, San Carlos, Calif; catalog No. 2300-12).13

### Microarray Hybridization

Samples were hybridized to the Human Genome U133 Plus 2.0 Array from Affymetrix (Santa Clara, Calif) without additional amplification steps. We judged the microarray experiments to be suitable when RNA isolation and microarray hybridization met all the indices of quality control as specified in the Affymetrix guidelines for assessing sample and array quality.7 Average background and noise of all chips registered within acceptable ranges, and hybridization efficiencies were similar for all samples.

### Statistical Analysis

Raw intensity values from microarray hybridization were normalized with the robust multiarray average implemented in the R package for statistical computing (available at www.R-project.org). In the next step, significance analysis of microarrays14 was used to identify phenotype-specific differences in gene expression. Significance analysis of microarrays defines significance with the q value, an adjusted probability value for multiple comparisons. For the development of a TBB, we used prediction analysis of microarrays15 to create a classifier in a training set (containing two thirds of the data, n=29), with subsequent validation in a test set (containing one third of the data, n=14).16 Prediction analysis of microarrays is software that enables one to find the minimum number of genes necessary to create a phenotype-specific “nearest shrunken centroid” for classification.15 This is done by a balanced 10-fold cross-validation in a training set, which enables one to choose a threshold that minimizes classification error.16 Overall accuracy of the discovered transcriptomic biomarker was assessed after 50 random partitions. To test for balanced baseline conditions of the cohort with good versus poor prognosis, we used Student’s t test for numerical variables and Fisher’s exact test for categorical variables.

The authors had full access to the data and take full responsibility for its integrity. All authors have read and agree to the manuscript as written.

### Results

We sought to test the hypothesis that the transcriptome, derived from a single EMB, contains sufficient information to develop a biomarker for prognosis in heart failure. Although substantial research has demonstrated the prognostic value of a variety of biomarkers and clinical prediction algorithms in heart failure,17,18 the ability to distinguish individual patients who will improve their functional status from those who will go on to circulatory collapse and require cardiac transplantation or LVAD placement remains an important clinical challenge.1,3

### Patient Characteristics

A total of 43 EMBs were analyzed with microarray technology to identify phenotype-specific differences in gene expression and to develop a prognostic TBB. Table 1 contains the baseline conditions of both cohorts, patients with good prognosis and those with poor prognosis. Importantly, there were no significant differences in age, gender, ventricular function, hemodynamics, or drug therapy between the 2 groups. The overall population with IDCIM presented an average age of 46±15 years, with slight overrepresentation of males (67%). Most patients were at an advanced New York Heart Association classification stage (at least stage II), with severely compromised ejection fraction of 23±13%, average left ventricular internal diastolic dimension of 6.1±1.5 cm, and pulmonary capillary wedge pressure of 15±9 mm Hg.

### Microarray Analysis

An average of 568±92 ng (mean±SEM) of total RNA was isolated from all EMBs and tested with the Agilent 2100 Bioanalyzer, which revealed high integrity and purity of RNA of all samples with consistent bands of 18S and 28S RNA.
Hall et al22 discovered a molecular signature of recovery that was specific for patients who recovered from heart failure during LVAD placement versus patients who continued to be listed for heart transplantation or died while waiting. Margulies et al23 concluded from a similarly designed study that transcriptomic changes in reperfusion tissue from explanted hearts or LVAD would not consistently explain their observed clinical outcomes, suggesting a bystander effect of volume unloading.23,25 Although some patients undergo complete recovery of their heart function within an average of 5 years, others progress into circulatory collapse within the first 2 years of presentation and require aggressive therapeutic interventions, such as mechanical circulatory assistance or cardiac transplantation.3 The ability to predict these patients has great clinical value in an era of multiple therapeutic options.

Previous transcriptomic studies16,22–24 investigated heart tissue from explanted hearts or LVAD. The results of these studies continue to be controversial and have led to the questioning of the feasibility of TBBs in cardiovascular disorders.23 Many previous studies evaluated gene expression changes during left ventricular functional recovery with mechanical circulatory assistance to discover genes associated with recovery. This approach is inherently limited, because gene expression changes could be either causative or due to a bystander effect of volume unloading.23,25 Although Hall et al23 discovered a molecular signature of recovery that was specific for patients who recovered from heart failure during LVAD placement versus patients who continued to be dependent on a device, Margulies et al23 concluded from a similarly designed study that transcriptomic changes in recovering LVAD patients would not consistently explain their functional improvement. These controversial findings moti-
<table>
<thead>
<tr>
<th>Affymetrix ID</th>
<th>Gene Title</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>232669_at</td>
<td>Hypoxia-inducible factor 3, α-subunit</td>
<td>1.8</td>
</tr>
<tr>
<td>214951_at</td>
<td>Solute carrier family 26, member 10</td>
<td>1.8</td>
</tr>
<tr>
<td>243482_at</td>
<td>Epidermal growth factor receptor pathway substrate 15-like 1</td>
<td>1.8</td>
</tr>
<tr>
<td>226210_s_at</td>
<td>Maternally expressed 3</td>
<td>1.7</td>
</tr>
<tr>
<td>232159_at</td>
<td>Epidermal growth factor receptor pathway substrate 15-like 1</td>
<td>1.7</td>
</tr>
<tr>
<td>233026_s_at</td>
<td>PDZ domain containing 2</td>
<td>1.6</td>
</tr>
<tr>
<td>211996_s_at</td>
<td>KIAA0220-like protein hypothetical gene LOC 283846</td>
<td>1.6</td>
</tr>
<tr>
<td>243774_at</td>
<td>Mucin 20, cell surface associated</td>
<td>1.6</td>
</tr>
<tr>
<td>242551_at</td>
<td>Chromosome 18 open reading frame 1</td>
<td>1.6</td>
</tr>
<tr>
<td>244548_at</td>
<td>Rho GTPase activating protein 28</td>
<td>1.6</td>
</tr>
<tr>
<td>244208_at</td>
<td>Checkpoint suppressor 1</td>
<td>1.6</td>
</tr>
<tr>
<td>239984_at</td>
<td>Sodium channel, voltage-gated, type VII, alpha</td>
<td>1.6</td>
</tr>
<tr>
<td>230683_at</td>
<td>CDNA:FLJ26892 fis, clone ADKA03430</td>
<td>1.5</td>
</tr>
<tr>
<td>241869_at</td>
<td>Apolipoprotein L, 6</td>
<td>1.5</td>
</tr>
<tr>
<td>241597_at</td>
<td>Arginine-glutamic acid dipeptide (RE) repeats</td>
<td>1.5</td>
</tr>
<tr>
<td>239887_at</td>
<td>Smg-6 homolog, nonsense mediated mRNA decay factor (C. elegans)</td>
<td>1.5</td>
</tr>
<tr>
<td>229957_at</td>
<td>Transmembrane protein 91</td>
<td>1.5</td>
</tr>
<tr>
<td>223546_x_at</td>
<td>LUC7L-like (S. cerevisiae)</td>
<td>1.5</td>
</tr>
<tr>
<td>239567_at</td>
<td>Rho GTPase activating protein 10</td>
<td>1.5</td>
</tr>
<tr>
<td>242194_at</td>
<td>Cullin 4A</td>
<td>1.5</td>
</tr>
<tr>
<td>1558525_at</td>
<td>Hypothetical protein LOC283901</td>
<td>1.4</td>
</tr>
<tr>
<td>227178_at</td>
<td>CUG triplet repeat, RNA binding protein 2</td>
<td>1.4</td>
</tr>
<tr>
<td>228198_s_at</td>
<td>Mitochondrial ribosomal protein S9</td>
<td>1.4</td>
</tr>
<tr>
<td>202379_s_at</td>
<td>Natural killer-tumor recognition sequence</td>
<td>1.4</td>
</tr>
<tr>
<td>224260_at</td>
<td>CDNA clone</td>
<td>1.4</td>
</tr>
<tr>
<td>238643_at</td>
<td>Neuroblastoma, suppression of tumorigenicity 1</td>
<td>1.4</td>
</tr>
<tr>
<td>232253_at</td>
<td>RAD50 homolog (S. cerevisiae)</td>
<td>1.4</td>
</tr>
<tr>
<td>227968_at</td>
<td>Parkinson disease 7 domain containing 1</td>
<td>1.4</td>
</tr>
<tr>
<td>233197_at</td>
<td>Kelch-like 9 (Drosophila)</td>
<td>1.4</td>
</tr>
<tr>
<td>244512_at</td>
<td>Transcribed locus, strongly similar to XP_001081342.1</td>
<td>1.4</td>
</tr>
<tr>
<td>233443_at</td>
<td>Hypothetical protein LOC389362</td>
<td>1.4</td>
</tr>
<tr>
<td>231275_at</td>
<td>FLJ42875 protein</td>
<td>1.4</td>
</tr>
<tr>
<td>226419_s_at</td>
<td>Hypothetical protein LOC45460</td>
<td>1.4</td>
</tr>
<tr>
<td>201221_s_at</td>
<td>Small nuclear ribonucleoprotein 70-kDa polypeptide</td>
<td>1.4</td>
</tr>
<tr>
<td>209354_at</td>
<td>Tumor necrosis factor receptor family member 14</td>
<td>1.4</td>
</tr>
<tr>
<td>226571_s_at</td>
<td>Protein tyrosine phosphatase receptor type, S</td>
<td>1.4</td>
</tr>
<tr>
<td>220728_at</td>
<td>EST</td>
<td>1.3</td>
</tr>
<tr>
<td>203071_at</td>
<td>Sema domain, immunoglobulin domain (Ig), short basic domain</td>
<td>1.3</td>
</tr>
<tr>
<td>213946_s_at</td>
<td>Obscurin-like 1, similar to titin isoform N2-B</td>
<td>1.3</td>
</tr>
<tr>
<td>201394_s_at</td>
<td>RNA binding motif protein 5</td>
<td>1.3</td>
</tr>
<tr>
<td>203748_x_at</td>
<td>RNA binding motif, single-stranded interacting protein 1</td>
<td>1.3</td>
</tr>
<tr>
<td>223147_s_at</td>
<td>WD repeat domain 33</td>
<td>1.3</td>
</tr>
<tr>
<td>213773_x_at</td>
<td>NOL1/NOP2/Sun domain family, member 5</td>
<td>1.3</td>
</tr>
<tr>
<td>1560049_at</td>
<td>CUG triplet repeat, RNA binding protein 2</td>
<td>1.3</td>
</tr>
<tr>
<td>243974_at</td>
<td>CDNA clone IMAGE:4821815</td>
<td>1.2</td>
</tr>
<tr>
<td>201510_at</td>
<td>E74-like factor 3(ets domain transcription factor, epithelial specific)</td>
<td>1.2</td>
</tr>
</tbody>
</table>

*All genes listed, except 1 EST (asterisk), were used as prognostic transcriptomic biomarker for patients with HF.
vated the present study with biopsy samples obtained at the
time of patient presentation.

Other investigators have attempted TBB studies from
explanted hearts rather than LVAD patients. Steenman et al24
investigated samples with ischemic cardiomyopathy versus
IDCM and failed to find significant differences in gene
expression. However, that study was limited by an extremely
small sample size (n=110052 in each group), and data were
evaluated by pooling the RNA from these samples. This
negative finding highlights the importance of sample size in
microarray experiments. Learning-curve constructions pro-
vided evidence that 20 to 30 samples are required per training
set to achieve significance in a classification problem.26 In a
previous study that involved a larger number of samples
(n=48), we discovered a cluster of genes that accurately
distinguished idiopathic from ischemic cardiomyopathy.16

Several features of this approach may have contributed to
the present results. First, the present study used biopsy
samples obtained at the time of patient presentation, not taken
from explanted hearts, a time at which disease could be
considered end-stage. Arguably this approach reduces tran-
scriptomic changes that could be considered compensatory
and that could have been activated during disease progres-
sion.23,25 A second key issue results from our technical
approach: We used a highly efficient RNA processing tech-
nique that allowed microarray analysis without additional
amplification and thus avoided possible bias due to different
binding preferences of primers and reduction of time and
costs with regard to a later clinical application.27

The transcriptomic biomarker reported here for cardiomy-
opathy has a similar degree of accuracy as those developed
for cancer.15,20,21 Previous proof-of-concept studies suggested
the feasibility of obtaining clinically useful TBBs from
cardiac biopsy samples.16 In this regard, we previously
demonstrated in explanted heart tissue obtained at the time
of heart transplantation that a transcriptomic signature could
be developed to delineate between ischemic and nonischemic
cardiomyopathy.16 Additional support for the concept of
TBBs for heart muscle disease arises from the successful
development of expression profiles to detect cardiac
inflammation.6,28

The developed TBB, which contains a molecular signature
of 45 genes, predicted the clinical phenotype of samples with
very high specificity (90%), which suggests its utility, par-
ticularly as a rule-out test in addition to current standard risk
assessment. Furthermore, results of the TBB might be in-
cluded in the assessment of a priority score in transplant
waiting lists. In combination with clinical data and laboratory
values, the TBB has the potential to offer guidance in patient
management.6 Furthermore, the present data offer insights
into novel molecular pathways involved in the recovery of
patients with heart failure.29 Of the 46 differentially expressed
genes (Table 2), all were overexpressed in patients who
recovered from heart failure. Many overexpressed genes in
these patients have RNA or DNA binding functions, eg,
RBMS1 and WDR33, playing a crucial role in DNA replica-
tion, gene transcription, cell cycle progression, and apopto-
sis.30–32 Various genes with critical regulatory functions were
identified, ie, SNRP70; the nuclear transcription factors
ELF3, CHES1, and RERE; NSUN4, a gene with methyltrans-
transcription factor HIF3A. HIF3A has been discussed as an inductor of glucose transporters, vascular endothelial growth factor, and erythropoietin, similar to HIF1A and HIF2A, whereas others postulated its counteraction with the other 2 subunits.33,34 The genes CUGBP2, LUC7-like, and SEMA3B are involved in neuromuscular development, axon guidance, and regulation of heart contractility (http://www.geneontology.org/). Closely related in its function is the overexpressed obscurin-like 1 gene (OBSL1), a linker that stabilizes cell contacts and organelles within the cytoskeleton35 and that is located at the intercalated discs in the adult cardiomyocyte.35 OBSL1 is supposed to have similar functions to obscurin, a multifunctional protein responsible for assembly of myofibrils and myocyte cellular organization.36 By its interaction with titin and ankyrin,36 as well as by linking sarcomere and sarcoplasmic reticulum,36 it provides the required alignments for contraction.36

Furthermore, we discovered 2 genes previously implicated in regenerative pathways, Rad50 and SMG6, important key regulators in telomerase activity and DNA repair.37 Rad50 is part of the Mre11/Rad50/Nbs1 (MRN) complex,37 a functional unit that generates t loops by inserting 3' G overhangs at human telomere ends.37 These t loops prevent chromosome ends from being recognized as damaged DNA and provide a template for telomerase and preservation of genome stability.38 The increase in telomerase activity might explain a protective effect against degenerative processes and aging in the hearts of patients with good prognosis. It is attractive to speculate that the reason for better recovery might be increased cellular regenerative capacity.39,40 Given observations that Rad50 is essential for viability of stem cells,38 Of note, 3 genes (SNRP70, OBSL1, and RBMS1) reported here were also reported in a study of 199 human LVAD samples that suggested genes involved in recovery.23 There was no overlap with results published by Lowes et al,41,42 who investigated genes that were overexpressed during recovery from heart failure. This may suggest that there are important differences between gene expression changes caused by therapy and characteristics of gene expression that may predict clinical outcome.

Validation of microarray experiments is an issue of considerable importance. In this context, it remains highly controversial whether independent RNA quantification is essential. Given the technical approach and motivation to avoid RNA amplification, we lacked sufficient RNA for independent reverse-transcription polymerase chain reaction quantification. Recent technical advances, however, obviate...
the need for polymerase chain reaction–based validation in clinical prediction studies. Indeed, the HG-U133 Plus 2.0 Gene Chip has been shown to perform with high consistency compared with established quantitative gene expression platforms. Moreover, numerous published studies have not used independent RNA quantification for validation. In this regard, recent TBB studies in oncology have used split-sample validation, which was the approach used in the present study. Additional validation of selected transcripts by reverse-transcription polymerase chain reaction was not considered mandatory, because the individual expression value of a single gene has less predictive power than the averaged overall gene expression pattern of a molecular signature.

Although endomyocardial biopsy is a low-risk procedure, alternative methods for obtaining transcriptomic biomarkers might be developed. There is evidence that affected tissue and circulating blood cells share a high percentage of common genes. Easy accessibility of peripheral blood mononuclear cells by a simple venous puncture makes those cells very attractive for clinical application. Several studies have proven a remarkably high correlation between the gene expression profiles of peripheral blood mononuclear cells and various tissues, and first diagnostic kits, based on blood samples, are commercially available. Future studies will show whether the biopsy-derived biomarkers are also present in the corresponding peripheral blood mononuclear cells of patients, whether peripheral blood mononuclear cell–derived markers perform even better, or whether a combination of both can potentiate the overall accuracy. Future studies will be also necessary to compare the utility of the transcriptomic biomarker in combination with existing biomarkers (eg, brain natriuretic peptide, uric acid) and established prediction algorithms (artificial neural networks, the Seattle Heart Failure Model). As a result of frequent comorbidity in patients with cardiomyopathy, a comprehensive profile of laboratory values and clinical data may be necessary for accurate prediction of an individual’s prognostic outcome.

Although the present study is one of the most extensive transcriptomic studies of heart tissue, the fact that the patient sample size was much smaller than the measured markers (>40 000 transcripts per sample) raises important statistical challenges. Power calculations are particularly challenging in this scenario, largely because the variance of each gene contained within the biomarker is impossible to assess, but learning-curve modulations have led to a general recommendation that samples of 20 to 30 in the training set are adequate. Nevertheless, logistic regression models indicate that under these circumstances, the probability of developing a nonrepresentative prediction algorithm rises. Sophisticated statistical modeling, however, indicates that several issues and approaches mitigate this problem. First, if the number of genes included in the predictive biomarker as predictors can be reduced to <50 and can be chosen according to evidence of differences in gene expression, methods of classification may be applied successfully. Critical to this approach is the practice of validating the biomarker obtained in the training set in an independent test set. The ultimate validation of a successful TBB requires prospective study in an independent clinical trial, an approach highly successful in various cancers.

In addition to the sample size issue, another study limitation is that some patients were lost to follow-up, and prognostic information was retrieved solely from the Social Security Death Index. Consequently, we were not able to assess echocardiographic data in all patients. Prospective studies are under way to address this issue.

In summary, we have developed a novel transcriptomic biomarker for prognosis in heart failure that is obtainable from a single EMB with potential for a direct clinical application. This approach should improve individualization of cardiac care and help identify patients at highest risk for circulatory collapse within the first years of presentation with heart failure.

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 assistance and support in the collection of patient samples. We thank Francisco Martinez Murillo, PhD, Linda Dorsch, BS, and Ira Maine, PhD, from the Johns Hopkins Microarray Core Facility for consultation and their assistance with sample processing.

Sources of Funding

This work was supported by National Institutes of Health grants HL-084275, AG-02017, and HL-65455 (Dr Hare).

Disclosures

None.

References


CLINICAL PERSPECTIVE

New technologies that measure expression levels of the entire complement of messenger RNAs in a cell or tissue have become highly useful for clinical prediction of disease origin, prognosis, and therapeutic response. Because they are highly comprehensive, they have the potential to be highly accurate. The present study shows that this approach could be very important to fulfill an unmet need in the field of heart failure: accurate prediction of the long-term clinical course of a patient. New-onset heart failure is very common and has a highly variable outcome; thus, the ability to accurately assess individual patient risk is of major significance. Using endomyocardial biopsy tissue obtained at the time of clinical presentation, we developed a molecular signature comprising 45 genes that predicted long-term clinical outcome in patients with new-onset heart failure. This transcriptomic biomarker distinguished patients who survived at least 5 years after first diagnosis from those who did poorly and required left ventricular mechanical assistance or cardiac transplantation or who died. These findings may provide the physician with important prognostic information about individual patients and could provide tools for personalized treatment or monitoring. Importantly, the biomarker can be obtained from a single endomyocardial biopsy and therefore is clinically practical. The biomarker contained biologically relevant genes, including those involved in regeneration and angiopoiesis, which suggests possible novel therapeutic targets.