Blood gene expression signatures associate with heart failure outcomes

Peter VanBuren, Jun Ma, Samuel Chao, Enkhtuyaa Mueller, David J. Schneider, and Choong-Chin Liew

1Department of Medicine, Cardiology Unit, University of Vermont, Burlington, Vermont; 2GeneNews Limited, Richmond Hill, Ontario, Canada; 3Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts

Submitted 7 September 2010; accepted in final form 20 January 2011

VanBuren P, Ma J, Chao S, Mueller E, Schneider DJ, Liew CC. Blood gene expression signatures associate with heart failure outcomes. Physiol Genomics 43: 392–397, 2011. First published January 25, 2011; doi:10.1152/physiolgenomics.00175.2010.—Gene expression signatures in blood correlate with specific diseases. Such signatures may serve as valuable diagnostic and prognostic tools in disease management. Blood gene expression signatures associated with heart failure may be applied to predict prognosis, monitor disease progression, and optimize treatment. Blood gene expression profiles were generated for 71 subjects with heart failure and 15 controls without heart failure, using the Affymetrix GeneChip U133Plus2.0. Survival analysis identified 197 “mortality genes” that were significantly associated with patient outcome. Functional categorization showed that genes associated with T cell receptor signaling were most significantly overpresented. Cluster analysis of these T cell receptor signaling genes significantly categorized heart failure patients into three risk groups (P = 0.031) that were distinct from the three risk groups categorized by New York Heart Association (NYHA) Classification (P = 0.0002). By combining the analysis of clinical assessment (NYHA class) with T cell receptor signaling gene expression, we proposed a model that demonstrated an even greater differentiation of patients at risk (P = 0.0001). In this discovery study, we identified blood expression signatures associated with heart failure patient outcomes. Characterization of these mortality genes helped identify a set of T cell receptor signaling genes that may be of utility in predicting survival of heart failure patients. These data raise the possibility of prospectively risk stratifying patients with heart failure by integrating blood gene expression signatures with current clinical assessment.

gene expression profiling; microarray; survival analysis

Heart failure has long been known to affect circulating levels of numerous neurohormones, cytokines, and inflammatory markers (3, 12). These factors directly and adversely affect intracellular signaling and ultimately gene expression in the myocardium. Genome-wide expression profiling has provided great insights into the pathophysiology of heart failure (13). Specifically, gene expression profiling studies in patients with heart failure have demonstrated a significant elevation in the expression of genes involved in cell growth, signal transduction, and cell defense in myocardial tissue (5, 7). Thus myocardial gene expression likely reflects direct tissue changes associated with the cardiomyopathic process as well as consequent alterations in gene expression secondary to the humoral response of the disease state.

Given the systemic nature of heart failure and the protean nature of neurohormonal signaling, other tissues are also affected by the heart failure state. In this study, we have taken advantage of this fact to discover and characterize differential expression of blood genes in patients with heart failure. Gene expression profiling of blood cells has shown promise in the identification of differentially expressed genes in coronary artery disease (9) and expression patterns correlated with blood lipid levels (8). In this study, we identified and characterized expression signatures of peripheral blood cells associated with outcomes in patients with heart failure. We hypothesize that blood gene expression profiling can contribute toward the determination of heart failure patient prognosis and thus could be used to optimize patient management.

METHODS

Study Population

Heart failure patients. Potential participants with heart failure were identified in our ambulatory clinics or at the time of hospital admission with a primary diagnosis of heart failure. Adult patients (age 21–89 yr) with history of heart failure were eligible for enrolment. All patients had left ventricular functional assessment as part of routine cardiac care prior to enrolment. Heart failure severity was clinically characterized in accordance with the New York Heart Association (NYHA) classification. Patients with NYHA class I–III heart failure were identified in our outpatient clinic at the time of routine follow up. NYHA class IV patients were identified at the time of hospital admission for decompensated heart failure. Exclusion criteria were acute myocardial infarction, unstable angina, dialysis-dependent renal failure, disseminated cancer, and sepsis. The goal of recruitment was approximately equal representation in each heart failure class.

Control group. Control subjects were selected from patients referred for evaluation of atypical or noncardiac chest pain and who had no prior diagnosis of cardiac disease. Normal ventricular function and absence of significant coronary disease in the control patient group were confirmed by stress echocardiogram or nuclear perfusion imaging.

This study was approved by the Committees in Human Research at the University of Vermont. All participants provided written, informed consent.

Blood Collection, RNA Extraction, and Gene Expression Profiling

Approximately 20 ml blood was taken from an antecubital vein after overnight fasting. Blood samples were collected with a Vacutainer tube (Becton Dickinson, Franklin Lakes, NJ) and stored on ice.
until RNA extraction (performed within 6 h after blood collection). Red blood cells were ruptured with hypotonic hemolysis buffer (1.6 mM EDTA, 10 mM KHCO₃, 153 mM NH₄Cl, pH 7.4), and white blood cells were separated by centrifugation. Total RNA of white blood cells was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA) in accordance with the manufacturer’s instructions. The quality of RNA samples was assessed on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) using RNA 6000 Nano Chips (Agilent Technologies). The quantity of RNA was determined by UV spectrophotometry (Beckman-Coulter DU640, Beckman-Coulter, Fullerton, CA). Five micrograms of total RNA from each sample was used for hybridization on a GeneChip U133Plus2 microarray (Affymetrix, Santa Clara, CA) in accordance with the manufacturer’s instructions.

Analysis of Blood Gene Expression Profiles

After scanning, probe-level expression data (CEL files) generated from the raw microarray hybridization images were processed by GC-robust multichip average (RMA) (20) using GeneSpring v7.3 software (Agilent Technologies). Genes with unreliable measurements, assessed by cross-gene error model (http://www.chem.agilent.com/cag/bsp/products/gsx/downloads/pdf/error_model.pdf), were removed from further analysis. Survival analysis was conducted using MedCalc (MedCalc Software, http://www.medcalc.be). The expression data for each gene and patient survival data were correlated, using a Cox proportional hazards regression model. Significance of associations between gene expression and survival time was assessed using a log-rank test. Multitest correction was performed using the q value package in Comprehensive R Archive Network (18). A q value using a log-rank test. Multitest correction was performed using the q value package in Comprehensive R Archive Network (18). A q value of 0.2 was chosen as the significance cutoff for the selection of mortality genes. Differentially expressed mortality genes between controls and NYHA I–II and between controls and NYHA III–IV were identified by a Welch t-test. Significance was identified when P was <0.05.

Functional categorization of the mortality genes was conducted using Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood City, CA). Genes in significantly overpresented functional group(s) were subjected to cluster analysis using Pearson correlation and average group linkage by Hierarchical Clustering Explorer (Human Computer Interaction Lab, University of Maryland). The 86 samples were then reclassified into three groups based on the cluster analysis of the mortality genes. Survival analysis was applied to the reclassified groups by cluster analysis. In addition survival analysis was also applied to the NYHA classified groups (control, NYHA I–IV). A Kaplan-Meier plot was drawn using MedCalc. Significance was assessed using a log-rank test.

A risk stratification model was proposed that integrated the clinical severity parameter (NYHA) and the gene expression signatures associated with patient outcomes. In this model, a clinical score was assigned according to each subject based on NYHA classification as follows: 0 (control), 1–4 (NYHA = I–IV); an expression score was assigned according to the cluster analysis of the 10 T cell receptor signaling genes (Fig. 2) as either 1 (group 1), 2 (group 2), or 3 (group 3); an integrated risk index (IRI) was created as defined as the sum of the clinical score and the expression score of each subject. IRI ranged from 1 to 7, and subjects were categorized into three risk groups by IRI as low risk (IRI = 1–3), intermediate risk (IRI = 4–5), and high risk (IRI = 6–7) (see Table 3). A Kaplan-Meier plot was applied to the three risk groups using the Survival package, and the significance was assessed by a log-rank test.

RESULTS

Subject Recruitment

A total of 87 subjects participated in this study: 72 heart failure patients and 15 nonheart failure controls. One heart failure patient was excluded after dying from a nonheart failure-related operation. All were Caucasian with the exception of one patient who was of Middle Eastern decent. Heart failure subjects were clinically categorized into the four NYHA classes. The clinical characteristics and medications for all subjects are summarized in Table 1. The underlying etiology of heart failure was ischemic heart disease in 39 patients and nonischemic in 32 patients. Valvular heart disease was a contributing factor in four patients (two with mitral regurgitation and two with aortic stenosis). The incidence of ischemic cardiomyopathy, diabetes, and atrial fibrillation was all higher in the NYHA III–IV group, reflecting either the effect of these confounding factors in advanced heart failure or the effect of heart failure on these comorbidities. Not surprisingly, loop diuretic use was also higher in more advanced heart failure. In the follow-up period (median 26 mo) 14 heart failure patients died. Thirteen of the 14 patients died of cardiovascular causes. In the 14th patient, heart failure was the secondary cause of death. In eight of the 14 deaths the cause of heart failure was ischemic heart disease.

Analysis of Blood Gene Expression Profiles

Affymetrix GeneChip U133Plus2.0 is a whole genome microarray containing more than 56,000 probe sets. A cross-gene
error model was applied to the 87 blood expression profiles processed with GC-RMA. After we removed probe sets with unreliable measurements, ~27,000 probe sets remained for further analysis.

Of these probe sets, survival analysis and subsequent multivariate correction identified a total of 197 genes significantly (q < 0.2) associated with patient survival, and these were defined as “mortality genes.” Table 2 shows a representative sample of the mortality genes, listing those associated with known signaling pathways (a complete list of the mortality genes is part of an online data supplement). Ten mortality genes were found to be differentially expressed (P < 0.05) in the compensated heart failure group, and 113 mortality genes were differentially expressed in the decompensated heart failure group. Of these 123 genes that were differentially expressed in the heart failure groups, the majority, 89 genes, were downregulated in heart failure patients. Interestingly, 83 of the mortality genes while associated with patient survival were not differentially expressed in heart failure groups as defined by NYHA class.

Functional categorization of the mortality genes showed that 11 canonical pathways were significantly overrepresented (P < 0.05, Fisher’s exact test, Benjamini-Hochberg multiple testing correction). The most significantly overrepresented canonical pathway involved T cell receptor signaling (Fig. 1). Ten mortality genes were identified in this functional group. Cluster analysis of these 10 genes divided the 86 subjects into three groups (Fig. 2). Survival analysis demonstrated that the survival of study subjects was associated with disease severity as described by NYHA classification as well as with the three groups defined by the 10 T cell receptor signaling-related pathways.
Row-by-Row normalization by
Standardization (Mean and Stdev)
Average Group Linkage
Pearson's r: Centered, Unabsolute
86 Items
10 Variables

Fig. 2. Cluster analysis of the 10 mortality genes involved in T cell receptor signaling segregated the patients into 3 groups.

mortality genes (Fig. 3B). In contrast, in this patient cohort left ventricular ejection fraction was not predictive of survival (data not shown). There was no sex difference in survival ($P = 0.2$). Survival analysis for NYHA-defined groups ($P = 0.0002$) was enhanced by mortality gene clustering ($P = 0.031$) such that the risk stratification model using IRI provided a marked improvement in differentiating risk groups with the $P$ value of a log-rank test being 0.0001, an increase from using either NYHA or the cluster analysis of 10 T cell receptor signaling-related genes (Fig. 4, Table 3). This approach provided good segregation of heart failure patients with 49% survival in the high risk group, 94% in the intermediate group, and 100% in the low risk group. Based on their gene expression clustering and confirmed by their survival to the end of the study period, nine heart failure patients were allocated to the low risk group.

**DISCUSSION**

Neurohormonal activation in heart failure is well established and involves activation of the renin-angiotensin-aldosterone network and the sympathetic nervous system. Associated with this process is the enhanced production of cytokines and other inflammatory factors that have been directly implicated in the pathophysiological progression of heart failure (2, 6, 11, 15). These factors have both direct and indirect effects on cardiac remodeling and can also affect gene expression in other tissues (19). Due to the systemic nature of the disease, we tested the hypothesis that noncardiac tissue can exhibit specific gene expression signatures that associate with cardiovascular events.

By performing blood gene expression profiling analysis, we analyzed ≥27,000 transcripts in blood and demonstrated for the first time that blood gene expression varies as a function of patient outcomes (see online data supplement). A total of 197 mortality genes were discovered. Intriguingly, a large number (eighty-three) of these genes were not differentially expressed between NYHA groups. As heart failure outcomes are determined by multiple factors, some of these factors are likely independent of NYHA class as is made evident by the correlation of expression the above 83 genes with patient survival in this study. Of the mortality genes that are differential expressed across NYHA groups, the differences in gene expression lay almost exclusively between the control group and the decompensated group (NYHA class III–IV). Aronow et al. (1) showed little concordance of altered gene expression in four separate cardiac hypertrophy models in the compensated state, whereas a greater overlap of differentially expressed genes was observed between groups exhibiting more severe forms of cardiomyopathy. Similarly, blood gene expression among patients with heart failure likely exhibits a convergence of transcriptional regulation as the disease progresses.

We identified and characterized an expression signature of 10 genes involved in T cell receptor signaling. Perhaps most striking is that by integrating NYHA and the gene expression signature associated with patient outcomes, risk stratification can be further improved ($P = 0.0001$). The commonly used clinical parameter of left ventricular ejection fraction was, by contrast, not predictive of survival in this patient cohort. These data suggest that blood gene expression profiling may be of benefit in assessing patient prognosis. Importantly, this analytical approach is capable of segregating out the patients that are at the highest risk for cardiac mortality and thus may be of utility in guiding the medical management of patients with heart failure in the future.

Blood leukocytes were the primary source of RNA isolated in this study. Therefore, the differential expression of leukocyte genes in heart failure may or may not directly reflect the primary disease process. For example, our transcriptomic data may reflect relative lower lymphocyte concentrations, as have been found to predict survival in advanced heart failure (14).
Irrespective, our data demonstrate that the expression profile can be used as a signature of heart failure severity. Previous studies investigating transcriptional gene expression of myocardial tissue in patients with heart failure have similarly demonstrated an overrepresentation of genes involved in inflammatory and cytokine signaling. In these studies, the gene expression in many of the inflammatory signaling pathways is enhanced, whereas in our study transcriptional expression of some of the genes involved cytokine signaling is reduced. Suppressed expression of genes in blood leukocytes may be a consequence of the overexpression and circulatory release of inflammatory factors by other tissues. Moreover, the high representation of inflammatory genes correlating with mortality shown in these data supports the concept that inflammation and cytokine signaling play an important role in the pathophysiology of heart failure (4). Many studies have demonstrated increased circulating mediators of inflammation in heart failure, including tumor necrosis factor-α, interleukin-2, and interleukin-6 (2, 4). In this study many of the intracellular signaling pathways involving the mortality genes, such as the Fc Epsilon R1 pathway and the T cell receptor signaling pathway, are directly involved in the production of tumor necrosis factor-α and interleukins.

Our results suggest that gene expression profiling in peripheral blood may predict prognosis in patients with heart failure. This approach, if validated in a larger cohort, could have direct implications in patient management. Several serum biomarkers have been shown to be elevated in heart failure and positively correlate with adverse patient outcomes. While these biomarkers may be predictive, they can be difficult to measure, with short half lives and wide sample variability, thus often limiting utility in standard clinical practice. While the commonly used heart failure biomarker brain natriuretic peptide has been demonstrated to be a sensitive discriminator in patients suspected of acute heart failure (10), a recent, large randomized trial showed brain natriuretic peptide-guided therapy to be of no benefit over standard clinical management (16). As such, accurate and reproducible biomarkers that can be used in the management and prognosis of patients with chronic heart failure remains an unmet need. As demonstrated in this study, gene expression profiling and in particular when combined

Table 3. Risk analysis with NYHA class and T cell receptor gene clustering

<table>
<thead>
<tr>
<th>NYHA</th>
<th>T Cell Receptor Gene Clustering</th>
<th>IRI</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk (n = 24)</td>
<td>control</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13 M/11 F</td>
<td>control</td>
<td>group2</td>
<td>12</td>
</tr>
<tr>
<td>100% survival</td>
<td>control</td>
<td>group3</td>
<td>2</td>
</tr>
<tr>
<td>Intermediate risk (n = 36)</td>
<td>I</td>
<td>group1</td>
<td>2</td>
</tr>
<tr>
<td>22 M/14 F</td>
<td>I</td>
<td>group2</td>
<td>13</td>
</tr>
<tr>
<td>94% survival</td>
<td>II</td>
<td>group3</td>
<td>5</td>
</tr>
<tr>
<td>High risk (n = 26)</td>
<td>III</td>
<td>group1</td>
<td>7</td>
</tr>
<tr>
<td>20 M/6 F</td>
<td>III</td>
<td>group2</td>
<td>6</td>
</tr>
<tr>
<td>49% survival</td>
<td>IV</td>
<td>group3</td>
<td>11</td>
</tr>
</tbody>
</table>

IRI, integrated risk index; M, male; F, female. See text for discussion.
with NYHA class may be of clinical utility in assessing the long term prognosis of patients with heart failure.

The small size of our control group and the larger number of female subjects in the control group limits our findings to some extent, and the study needs to be reproduced in a larger sample. However, no differences between males and females were found in mortality, our end point.

The death rate from cardiovascular diseases has steadily decreased over the past two decades. However, greater survival combined with the aging of the population has directly contributed to the increasing incidence of heart failure in western countries. Although treatment of heart failure has advanced, little progress has been made on the development of biological metrics that aid in predicting prognosis in patients with heart failure and optimizing their management. Blood gene expression profiling particularly when combined with the clinical assessment of NYHA class may provide a better delineation of patient’s risk in heart failure, and thus be a valuable tool in optimizing medical management.

ACKNOWLEDGMENTS

We thank Patricia Bauman and Kelly Begin for technical assistance.

GRANTS

This work was supported in part by National Heart, Lung, and Blood Institute Grant 5R01HL-077637 (P. VanBuren).

DISCLOSURES

C.-C. Liew and J. Ma are both employees and C.-C. Liew is chief scientist of GeneNews, which partly sponsored this research.

REFERENCES