Transcriptional Genomics Associates FOX Transcription Factors With Human Heart Failure

Sridhar Hannenhalli, PhD; Mary E. Putt, PhD, ScD; Joan M. Gilmore, BS; Junwen Wang, PhD; Michael S. Parmacek, MD; Jonathan A. Epstein, MD; Edward E. Morrisey, PhD; Kenneth B. Margulies, MD; Thomas P. Cappola, MD, ScM

Background—Specific transcription factors (TFs) modulate cardiac gene expression in murine models of heart failure, but their relevance in human subjects remains untested. We developed and applied a computational approach called transcriptional genomics to test the hypothesis that a discrete set of cardiac TFs is associated with human heart failure.

Methods and Results—RNA isolates from failing (n=196) and nonfailing (n=16) human hearts were hybridized with Affymetrix HU133A arrays, and differentially expressed heart failure genes were determined. TF binding sites overrepresented in the −5-kb promoter sequences of these heart failure genes were then determined with the use of public genome sequence databases. Binding sites for TFs identified in murine heart failure models (MEF2, NKX, NF-AT, and GATA) were significantly overrepresented in promoters of human heart failure genes (P<0.002; false discovery rate 2% to 4%). In addition, binding sites for FOX TFs showed substantial overrepresentation in both advanced human and early murine heart failure (P<0.002 and false discovery rate <4% for each). A role for FOX TFs was supported further by expression of FOXC1, C2, P1, P4, and O1A in failing human cardiac myocytes at levels similar to established hypertrophic TFs and by abundant FOXP1 protein in failing human cardiac myocyte nuclei.

Conclusions—Our results provide the first evidence that specific TFs identified in murine models (MEF2, NKX, NFAT, and GATA) are associated with human heart failure. Moreover, these data implicate specific members of the FOX family of TFs (FOXC1, C2, P1, P4, and O1A) not previously suggested in heart failure pathogenesis. These findings provide a crucial link between animal models and human disease and suggest a specific role for FOX signaling in modulating the hypertrophic response of the heart to stress in humans. (Circulation. 2006;114:1269-1276.)

Key Words: genomics ■ heart failure ■ hypertrophy ■ remodeling ■ transcription factors

Cardiac myocyte hypertrophy is a central feature of heart failure, the most common reason for adult hospitalization in the industrialized world.1 Research performed over the past decade has elucidated a complex network of signaling pathways and transcription factors that converge and modulate the hypertrophic response in murine models.2,3 By contrast, the role of transcription factors in human heart failure remains untested because of methodological limitations for assessment of transcription factor function in human subjects. As a result, very little of our detailed understanding of cardiac transcription factors in murine models has been translated into an understanding of human heart failure.

Clinical Perspective p 1276

We have developed a computational approach, which we call transcriptional genomics, that provides fundamental insights into transcription factor function in the failing human heart by integrating gene expression data from microarray experiments with readily available genome sequence data. We used this approach to perform an unbiased assessment of transcription factors associated with advanced heart failure in a large human cohort and with early postinfarction heart failure in a murine model. We hypothesized that despite thousands of observed gene expression changes, a discrete set of cardiac transcription factors regulates myocardial gene expression in the pathogenesis of heart failure.

Methods

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.

Transcriptional Genomics Overview

Our approach analyzes gene expression data with the following question in mind: Among all genes that are changed between 2 states (heart failure and control), is there evidence that particular transcription factors are directing the observed changes? We can answer this...
question by using a computational approach that incorporates promoter sequences and transcription factor binding sites into the gene expression analysis (Figure 1). First, a set of cardiac genes is determined from microarray data by selecting genes expressed above background levels in the myocardium. Each cardiac gene is then mapped to its corresponding promoter region sequence, defined as the 5 kb of genomic sequence upstream from the transcription start site, with the use of the RefSeq genome sequence database. Transcription factor binding sites are determined within these promoters with the TRANSFAC database of vertebrate transcription factor binding sites, with a focus on promoter regions that show human-mouse evolutionary sequence conservation. We then determine which binding sites are statistically overrepresented among genes that show altered expression in heart failure. In essence, our approach regards a set of differentially expressed heart failure genes as reporter genes and asks which transcription factors are driving the expression changes. We applied this approach to human and murine cardiac gene expression data as described below.

**Advanced Human Heart Failure**

Human myocardium was collected by a protocol approved by our institutional review board. Myocardium was obtained from patients undergoing heart transplantation for advanced systolic heart failure (n = 194) and from nonfailing controls deemed unsuitable for transplantation (n = 16), as described previously.

Patients had heart failure due to either ischemic (n = 86) or idiopathic dilated (n = 108) cardiomyopathy. No subjects received mechanical support with left ventricular assist devices. All heart failure patients had New York Heart Association class 3 to 4 symptoms and left ventricular systolic dysfunction, with mean ± SD ejection fraction of 14 ± 8%. Nonfailing controls had normal left ventricular function with mean ejection fraction of 56 ± 7% (P = 0.0001 versus failing). Ages were comparable in subjects with heart failure (57 ± 12 years) and nonfailing controls (54 ± 12; P = 0.3).

At the time of transplantation or donor harvest, whole hearts were removed after preservation in cold cardioplegia, and segments of noninfarcted left ventricular free wall were snap-frozen in liquid nitrogen. For each sample, RNA was isolated with the use of the Trizol reagent, and cRNA probes were prepared using an in vitro transcription reaction. Probes were then hybridized with individual Affymetrix HGU133A arrays according to the manufacturer’s instructions. Arrays were scanned and converted to .cel file data with the use of MAS 5.0 software. Data are available for download via Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE4006. A subset of these data (n = 157 failing, n = 14 control) was analyzed in a previous study that focused on reverse remodeling induced by left ventricular assist devices.

In addition, cardiac myocytes were isolated from selected patients by methods previously described, enriched via gentle centrifugation, and frozen in liquid nitrogen. RNA was subsequently isolated from these specimens with the use of the Trizol reagent for subsequent quantitative reverse transcriptase–polymerase chain reaction (RT-PCR) analysis.

**Murine Postinfarction Heart Failure**

We used a publicly available gene expression dataset from the Genomics of Cardiovascular Development, Adaptation, and Remodeling, NHLBI Program for Genomic Applications, Harvard Medical School. Details about the experimental protocol to induce myocardial infarction reside at http://cardiogenomics.med.harvard.edu/groups/proj1/pages/mi_web.html. Briefly, FVB mice were subjected either to thoracotomy and surgically induced myocardial infarction (n = 18) or to sham thoracotomy alone (n = 18), and expression profiles from noninfarcted regions of the left ventricle were measured with the use of Affymetrix MGLU74AV2 arrays at specific time points as heart failure developed (1 hour, 4 hours, 24 hours, 48 hours, 1 week, and 8 weeks; n = 3 arrays at each time point in each group). Data were downloaded as .cel files for subsequent analysis.

**Statistical Analysis of Microarray Data**

Analyses were performed in R software version 1.9-2.2 (www. R-project.org). All .cel files were normalized with the use of robust multivariate analysis (RMA) in the affy package version 1.6.7 (www.bioconductor.org). For both the human and the murine data, probe sets were removed if they displayed expression values |log2| > 6.5 on all arrays. This filtering yielded sets of cardiac genes present well above background levels in either the human or the murine heart. For the human data, we used significance analysis of microarrays (SAM), implemented with the use of the siggenes package in R, to select genes that were differentially expressed in failing hearts compared with nonfailing controls. These analyses were stratified by cause of heart failure, ie, ischemic or idiopathic cardiomyopathy.

For the murine data, we focused on cardiac genes whose temporal expression patterns differed between the infarcted and the sham-operated mice. A 2-way ANOVA model that included terms for the effects of time, group (infarcted versus sham), and time-by-group interaction was fit to the data for each gene. Cardiac genes were initially selected on the basis of the probability value for the F test of the time-by-group interaction with a conservative threshold of P < 0.0001. These genes thus showed the strongest evidence of changes over time that differed between the sham and infarct groups. We then used a latent class method to determine the predominant temporal patterns of gene expression as heart failure developed (see Methods in online-only Data Supplement). Four temporal patterns emerged from this analysis, all with striking changes at the 1-week time point. For this reason, and to parallel the analysis of a single time point in the human data, we performed a single 2-group SAM analysis comparing infarcted and sham mice at the 1-week time point alone.

**Transcription Factor Binding Sites in Promoters of Cardiac Genes**

Using our previously published, phylogenetic-footprinting approach, we determined a comprehensive database of putative transcription factor binding sites for both the human and the murine cardiac gene sets. Briefly, for each gene, we extracted 5 kb of genomic sequence immediately upstream from the transcription start site using the RefSeq genome sequence database (www.genome.ucsc.edu). We searched these promoter regions using the 546 transcription factor binding site motifs obtained from the TRANSFAC database version 8.4. A binding site motif is represented as a "position weight matrix" (PWM), which is a 4× k matrix for a k-bases-long binding site and provides, for each of the 4 positions,
the preferences for the 4 nucleotide bases at that position. Matches between TRANSFAC PWMs and promoter regions of cardiac genes were determined with the use of the tool PWMSCAN.\(^1\) The match for a PWM was probability value cutoff of \(2 \times 10^{-4}\), corresponding to an expected frequency of \(\leq 1\) random match in 5 kb. We filtered these matches further using human-mouse genome sequence alignments to focus on promoter regions that showed evolutionary conservation. For each TRANSFAC match, let \(c\) be the fraction of binding site bases that were identical between human and mouse. We retained matches such that either probability value \(\leq 0.00002\) (expected frequency of 1 in 50 kb) or \(c \geq 0.8\). These criteria for matching have been evaluated previously and were shown to accurately detect experimentally verified binding sites with a low false-positive rate of 1 random match in every 50 kb of genomic sequence searched.\(^1\)

### Overrepresented Transcription Factor Binding Sites in Heart Failure

We determined overrepresentation of TRANSFAC binding sites by counting the frequency with which a given binding site was present in promoters of differentially expressed heart failure genes relative to the frequency in the reference set of cardiac genes (Figure 1). Statistical significance was then assessed with the use of random resampling or permutation tests. For each set of heart failure genes, let \(P\) denote the corresponding set of 5 kb upstream promoter sequences of these genes. Let \(C\) represent the promoter sequences for the entire set of cardiac-specific genes. For each of the 546 transcription factor PWMs in TRANSFAC, define overrepresentation of the PWM \(x_i\) as:

\[
S_i = \frac{\text{number of } x_i \text{ in } P \cap |C|}{\text{number of } x_i \text{ in } C} \cdot \frac{|P|}{|C|}
\]

where \(|P|\) and \(|C|\) are the number of sequences in \(P\) and \(C\), respectively. Let \(P]\) be a set of \(P\) sequences, randomly selected from \(C\). Analogous to \(s_i\), we calculate overrepresentation \(s_i'\) in \(P\) relative to \(C\). Assume that \(s_i \geq 1\). In 1000 such random samplings, the fraction of times in which the overrepresentation \(s_i' \geq s\) estimates the significance of \(s_i\). Similarly, if \(s_i < 1\), we call this an underrepresentation and in this case the fraction of random samples in which the \(s_i' < s\) estimates the significance of underrepresentation \(s_i\). A 2-sided probability value, which estimates the chance probability of the observed data, or a more extreme result, is constructed by doubling the overrepresentation or underrepresentation significance level and truncating, if needed, at 1.0. To account for multiple comparisons, false discovery rates were estimated from all probability values by a direct approach.\(^1\) We report here only the results for those PWMs with 2-sided probability values \(\leq 0.002\), which corresponds to an false discovery rate (FDR) of \(<5\%\) in all datasets (Tables 1 and 2).

We used \(Z\) score transformation of the overrepresentation ratio as a way of quantifying the degree of overrepresentation of the selected PWMs relative to all 546 PWMs in TRANSFAC. Given the overrepresentation values \(s_i\) for the \(i\)th PWM, 1\(\leq i \leq 546\), we computed the sample mean (\(\bar{s}\)) and the sample SD (\(\sigma\)) of the overrepresentation values for the 546 PWMs. We define \(Z\) score as follows:

\[
Z_i = \frac{s_i - \bar{s}}{\sigma}
\]

Thus, the \(Z\) score indicates the number of SDs of overrepresentation for a given transcription factor binding site.

### Quantitative RT-PCR

RNA isolates from human cardiac myocytes were treated with DNase to remove any contaminating genomic DNA and subsequently converted to cDNA with an in vitro transcription reaction. cDNAs were used as templates for Taqman RT-PCR with ABI Assays-on-Demand on an ABI 7500 sequence detection system. The specific assays used were FOXC1 (Hs00559473_m1), FOXC2 (Hs00270951_s1), FOXD4L1 (Hs00846606_s1), FOXH1 (Hs00182690_m1), FOXJ1 (Hs00230964_m1), FOXJ2 (Hs00218236_m1), FOX13 (Hs00208978_m1), FOXK2 (Hs00189612_m1), FOXM1 (Hs00153543_m1), FOXO1A (Hs00231106_m1), FOXO3A (Hs00818121_m1), FOXP1 (Hs00212860_m1), FOXP4 (Hs00405889_m1), FOXP1 (Hs00271535_m1), FOXD2 (Hs00231149_m1). All samples were run in duplicate, and 18s rRNA (Hs99999901_s1) was used as an internal control to normalize transcript abundance. Duplicates were averaged to calculate an expression value for each sample.

### Immunohistochemistry

Paraffin-embedded failing human heart tissue was sectioned and stained with the polyclonal anti-FOX1 antibody at 1:500 dilution as previously described.\(^3\)

### Results

#### Advanced Human Heart Failure

Using the entire dataset of 210 human hearts, we determined a set of 13 639 probe sets expressed above background levels in the heart that mapped to 8440 unique cardiac genes in the RefSeq genome sequence database (Table I in online-only Data Supplement). For each of these cardiac genes, we...
extracted 5 kb of 5' flanking genomic sequence, containing the promoter sequence, from the RefSeq database (www.genome.ucsc.edu). We identified transcription factor binding sites in these promoter regions using human-mouse sequence conservation and the 546 PWMs for vertebrate transcription factors in the TRANSFAC. We then determined which binding sites were overrepresented among differentially expressed heart failure genes compared with the entire set of 8440 cardiac genes. Statistical significance was assessed with the use of random resampling, and FDRs were calculated to correct for multiple comparisons.

We focused on the 2 most common forms of advanced heart failure: idiopathic dilated cardiomyopathy and ischemic cardiomyopathy. Using SAM, we identified 1002 and 1020 probe sets that were differentially expressed in idiopathic and ischemic cardiomyopathies, respectively, with FDR = 0.01% (Table I in online-only Data Supplement). These probe sets subsequently mapped to 815 and 808 unique genes in the RefSeq database. We then determined overrepresented binding sites separately for cardiac genes that were overexpressed or underexpressed in ischemic or nonischemic cardiomyopathy (4 analyses in total). As shown in Table 1, several binding sites were significantly overrepresented in promoters of overexpressed heart failure genes (with resampling probability value <0.002 and FDR of 2% to 4%). Of note, underexpressed cardiac genes did not show statistically significant enrichment of binding sites in their promoters.

There is a great deal of sequence similarity among DNA binding sites for related transcription factors. To account for this, we grouped transcription factors into families on the basis of the sequence similarity among their TRANSFAC PWNs (Figure 2 and Methods in the online-only Data Supplement). Overrepresentation of any one binding site in a TRANSFAC family suggests increased activity of 1 or more of the family members (Table 1), but determining which specific transcription factor is not possible by computational means alone. As such, these analyses serve to identify transcription factor families rather than specific factors that are activated in the failing heart. Figure 3 summarizes the degree of overrepresentation of TRANSFAC families among differentially expressed genes in human heart failure. Several families that have been implicated in the pathogenesis of cardiac hypertrophy in animal models (MEF-2, NKX, NF-}

<table>
<thead>
<tr>
<th>TRANSFAC PWM</th>
<th>Binding Site</th>
<th>TRANSFAC Family</th>
<th>1 Week (FDR = 0.04)</th>
<th>1 Week (FDR = 0.04)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00451</td>
<td>NKX3A</td>
<td>23</td>
<td>Nkx</td>
<td>X</td>
</tr>
<tr>
<td>M00292</td>
<td>Freq-A</td>
<td>34</td>
<td>FOX</td>
<td>X</td>
</tr>
<tr>
<td>M00232</td>
<td>MEF-2</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00407</td>
<td>RSRFC4</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00269</td>
<td>RSRFC4</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00495</td>
<td>Bach1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00517</td>
<td>AP-1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00926</td>
<td>AP-1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00657</td>
<td>PTF1-β</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FDR indicates false discovery rate for the entire column; X, P<0.002 by permutation. Analyses not shown (latent classes 1 to 4) had FDRs >0.05. TRANSFAC families group PWNs that share a high degree of sequence similarity and are indicated by a number and name (see Table III in the online-only Data Supplement).

---

**Table 2. Overrepresented Transcription Factor Binding Sites in Promoters of Murine Heart Failure Genes Early After Myocardial Infarction**

<table>
<thead>
<tr>
<th>TRANSFAC PWM</th>
<th>Binding Site</th>
<th>TRANSFAC Family</th>
<th>1 Week (FDR = 0.04)</th>
<th>1 Week (FDR = 0.04)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00451</td>
<td>NKX3A</td>
<td>23</td>
<td>Nkx</td>
<td>X</td>
</tr>
<tr>
<td>M00292</td>
<td>Freq-A</td>
<td>34</td>
<td>FOX</td>
<td>X</td>
</tr>
<tr>
<td>M00232</td>
<td>MEF-2</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00407</td>
<td>RSRFC4</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00269</td>
<td>RSRFC4</td>
<td>50</td>
<td>MEF-2</td>
<td>X</td>
</tr>
<tr>
<td>M00495</td>
<td>Bach1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00517</td>
<td>AP-1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00926</td>
<td>AP-1</td>
<td>62</td>
<td>AP-1</td>
<td>X</td>
</tr>
<tr>
<td>M00657</td>
<td>PTF1-β</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FDR indicates false discovery rate for the entire column; X, P<0.002 by permutation. Analyses not shown (latent classes 1 to 4) had FDRs >0.05. TRANSFAC families group PWNs that share a high degree of sequence similarity and are indicated by a number and name (see Table III in the online-only Data Supplement).

**Figure 2.** TRANSFAC families. We assessed the specificity of TRANSFAC PWMs by clustering all 546 PWMs on the basis of their degree of sequence similarity (see online-only Data Supplement). Similar PWMs clustered into families that correspond to related transcription factors. Here GATA and FOX families are displayed, with a complete display for the entire TRANSFAC v8.4 database in Figure 1 in the online-only Data Supplement.

**Figure 3.** Overrepresentation of transcription factor families in advanced human heart failure. We mapped the ratios of overrepresentation for each PWM to a Z score, which indicates the number of SDs of overrepresentation of each PWM. The median Z score for TRANSFAC families listed in Table 1 is shown. Families previously implicated in murine models of cardiac hypertrophy (black bars) were substantially overrepresented in human heart failure. Novel families (gray bars), including the FOX family, also showed substantial overrepresentation, suggesting activation of FOX signaling in human heart failure.
AT, and GATA; reviewed by Frey and Olson2) were >3 SDs overrepresented in the promoters of overexpressed heart failure genes. Interestingly, transcription factors not previously suggested in the pathogenesis of human heart failure were also implicated. Most notable among these was the FOX family of transcription factors, which showed nearly 4 SDs of overrepresentation. These data suggest increased activity of 1 or more FOX transcription factors in the failing human heart.

To assess the robustness of our method to identify transcription factor families involved in heart failure, we performed a sensitivity analysis, in which we varied the thresholds to select the cardiac-specific gene set and the heart failure gene sets (Methods in the online-only Data Supplement). The MEF2, NKX, NF-AT, GATA, and FOX families were implicated at all thresholds tested, indicating robust results. A byproduct of our analysis is a database of genes that both are induced in human heart failure and are putative transcriptional targets of factors activated in human heart failure (http://www.cagr.pcbi.upenn.edu/CardioTransGen-Suppl2/). These data include the location of transcription factor binding sites in the promoter regions of differentially expressed human heart failure genes.

Murine Postinfarction Heart Failure
For our analysis in human subjects, mRNA was obtained at a single point in time from patients with advanced heart failure. Therefore, it is not possible to discern whether activation of the identified transcription factors occurs before, during, or in response to the development of heart failure. To determine whether the factors in Table 1 are activated early in heart failure progression, we used a dataset from a murine model of heart failure in which transcriptional changes were assessed in response to controlled myocardial infarction (www.cardiogenomics.org). Using the entire dataset of 38 murine hearts, we determined a set of 9268 cardiac-specific probe sets that mapped to 5723 unique genes in the RefSeq database (Table II in the online-only Data Supplement). We focused on the subset of these cardiac genes whose temporal expression patterns differed between the infarcted and the sham-operated mice. Figure 4 demonstrates the temporal expression patterns of the selected genes with the use of a latent class method to group genes with similar patterns. Striking changes in gene expression occurred 1 week after myocardial infarction in all 4 latent classes. We then analyzed genes within each of the 4 classes, and at the 1-week time point alone, to determine overrepresented transcription factor binding sites. Only the analysis of the 1-week time point yielded results with FDR <5%. As expected, transcription factor families previously identified in the hypertrophic response in murine models, such as MEF2, NKX, and AP-1, were identified by this approach (Table 2 and Figure 5). As in the advanced human heart failure cohort, there was substantial overrepresentation of FOX binding sites, suggesting activation of FOX signals early in the pathogenesis of heart failure.

Figure 4. Longitudinal patterns of gene expression in noninfarcted segments of murine myocardium after myocardial infarction (n=3 infarcted mice and 3 sham controls at each time point). Genes were selected that changed differently over time in infarcted mice compared with sham controls and then were grouped into classes that shared similar temporal patterns. Data are displayed as a ratio of mean-centered expression levels in infarcted mice compared with control. Four classes were apparent, containing 167, 82, 166, and 125 genes, respectively (Table II in the online-only Data Supplement). In all classes, the most striking transcriptional changes occurred 1 week after myocardial infarction. Each class, as well as the 1-week time point alone, was then subjected to transcriptional genomics analysis (Table 2).

FOX Transcription Factors in Human Heart Failure
The FOX family of transcription factors is large, with at least 44 members identified in humans.14 Many FOX factors are critical regulators of tissue-specific gene expression and morphogenesis, including cardiovascular development.15–19 To identify candidate FOX factors associated with heart failure, we determined which FOX factors were present in human myocardium. On the basis of microarray data, a minimum of 15 FOX factors were present in left ventricular myocardium at levels comparable to those of established hypertrophic factors such as MEF2 (Figure 6A). To determine whether changes in expression of FOX factors themselves occurred in the failing heart compared with control, we returned to our SAM analysis (Table II in the online-only Data Supplement). There were no changes in expression of FOX factors, nor were there changes in expression of NFAT, GATA, MEF2, and NKX, with the exception of minor increases in MEF2A and MEF2C. These findings suggest that mechanisms other than alterations in transcription factor expression are responsible for our observations and are
consistent with basic studies indicating that stress signals regulate hypertrophic transcription factors via posttranslational mechanisms such as phosphorylation, transport into and out of the nucleus, or interaction with coactivators and repressors.

These data suggest a role for specific FOX transcription factors but do not determine the specific cell types involved. We addressed this by performing quantitative RT-PCR in isolated human cardiac myocytes (Figure 6B). A subset of FOX factors were expressed at levels similar to MEF2, including FOXJ2, FOXO1A, FOXP1, FOXP4, FOXC1, and FOXC2. By immunohistochemistry, FOXP1 protein was present in nuclei of human cardiac myocytes at high levels, making FOXP1 a particularly promising candidate for modulating the hypertrophic response in human heart failure (Figure 7).

**Discussion**

Our findings provide the first evidence associating MEF2, NKX, GATA, and NF-AT transcription factors with common forms of human heart failure, thereby providing a crucial link between a decade of research in murine models and human disease. We accomplished this by developing and applying a computational method that integrates multiple sources of genomic data (Figure 1). Using this same approach, we determined that novel transcription factors, in particular the FOX family of transcription factors, are associated with early heart failure in mice and advanced heart failure in humans. This finding is supported further by the expression of specific FOX factors at levels similar to established hypertrophic factors in the failing human heart and by abundant expression of FOXP1 protein in nuclei of failing human cardiac myocytes. Taken together, these data suggest that FOX signaling may play a critical role in human heart failure.

Previous studies have suggested a role for FOXO transcription factors, acting downstream from GH/IGF-1/AKT signals, in the regulation of cell and organ size during normal development and aging. In particular, exogenously transfected FOXO3A has the capacity to regulate myocyte size in rodent models. However, a link between endogenous FOX signals and pathological hypertrophy in humans has not been demonstrated previously. Moreover, loss-of-function studies have shown that other members of the FOX family modulate specific aspects of cardiac development, including FOXC1,
C2, H1, J1, P1, and P4. Four of these 6 factors were identified as candidate regulators of pathological hypertrophy in humans by our analysis (Figure 6B). In particular, loss of FOXP1 results in defective cardiomyocyte proliferation and differentiation. Thus, we propose that signaling by a specific subset of FOX factors may play an important role in adult human heart disease, a finding that adds to a paradigm in which pathological cardiac hypertrophy is a recapitulation of fetal/developmental pathways in an effort to compensate for injury.

Cardiac transcription factors integrate a wide range of upstream stress signals, including mechanical load, neurohormones, and cytokines, and therefore represent a point of convergence in the causal pathway from cardiac stress to cardiac remodeling and failure. Table 1 supports this concept as there is substantial overlap in the transcription factors identified despite different heart failure causes. However, we also found that a minority of factors were exclusively associated with either idiopathic or ischemic cardiomyopathy. NFAT was associated with idiopathic cardiomyopathy, suggesting a more prominent role for abnormal calcium signaling and calcineurin-mediated hypertrophy, whereas C/EBP, which is modulated by inflammation and mitogen-activated protein kinases, was specifically associated with ischemic heart failure. These cause-specific findings corroborate previous work showing that distinct subsets of genes are altered in different types of human heart failure.

Previous studies of the human cardiac transcriptome have determined sets of genes associated with heart failure. However, interpretation of such studies can be problematic because of the enormous number of gene associations identified, small sample sizes, heterogeneous patient characteristics, and different approaches to data analysis. We sought to overcome these limitations by shifting the focus away from gene expression changes themselves and asking whether the observed changes in gene expression, taken in aggregate, suggest activation of specific transcription factors. Of note, Zhao et al piloted a similar method to study transcription factors associated with cardiac atria and ventricles in mice. These approaches distill a large body of expression data into discrete results within the framework of conserved transcription factor/target gene relationships and can be used to study a broad range of disease phenotypes.

Our analyses are strengthened by a large patient cohort, a rigorous analytical approach, and congruent findings in animal models and humans. They are limited by incomplete knowledge of transcription factor binding sites, but it is likely that most transcription factor families are well represented in TRANSFAC and should thus be captured by our approach. More importantly, regulation of cardiac gene expression is a complex phenomenon mediated by interactions among transcription factors, coactivators, corepressors, and epigenetic modifications such as histone acetylation. Hence, the specific mechanisms responsible for our findings cannot be determined by computational means alone. For example, our results with regard to FOX transcription factors could be explained by activation of specific FOX factors or by inactivation of FOX corepressors. Discerning these underlying mechanisms will require experimental approaches that at this time are feasible only in model systems. Mechanistic studies in human subjects await the development of pharmacological agents that alter function of specific transcription factors to test whether such alterations change the course of heart failure. Until then, transcriptional genomics and other integrative approaches may well provide the critical platform necessary to identify therapeutic targets and to extend our understanding of complex pathways in animal models to an understanding of molecular events in human heart disease.

**Sources of Funding**

This study was supported by research grants from the National Institutes of Health (K23HL071562, P20RR020741) and from Gene-Logic and by an allocation from the Commonwealth of Pennsylvania.

**Disclosures**

None.

**References**

Cardiac myocyte hypertrophy is a central feature of virtually all forms of heart failure. A large body of work in animal models indicates that stresses such as pressure or volume load, neurohormones, cytokines, and reactive oxygen species promote myocyte hypertrophy via activation of a discrete set of cardiac transcription factors. These “hypertrophic factors” in turn alter cardiac gene expression to cause cardiac myocyte hypertrophy and failure. Despite these findings in animals, the role of cardiac transcription factors in human heart failure has been largely untested because of limited methods to assess transcription factors in human subjects. Our study used novel genomic approaches to test whether specific cardiac transcription factors were associated with human heart failure due to ischemic or idiopathic dilated cardiomyopathy. We found that transcription factor families previously identified in murine models of heart failure (MEF2, NKK, NFAT, and GATA) were strongly associated with human heart failure. In addition, our analysis suggested a role for specific members of the FOX family of transcription factors (FOXC1, C2, P1, P4, and O1A) not previously suggested in heart failure pathogenesis. This research provides some of the first evidence in humans supporting a role for hypertrophic transcription factors in the pathogenesis of heart failure and identifies FOX transcription factors as new candidates for further research. From a clinical standpoint, this study provides a “translational” link between animal models and human disease and supports the idea that cardiac hypertrophy and cardiac gene expression are potential therapeutic targets in human heart failure.
Transcriptional Genomics Associates FOX Transcription Factors With Human Heart Failure
Sridhar Hannenhalli, Mary E. Putt, Joan M. Gilmore, Junwen Wang, Michael S. Parmacek, Jonathan A. Epstein, Edward E. Morrisey, Kenneth B. Margulies and Thomas P. Cappola

Circulation. 2006;114:1269-1276; originally published online September 4, 2006;
doi: 10.1161/CIRCULATIONAHA.106.632430

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/114/12/1269

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2006/08/31/CIRCULATIONAHA.106.632430.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/
Supplementary Figure 1. Dendrogram showing Transfac PWM Clusters. The 546 positional weight matrices were grouped into 103 clusters based on sequence similarity as described in Methods. Cluster assignments for each PWM are in Supplementary Table 3. To display the relationships among the clusters, we constructed a neighbor-joining tree of TRANSFAC PWMs based on the PWM-pair similarity. Clusters of highly similar PWMs are apparent as groups connected by short branches in the dendrogram. Since PWMs are highly similar within a given family, over-representation of any member implicates the entire family rather than a specific member of the family. Each branch is labeled with a PWM ID and the corresponding transcription factor name. Zoom in to view individual labels. Branches for the GATA and FOX families of transcription factors are displayed in Figure 2 of the manuscript.