Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Evidence for co-regulation of myocardial gene expression by MEF2 and NFAT in human heart failure

Mary E. Putt, Sridhar Hannenhalli, Yun Lu, Philip Haines, Hareesh R. Chandrupatla, Edward E. Morrisey, Kenneth B. Margulies and Thomas P. Cappola

Circ Cardiovasc Genet published online Mar 31, 2009;
DOI: 10.1161/CIRCGENETICS.108.816686
Circulation: Cardiovascular Genetics is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
Copyright © 2009 American Heart Association. All rights reserved. Print ISSN: 1942-325X. Online ISSN: 1942-3268

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

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at
http://circgenetics.ahajournals.org/subscriptions/

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail:
journalpermissions@lww.com

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

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Evidence for co-regulation of myocardial gene expression by MEF2 and NFAT in human heart failure

Mary E. Putt, PhD, ScD[1], Sridhar Hannenhalli, PhD[2], Yun Lu, PhD[1], Philip Haines, MBBCh, MPH[3], Hareesh R. Chandrupatla, MS[3], Edward E. Morrisey, PhD[3], Kenneth B. Margulies, MD[3]

Thomas P. Cappola, MD, ScM[3]*

Short Title: MEF2 / NFAT co-regulation in human heart failure

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Author Affiliations:

[1]Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics,

[2]Department of Genetics and Penn Center for Bioinformatics, and

[3]Penn Cardiovascular Institute, all at the University of Pennsylvania School of Medicine

*Correspondence:
Thomas, Cappola, MD, ScM
Perelman Center for Advanced Medicine
2nd Floor East Pavilion
3400 Civic Center Blvd.
Philadelphia, PA 19104
Phone: 215-615-0805
Fax: 215-615-0829
Email: thomas.cappola@uphs.upenn.edu

Word count: 6,021 total. 245 Abstract

Subject Heads: genomics, gene expression, transcription factors, heart failure

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Abstract:

Background: Pathologic stresses induce heart failure in animal models through activation of multiple cardiac transcription factors (TFs) working cooperatively. However, interactions among TFs in human heart failure are less well understood. Here we use genomic data to examine the evidence that five candidate TF families co-regulate gene expression in human heart failure.

Methods and Results: RNA isolates from failing (n=86) and non-failing (n=16) human hearts were hybridized with Affymetrix HU133A arrays. For each gene on the array, we determined conserved MEF2, NFAT, NKX, GATA, and FOX binding motifs within the -1 kb promoter region using human-murine sequence alignments and the TRANSFAC database. Across 9,076 genes expressed in the heart, TF binding motifs tended to cluster together in nonrandom patterns within promoters of specific genes (P-values ranging from 10^{-2} to 10^{-21}), suggesting co-regulation. We then modeled differential expression as a function of TF combinations present in promoter regions. Several combinations predicted increased odds of differential expression in the failing heart, with highest odds ratios noted for genes containing both MEF2 and NFAT bindings motifs together in the same promoter (peak OR 3.47, P=0.005).

Conclusions: These findings provide genomic evidence for co-regulation of myocardial gene expression by MEF2 and NFAT in human heart failure. In doing so, they extend the paradigm of combinatorial regulation of gene expression to the human heart and identify new target genes for mechanistic study. More broadly, we demonstrate how integrating diverse sources of genomic data yields novel insights into human cardiovascular disorders.

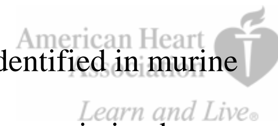
Keywords: heart failure; hypertrophy; remodeling; genes; transcription factors

American Heart Association Learn and Live
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Basic research has shown that pathologic stresses promote heart failure via activation of cardiac transcription factors (TFs)[1, 2]. These TFs integrate upstream stress signals and interact with each other and with co-regulators to induce a fetal gene program that mediates cardiac myocyte hypertrophy and failure. However, the role of TFs in human heart failure has been difficult to study because transgenic techniques used to study animal models cannot be applied to human subjects.

In a previous study we utilized computational approaches that integrate human gene expression data with genome sequence data in order to identify TFs associated with advanced human heart failure[3]. We found that many of the hypertrophic TFs identified in murine models were associated with human disease, and we identified a new association between forkhead (FOX) TFs and human heart failure. This initial study focused on identifying individual TFs associated with differentially expressed heart failure genes. However, there is substantial evidence from animal models that cardiac TFs function cooperatively with each other and with co-activators and repressors in their regulation of gene expression[4]. Here we extend our previous findings by examining the evidence that combinations of specific cardiac TFs (MEF2, NFAT, NKX, GATA, and FOX) regulate gene expression in human heart failure.



Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Methods

Patients

Human myocardium was collected using a protocol approved by our institutional review board. Myocardium was obtained from patients undergoing heart transplantation for advanced systolic heart failure due to idiopathic dilated cardiomyopathy (n=86) and from non-failing controls deemed unsuitable for transplantation (n=16) as described previously[5]. No subjects received mechanical support with left ventricular assist devices. All heart failure patients had New York Heart Association Class 3-4 symptoms and left ventricular systolic dysfunction, with mean ± standard deviation ejection fraction of 11±5%. Non-failing controls had normal left ventricular function with mean ejection fraction of 56±7% ($p=0.0001$ vs. failing by t-test). Subjects with idiopathic dilated cardiomyopathy (50±12 yrs) were slightly younger than controls (57±12; $p=0.03$ by t-test). RNA isolation and microarray hybridization with Affymetrix HU133A arrays were performed as described previously[5]. These data represent a subset of those in our previous analysis, which focused on associations between individual TFs and gene expression in heart failure[3]. Expression data are available for download via NCBI gene expression omnibus (GSE472).

American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Study Design

Our study design is summarized in Figure 1. For each gene on the Affymetrix HU133A array, the corresponding promoter region sequence was pulled from the UCSC genome database (www.genome.ucsc.edu) using the RefSeq gene ID, and the TRANSFAC database (www.transfac.de) was utilized to determine TF binding motifs within portions of these promoters that showed evolutionary conservation. We restricted our focus to binding motifs for the five TF families of interest: MEF2, NFAT, NKX, GATA, and FOX. Our

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

underlying assumption was that in order for TFs to cooperatively regulate the expression of a given gene, their binding motifs must occur together within the promoter region of that gene. Hence, we first examined patterns of co-occurrence of TF binding motifs in promoter regions across all genes expressed in the myocardium, and examined the evidence that these patterns were non-random. We then modeled differential expression as a function of specific combinations of TF binding motifs to identify groups of TFs that appear to regulate myocardial gene expression cooperatively. All analyses were carried out in R 2.6 (www.r-project.org; MASS and maanova libraries) as well as Bioconductor tools (www.bioconductor.org; affy library) and permutation programs developed specifically for the project[6].

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Integration of datasets

Affymetrix gene expression data were normalized and adjusted for background signal using robust multi-array analysis (RMA)[7]. Probesets with expression values lower than 91 units ($\log_2$=6.5) in all patients were removed. Based on prior work, this filtering removes genes that are near background levels of detection[3, 8]. These normalized, filtered datasets were then linked to promoter region binding motif data. For each unique gene represented on the HU133A array, we determined the -1 kb putative promoter sequences utilizing the RefSeq gene annotation in the UCSC database. We then utilized the positional weight matrices (PWMs) in the TRANSFAC database of vertebrate transcription factors to determine TF binding motifs within these promoters. Of note, in our prior study we used a 5 kb promoter region. Here we chose a 1 kb promoter region based on an analysis of the frequency of TF binding motifs (Supplementary Data). Our analysis utilized all PWMs in TRANSFAC that correspond to one of the five candidate TF families of interest.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Assignment of PWMs to TF families was based on our previously reported clustering of all PWMs TRANFAC[3]. The presence/absence call for a PWM in a promoter region was made using our previously reported, phylogenetic footprinting approach, which combines significance of PWM match score with the human-mouse conservation[3, 9].

Co-occurrence of TF binding motifs in promoters of cardiac genes

We first summarized the data according to the specific combinations of TFs that were identified in the promoters of cardiac genes. Based on the overall frequencies of each TF family across all genes, we computed the frequency of each TF combination expected under the assumption that the occurrences of each TF family were randomly, and independently, distributed across the genes. We then used an exact binomial test to test the null hypothesis that the observed frequency of genes for each specific TF combination was consistent with that expected under the random distribution model.

American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Assessment of Differential gene expression

To compare expression in failing versus non-failing hearts we computed two-sample T-statistics for each probeset using a shrinkage estimate of the variance[10]. Based on the magnitude of the maximum observed T statistic (Tmax) across probesets that mapped to each unique gene, we assigned each gene to one of two groups ('differentially expressed' or 'unchanged'). In considering how TFs regulate gene expression in heart failure, we hypothesized that TFs regulate multiple genes simultaneously, with the involved genes demonstrating a wide range of levels of differential expression. Thus rather than modeling a continuous measure of differential expression we used two cutpoints as a threshold for differential expression: $Tmax>2.00$ or $Tmax>3.16$ to designate a gene as either differentially

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

expressed or unchanged. Because our determination of differential expression is subject to error, we used two different thresholds to assess the sensitivity of our results to this determination. We repeated this procedure 5000 times after permuting the failing/non-failing status of the subjects using sampling without replacement. We used the resulting subject-level permutation-based Tmax values, an approach which accounts for dependencies among genes, to estimate the proportion of genes misclassified as differentially expressed (false discovery rate, FDR), the proportion of genes misclassified as unchanged, and the overall misclassification rate at the two thresholds for differential expression[11, 12].

Modeling differential expression as a function of TF combinations

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

The binary indicators described above were used as the outcome in logistic regression models to indicate a gene that demonstrated differential expression in failing versus non-failing hearts. Models were constructed to examine the association of different patterns of TF family occurrence and co-occurrence with the odds of differential expression. Associations were examined between the odds of differential expression and the presence of (1) at least one PWM associated with each specific TF family, (2) increasing numbers of TF families, and (3) specific combinations of TF families. The models were specified *a priori* and no variable selection was carried out. For all models, the reference group was genes that lacked PWMs from all five TF families.

The assumption of independent outcomes needed for valid inference from a logistic regression model is violated by the correlations between genes. To address this, we used a modified logistic regression approach, similar to that of the 'SAFE' procedure, based on permutations of subject status[13]. For models (2) and (3), global p-values were determined by comparing the deviance change for the null and full models from the observed data, to the

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

distribution of the deviance change for each of the 5000 subject-level permutations. If the global p-value achieved significance, subject-level permutation p-values for individual terms in each model were determined using Wald statistics.

We performed additional analyses to determine whether the five TF families chosen for this study revealed associations beyond randomly selected five-family TF combinations. We grouped all 546 TFs in TRANFAC into 190 unique TF families based on similarity of TF binding motifs as previously described[3]. Using 25000 replications, we first randomly sampled five TF families and determined the number of associated genes. For each replicate we computed the odds of differential expression for genes with the five TF families relative to the odds of differential expression for genes with none of the five TF families.

American Heart Association
Learn and Live®
Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Hypothesis Testing

Statistical significance was declared for p-values < .05, and marginal significance was noted for p-values between .05 and .10. Exact binomial tests were corrected for multiple hypothesis testing using a Bonferroni correction. For the logistic regression models, we addressed the issue of false positives by carrying out a single global test of significance for each model. Only if that test was significant, were test statistics for the individual terms in the model considered. To retain statistical power, p-values for these individual tests were not adjusted for multiple comparisons, but were instead used in combination with the odds ratio to determine combinations of TFs of particular interest.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Results

Co-occurrence of TF binding motifs in promoters of cardiac genes.

After normalization, filtering, and mapping of microarray data to RefSeq, 9076 genes were detected in human myocardium at levels above background. Tables 1 and 2 describe the prevalence of TF families within the promoters of these genes and the PWMs that comprise these families. TF families are represented by different numbers of unique PWMs in TRANSFAC; FOX was most heavily represented with a total of 27 PWMs, whereas NFAT was represented by only 2 and NKX by only 8. A total of 87% of the 9076 cardiac genes were associated with at least one of the five TF families of interest. GATA (62%) and FOX (40%) were most common whereas NFAT (13%) and MEF2 (19%) occurred less frequently. For each TF family, the median number of matching PWMs within a particular promoter region was one, except for the GATA family which had a median of two PWMs per promoter. Table 2 shows that the majority of genes (64%) were associated with either one or two TF families, whereas less than 1% of genes were associated with all five TF families.

Figure 2A indicates that the TF binding motifs of interest were not distributed randomly across the genes, but tended to cluster together within a subset of cardiac genes. In particular, the number of genes with binding sites for four or five TF families within the same promoter was substantially higher than expected by chance (~ 2-3 fold increase). Figure 2B shows this finding in more detail by comparing the observed versus expected frequency of TF co-occurrence within the same promoter for all possible combinations of our candidate TFs. For the majority of combinations with four or five TF families, the number of observed genes greatly exceeded that expected by chance, with Bonferroni-corrected p-values ranging from 10^{-3} to 10^{-21}. By contrast, the observed versus expected values tended to

American Heart Association Learn and Live®
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

be similar for combinations of three TFs, and for two or one TF, we often found that the observed value was smaller than expected. Thus, binding sites for the five cardiac TF families investigated tend to cluster together within promoters of specific cardiac genes, with strong evidence for higher order clusters of four or five TFs. We repeated our analysis using combinations of five randomly selected TFs and found that clustering of five binding motifs within promoters is rare and unlikely to occur by chance (Supplementary data). These findings raise the possibility that MEF2, NFAT, NKX, GATA, and FOX cooperatively regulate gene expression in the human heart.

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Differential gene expression

We next determined which among the 9076 cardiac genes studied showed evidence for differential expression in heart failure compared to controls, and which genes appeared unchanged. In practice, choosing a threshold for differential expression involves tradeoffs. Choosing a stringent threshold results in a differentially expressed group with little contamination by unchanged genes (i.e., low false discovery rate); however, a number of differentially expressed genes will be misclassified into the 'unchanged' group. Conversely, choosing a less stringent threshold correctly assigns more of the differentially expressed genes to the correct group, but at the price of increased contamination of the differentially expressed group by unchanged genes. Here we chose two thresholds, Tmax>2.0 and Tmax>3.2, carrying out separate analyses with the rationale that important results with respect to TF co-regulation should be consistent across both analyses.

Using a cutoff of Tmax>2.0 resulted in assignment of 3579 genes (39%) to the differentially expressed group compared to a substantially smaller number, 1772 genes (20%), for the more stringent cutpoint of Tmax>3.2 (Table 3). Based on subject-level

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

permutations, the stringent cutpoint of Tmax>3.2 yielded a low FDR of 1% for the differentially expressed group, but a high misclassification rate of 34% for the unchanged group, indicating that about one third of the genes classified as unchanged are actually differentially expressed. By contrast, using a more liberal cutpoint of Tmax>2 yielded an FDR of 9% for the differentially expressed group and a more reasonable misclassification rate of 18% for the unchanged group. The overall misclassification rate was lower for the more liberal Tmax >2 cutpoint.

Association between TF combinations and gene expression

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

After classifying genes as differentially expressed or unchanged, we then utilized a series of logistic regression models to test the association between presence of specific combinations of TF binding motifs in the promoter region and odds of differential expression in heart failure across all 9076 cardiac genes. Table 4 summarizes the odds of differential expression associated with the presence of each individual TF family and with the total number of TF families present in the promoter. Results for the two Tmax cutpoints were similar, although the OR and the associated statistical significance were generally greater for the higher cutpoint. Compared to a reference set of genes that do not contain any binding sites for our TFs of interest, the presence of at least one PWM from any of the candidate TF families was associated with a small, but statistically significant or marginally significant odds of increased differential expression. Increasing numbers of TF families were associated with increasing odds of differential expression (Global P=.035 for Tmax>2.0; Global P=.0006 for Tmax>3.2). Genes that contained PWMs from all five TF families had a 2.1 (Tmax>2) to 3.2 (Tmax>3.2) fold increase in the odds of differential expression compared to a reference set with none of the TF families of interest. Because increasing numbers of TF

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

families implies increasing numbers of PWMs (Table 2) we added a term describing the number of PWMs to the model for TF family number. The addition of this term to this model did not substantively alter the findings in Table 4 (data not shown). These results suggest that presence of multiple binding motifs together within the same promoter region substantially increases the likelihood of differential expression in the failing human heart.

Table 5 shows the association between specific TF family combinations and odds of differential expression. Here each gene was assigned to either the baseline reference group or to one of 32 categories shown in Table 5. We then used two multivariable logistic regression models (one for each Tmax threshold) to determine associations between these TF categories and odds of differential expression. Global P-values for each of the two models indicated evidence of significant associations for both thresholds (P=.028 for Tmax>2.0; P=.001 for Tmax>3.2). As with the earlier models (Table 4), the association with a specific TF combination was typically more pronounced for the higher Tmax threshold. Five combinations showed statistically significant associations ($P < 0.05$) at both thresholds and are indicated in bold in the table. Of these, the largest ORs were for the two-family combination of MEF2 with NFAT (OR=3.47/3.03 for Tmax>2.0/Tmax>3.2) and for all five TF families (OR=2.1/3.2). For comparison, we randomly generated sets of five TFs using data from TRANSFAC and repeated our analysis (Supplementary Data). Among random TF combinations with at least 50 associated genes, the median OR for association with differential expression was only 0.9, and the upper 95th percentile was 1.4. This finding further supports the hypothesis that our combination of five cardiac TF families is associated with an unusually high odds of differential gene expression.

Further examination of all the results in Table 5 suggested that any TF combination containing both MEF2 and NFAT tended to fall among the higher ORs. For example, four

American Heart Association Learn and Live
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

out of five of the highest ORs for either Tmax cutpoint contained both MEF2 and NFAT. Figure 3 explores this finding further by plotting the ORs from both models and highlighting all TF combinations that contain both MEF2 and NFAT. As shown, any TF combination containing both MEF2 and NFAT tended to consistently rank among the highest ORs in our study. We note that the findings in Table 5 and Figure 3 were very similar when we included a term in the model that adjusted for the total number of PWMS associated with each gene (results not shown). Thus, among the candidate TF combinations investigated, our analysis suggests that genes jointly targeted by MEF2 and NFAT show the strongest evidence for differential expression in the failing human heart.

In the Supplementary Data, we provide a browseable database of the human cardiac MEF2/NFAT target genes and promoters identified by our analysis. We also provide a heatmap that summarizes changes in expression of these genes in across samples, and an ontology analysis of biological processes to which they contribute.

Circulation Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Discussion

In this study we integrated cardiac gene expression data, promoter sequence data, and TF binding motif data to ask whether candidate hypertrophic TFs (MEF2, NFAT, NKX, GATA, and FOX) co-regulate gene expression in the failing human heart. We found that (1) binding motifs for these TFs tend to cluster in a subset of cardiac promoters, with strong evidence of higher-order clusters; (2) the greater the number the of binding motifs, the greater the odds of differential expression in heart failure; and (3) genes containing binding motifs for both MEF2 and NFAT show the strongest likelihood of differential expression among the combinations studied. Taken together, these findings suggest that regulation of myocardial gene expression by combinatorial action of MEF2 and NFAT may play an important role in human heart failure.

Numerous animal studies suggest that cardiac hypertrophy and heart failure are mediated to a large extent by cardiac TFs integrating upstream stress signals and acting in a combinatorial fashion to regulate the expression of "fetal" genes in cardiac myocytes[14-16]. This paradigm has developed largely based on evidence from murine models and cell culture through the use of transgenic techniques that explore the function of specific TFs. However, these techniques cannot be applied to human subjects and, as such, it has been difficult to investigate the overall paradigm, and the specific TFs identified, in human disease.
Our study provides the first evidence supporting combinatorial regulation of myocardial gene expression in the human heart. More broadly, we show that integrating diverse sources of whole genome data can provide insights into the pathogenesis of human cardiovascular disorders.

Our finding regarding MEF2 and NFAT is of particular interest since these factors play a critical role in the pathogenesis of heart failure in murine models. MEF2 (myocyte

American Heart Association Learn and Live®
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

enhancer factor 2) TFs are essential regulators of muscle gene expression in cardiac development and stress-induced cardiac hypertrophy. There are four MEF2 TFs (MEF2A-D), which share overlapping functions and interact with the same promoter-region binding site. MEF2A$^{-/-}$ and MEF2C$^{-/-}$ mice die from a variety of cardiac abnormalities[17, 18], and Kim et al recently demonstrated that MEF2D$^{-/-}$ mice show a blunted hypertrophic response to cardiac stress[19]. Conversely, transgenic over-expression of MEF2A or MEF2C enhances the hypertrophic response to stress[20], and *in vivo* reporter-gene studies demonstrate endogenous MEF2 activation in response to hypertrophic stimuli[21]. These and other studies have established MEF2 TFs as critical mediators of hypertrophy and heart failure in mice.

Likewise, NFAT TFs are critical for cardiac development and are well-established mediators of stress-induced hypertrophy and failure. The four NFAT TFs (NFATc1-4) are expressed in the myocardium and localized to the cytoplasm. In response to increases in intracellular calcium, the phosphatase calcineurin dephosphorylates NFAT TFs, which are then translocated to the nucleus where they alter gene expression and induce hypertrophy and failure in animals[14]. Interestingly, previous *in vitro* studies also found that MEF2 and NFAT can cooperatively regulate gene expression. In T-cells, stimulation of the T-cell antigen receptor (TCR) induces expression of Nu77 via MEF2/NFAT interactions that are mediated by physical association between MEF2, NFAT, and p300, a histone acetyl transferase that enhances transcription via chromatin unfolding[22]. NFAT-MEF2 interactions have also been identified in the development of slow-twitch skeletal muscle fibers and expression of myoglobin[23, 24]. Thus, our findings in human heart failure are consistent with experimental data from a variety of different model systems, and extend the paradigm of co-regulation by MEF2/NFAT to the failing human heart.

Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Although the principal aim of this study was to explore evidence for co-regulation by cardiac TFs, we have also identified a number of specific target genes that show altered expression in the failing heart and contain binding sites for cardiac TFs in their 1kb promoter region (see Supplementary Data for a browseable database of target genes). These genes may be novel contributors to the pathophysiology of heart failure and warrant focused study. An analysis of ontological terms associated with these genes showed enrichment of terms related to muscle and vascular tissue as well as organ development and morphogenesis (Supplementary Data). This is consistent with the concept that cardiac remodeling in humans represents, in part, a reactivation of fetal/developmental pathways. Many of these specific target genes encode components of the sarcomere (e.g., MYOT, NEBL, ITGB1BP2), the cytoskeleton (e.g., ADD3, PALLD, SORBS2), and the interstitial matrix (e.g., LUM, COL3A1,TNXB, OGN, COL16A1). However, a number of transcription factors and co-regulators were also identified as MEF2 / NFAT targets (e.g., HMGN3, E2F, MAX, SNW1, TCF12, SMAD7, GTF2E1, KLF6 and ATF6). Of note, MEF2C and NFATC1 were, themselves, identified as target genes that contain upstream NFAT and MEF2 binding sites and that were induced in the failing heart. These findings are consistent with a model in which hypertrophic TFs induce the expression of a network of secondary TFs to mediate their effects, and offer the intriguing possibility that feedback loops may exist between MEF2/NFAT activity and transcription of the MEF2C and NFATC1 genes.

American Heart Association Learn and Live® Circulation Cardiovascular Genetics Journal of the American Heart Association

Although this is a human study performed with rigorous approaches, several limitations warrant discussion. First, this is an observational study that associates groups of TFs with heart failure, but cannot prove a causal relationship. Such proof requires experimental manipulation of myocardial TF activity in the human heart, which is currently not feasible. Second, although the total sample size is large for a gene expression study, our

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

control group is relatively small owing to the difficulty in acquiring non-failing human cardiac tissue. In addition, the study population had advanced stage heart failure, and findings might differ in earlier stages of disease. Third, to build on results from animal studies, we selected five candidate TF families based on the current state of knowledge and on prior analyses[3]. We may have ignored the potential contributions of other TFs, and ongoing work using an inclusive exploratory approach across the entire TRANSFAC database should help resolve this question.

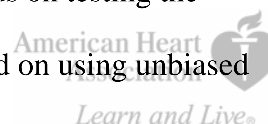
Methodological challenges arose during the course of our analysis. As discussed in our prior work[3], studying TF binding motifs prevents us from distinguishing the effects of individual members of a TF family, since all members recognize highly similar binding motifs. This could bias our analysis toward finding no associations since we group TFs that may not contribute to heart failure (e.g., GATA1) along with those that probably do (e.g., GATA4). Thus, our results may underestimate the extent of co-regulation by cardiac TFs. While we used logistic regression to model differential expression as a binary outcome, a proportion of the genes in our study were undoubtedly misclassified (Table 3). Our previous study used a different method to assess differential expression and a more stringent threshold that yielded a smaller group of differentially expressed genes[3]. Despite these differences, the results of the current and original studies were consistent in identifying significant associations between the five individual TFs and differential gene expression, and the current study extends these findings by identifying groups of TFs that appear to co-regulate gene expression. Another challenge was the natural variation in the frequency of the individual and combinations of TFs and the implications for statistical power. Thus if two TF combinations both have the same association with differential expression, but one combination occurs more frequently, the more frequent combination will tend to yield the

American Heart Association Learn and Live
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

lower p-value in a hypothesis test. This emphasizes the need to focus on both the effect size (odds ratio) and the level of statistical significance (p-value) in the interpretation of our results. In the case of combinations containing MEF2 and NFAT, both the odd ratios (Figure 3) and p-values (Table 5) indicated strong associations.

In conclusion, we provide genomic evidence for co-regulation of myocardial gene expression by MEF2 and NFAT in advanced human heart failure. In doing so, this work extends the paradigm of combinatorial regulation of gene expression to the human heart and has identified target-genes of biologic interest. Future work will focus on testing the mechanistic role of these target genes using experimental models, and on using unbiased approaches to broadly screen combinations of all known transcription factors for association with human heart failure.



Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Sources of Funding

Supported by research grants from the National Institutes of Health (5R01HL088577, 1R21HL092379, 5U01AI063589), and from the Penn Cardiovascular Institute.

Conflict of Interest Disclosures

None.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

References

1. Hill JA, Olson EN. Cardiac plasticity. *N Engl J Med.* 2008; 358:1370-80.

2. Olson EN. A decade of discoveries in cardiac biology. *Nat Med.* 2004; 10:467-74.

3. 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.

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

4. Olson EN. Gene regulatory networks in the evolution and development of the heart. *Science.* 2006; 313:1922-7.

5. 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.

6. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. *R 2 6* 2008;Available at: URL: www.r-project.org.

7. 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. *Biostat.* 2003; 4:249-64.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

8. Horwitz PA, Tsai EJ, Putt ME, Gilmore JM, Lepore JJ, Parmacek MS, Kao AC, Desai SS, Goldberg LR, Brozena SC, Jessup ML, Epstein JA, Cappola TP. Detection of cardiac allograft rejection and response to immunosuppressive therapy with peripheral blood gene expression. *Circulation.* 2004; 110:3815-21.

9. Levy S, Hannenhalli S. Identification of transcription factor binding sites in the human genome sequence. *Mamm Genome.* 2002; 13:510-4.

10. Cui X, Hwang JTG, Qui J, Blades NJ, Churchill GA. Improved statistical tests for differential gene expression by shringkin variance components estimates. *Biostat.* 2005; 6:59-75.

11. Ge Y, Dudoit S, Speed T. Resampling-based multiple testing for microarray data analysis. *Test.* 2003; 12:1-44.

12. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics.* 2003; 31:2013-35.

13. Barry WT, Nobel AB, Wright FA. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics.* 2005; 21:1943-9.

14. Molkentin JD, Lu JR, Antos CL, Markham B, Richardson J, Robbins J, Grant SR, Olson EN. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell.* 1998; 93:215-28.

American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

15. Belaguli NS, Sepulveda JL, Nigam V, Charron F, Nemer M, Schwartz RJ. Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators. *Mol Cell Biol.* 2000; 20:7550-8.

16. Sepulveda JL, Vlahopoulos S, Iyer D, Belaguli N, Schwartz RJ. Combinatorial expression of GATA4, Nkx2-5, and serum response factor directs early cardiac gene activity. *J Biol Chem.* 2002; 277:25775-82.

17. Naya FJ, Black BL, Wu H, Bassel-Duby R, Richardson JA, Hill JA, Olson EN. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med.* 2002; 8:1303-9.

American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
Journal of the American Heart Association

18. Lin Q, Schwarz J, Bucana C, Olson EN. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science.* 1997; 276:1404-7.

19. Kim Y, Phan D, van RE, Wang DZ, McAnally J, Qi X, Richardson JA, Hill JA, Bassel-Duby R, Olson EN. The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J Clin Invest.* 2008; 118:124-32.

20. Xu J, Gong NL, Bodi I, Aronow BJ, Backx PH, Molkentin JD. Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. *J Biol Chem.* 2006; 281:9152-62.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

21. Passier R, Zeng H, Frey N, Naya FJ, Nicol RL, McKinsey TA, Overbeek P, Richardson JA, Grant SR, Olson EN. CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. *J Clin Invest.* 2000; 105:1395-406.

22. Youn HD, Chatila TA, Liu JO. Integration of calcineurin and MEF2 signals by the coactivator p300 during T-cell apoptosis. *EMBO J.* 2000; 19:4323-31.

23. Wu H, Naya FJ, McKinsey TA, Mercer B, Shelton JM, Chin ER, Simard AR, Michel RN, Bassel-Duby R, Olson EN, Williams RS. MEF2 responds to multiple calcium-regulated signals in the control of skeletal muscle fiber type. *EMBO J.* 2000; 19:1963-73.

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

24. Chin ER, Olson EN, Richardson JA, Yang Q, Humphries C, Shelton JM, Wu H, Zhu W, Bassel-Duby R, Williams RS. A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev.* 1998; 12:2499-509.

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Table 1. Characteristics of specific TF families

	TF Family					
	MEF2	**GATA**	**NKX**	**NFAT**	**FOX**	**ANY**
Number of PWMs in TRANSFAC	10	23	8	2	27	70
Number (percent) of cardiac genes matching at least 1 PWM	1727 (19%)	5602 (62%)	3340 (37%)	1220 (13%)	3582 (40%)	7872 (87%)
Median (range) number of PWM matches per cardiac gene	1(1,10)	2(1,15)	1(1,5)	1(1,2)	1(1,18)	3(1,32)

Table 2. Characteristics of cardiac genes with specific numbers of associated TF families.

	Number of TF Families					
	0	**1**	**2**	**3**	**4**	**5**
Number (percent) of cardiac genes	1204 (13%)	2961 (33%)	2836 (31%)	1481 (16%)	515 (6%)	67 (0.7%)
Median (range) number of PWM matches per cardiac gene	0 (0,0)	1 (1,17)	3 (2,22)	6 (3,28)	9 (4,32)	11 (5,24)

Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Table 3. Characteristics of gene groups defined by thresholds of Tmax

	Threshold for differential expression	
	Tmax>2.0	**Tmax>3.2**
Number of genes (% of total) classified as differentially expressed	3579 (39%)	1772 (20%)
Number of genes (% of total) classified as unchanged	5497 (61%)	7304 (80%)
Proportion of genes misclassified as differentially expressed (FDR)*	9%	1%
Proportion of genes misclassified as unchanged*	18%	34%
Overall misclassification rate*	14%	28%

*Based on subject-level permutations, which estimated a total of 4247 genes (47%) as differentially expressed out of a total of 9076.

American Heart Association
Learn and Live®
Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Table 4. Odds of differential expression associated with individual TF families and numbers of TF families

		OR (p-values)	
		Tmax>2	**Tmax>3.16**
TF Family*	**MEF2**	**1.26 (.010)**	**1.47 (<.0002)**
	NFAT	1.15 (.066)	**1.31 (.001)**
	GATA	1.13 (.073)	**1.24 (.009)**
	FOX	1.15 (.067)	**1.30 (.004)**
	NKX	**1.18 (.024)**	**1.29 (.001)**
Number of Families†	0	1.00	1.00
	1	**1.15 (.041)**	**1.23 (.004)**
	2	1.12 (.121)	**1.23 (.005)**
	3	1.11 (.237)	**1.24 (.008)**
	4	**1.32 (.033)**	**1.47 (.0008)**
	5	**2.14 (.013)**	**3.19 (.0002)**
	Global p	.035	.0006

P-values < 0.05 are in bold; row entries in bold indicate p<.05 for both Tmax cutpoints

* OR for genes with specific TF Family member versus no TF family members from any family. Each row represents a separate model.

† OR for genes with each increase in TF family number versus no TF family members. A single model was constructed.

Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Table 5. Odds of differential expression associated with specific TF combinations.

Number of TF families	TF combination (number of genes)	OR (p-value)	
		Tmax>2.00	Tmax>3.16
0	none (1216)	1.00	1.00
1	**MEF2 (178)**	**1.38 (.044)**	**1.56 (.0046)**
	NKX (511)	**1.33 (.007)**	**1.34 (.003)**
	GATA (1576)	1.10 (.205)	**1.18 (.0250)**
	FOX (578)	1.09 (.418)	**1.20 (.0350)**
	NFAT (118)	1.07 (.737)	1.22 (.1222)
2	**MEF2.NFAT (24)**	**3.47 (.005)**	**3.03 (.011)**
	MEF2.NKX (86)	1.31 (.199)	**1.84 (.004)**
	GATA.NFAT (212)	1.30 (.082)	1.17 (.134)
	MEF2.GATA (279)	1.25 (.113)	**1.41 (.001)**
	NKX.FOX (296)	1.13 (.351)	**1.28 (.027)**
	MEF2.FOX (142)	1.13 (.502)	1.13 (.152)
	NKX.NFAT (62)	1.09 (.653)	1.12 (.109)
	GATA.NKX (837)	1.07 (.443)	1.14 (.090)
	GATA.FOX (836)	1.03 (.745)	**1.17 (.046)**
	NFAT.FOX (62)	1.02 (.822)	1.47 (.055)
3	MEF2.NKX.NFAT (17)	2.48 (.068)	**2.75 (.029)**
	MEF2.GATA.NFAT (39)	1.65 (.073)	1.51 (.059)
	GATA.NKX.FOX (569)	**1.31 (.011)**	**1.47 (<.0002)**
	MEF2.NFAT.FOX (21)	1.30 (.300)	**2.02 (.032)**
	MEF2.GATA.FOX (243)	1.13 (.455)	**1.51 (.003)**
	MEF2.NKX.FOX (82)	1.05 (.764)	0.78 (.100)
	GATA.NFAT.FOX (146)	0.99 (.930)	0.94 (.201)
	MEF2.GATA.NKX (164)	0.95 (.756)	1.00 (.259)
	NKX.NFAT.FOX (57)	0.72 (.294)	**0.48 (.0006)**
	GATA.NKX.NFAT (143)	**0.70 (.047)**	0.77 (.093)
4	GATA.NKX.NFAT.FOX (130)	1.73 (.097)	1.39 (.066)
	MEF2.NKX.NFAT.FOX (24)	1.39 (.107)	1.68 (**.048**)
	MEF2.GATA.NFAT.FOX (66)	1.36 (.121)	1.75 (**.028**)
	MEF2.GATA.NKX.FOX (263)	1.27 (.211)	1.43 (**.011**)
	MEF2.GATA.NKX.NFAT (32)	1.19 (.461)	1.41 (.059)
5	**MEF2.GATA.NKX.NFAT.FOX (67)**	**2.13 (.013)**	**3.20 (.0002)**
	Global p	.028	.001

Within each TF family number specific TF combinations are ordered by OR estimate for Tmax>2.
P-values < 0.05 are in bold; row entries in bold indicate p<.05 for both Tmax cutpoints

American Heart Association Learn and Live
Circulation Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Figure Legends

Figure 1. Overall design.

Figure 2. Observed versus expected frequency of TF combinations in promoters of N=9076 cardiac genes. The expected frequency for a specific TF combination was determined by assuming that the occurrence of each TF family in the combination was distributed randomly across all genes. **A,** Ratio of observed to expected frequency of genes with specific TF families as a function of the overall number of TF families. **B,** Observed (black) vs. expected (gray) frequencies of specific TF combinations in promoters of cardiac genes. Headings indicate the overall number of TF families. Bonferroni-corrected p-values for the exact binomial test comparing observed versus expected frequencies appear below each combination; P-values >.05 are indicated by a dashed line.

American Heart Association Learn and Live
Circulation Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Figure 3. Comparison of odds ratios for differential expression using two different Tmax cutpoints across all TF combinations. TF combinations that include both MEF2 and NFAT (black) show the highest odds of differential expression.

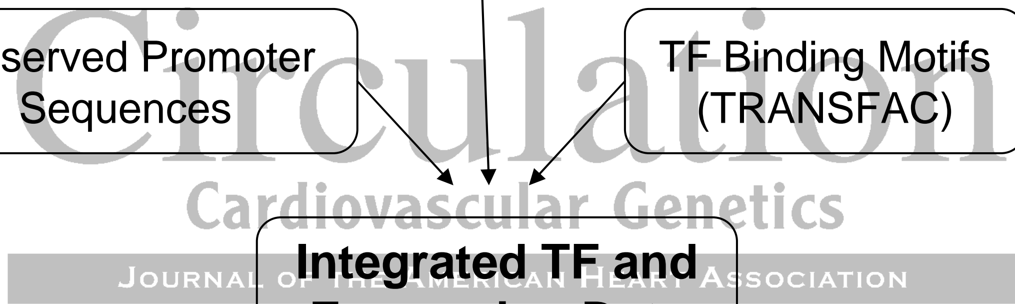
Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009

Whole Genome Expression Data (Heart Failure/Controls)

American Heart Association
Learn and Live

Conserved Promoter Sequences

TF Binding Motifs (TRANSFAC)


Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Integrated TF and Expression Data

Model differential expression as a function of TF combinations (MEF2, NFAT, NKX, GATA, FOX)

Observed:Expected

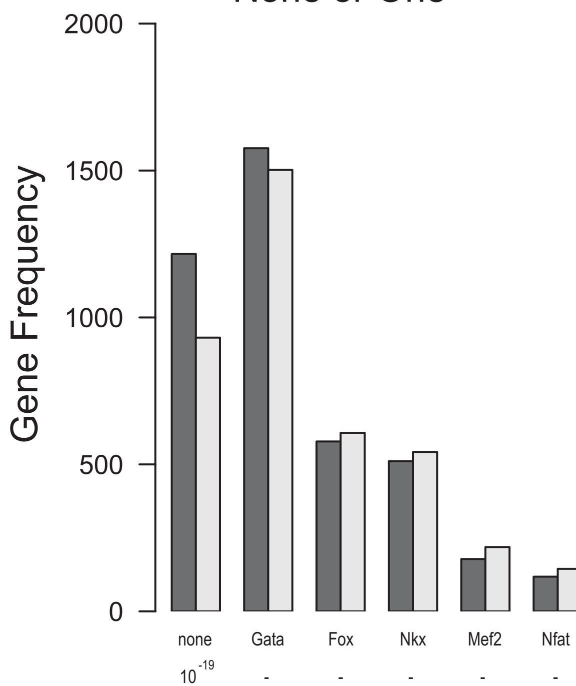
3.0
2.5
2.0
1.5
1.0

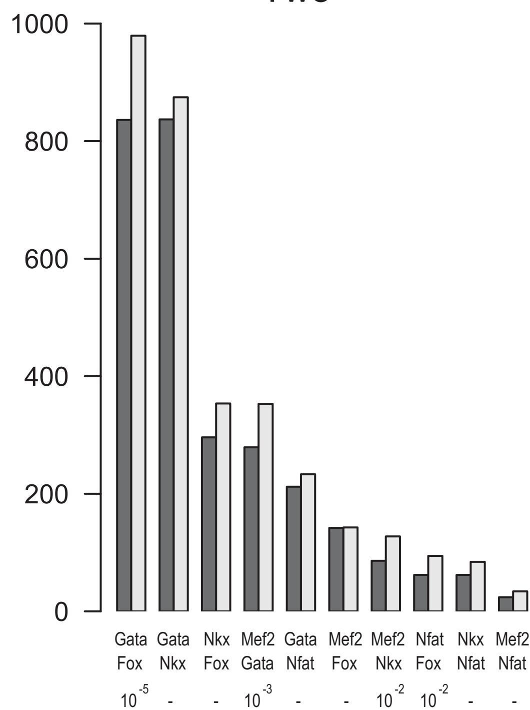
0 1 2 3 4 5

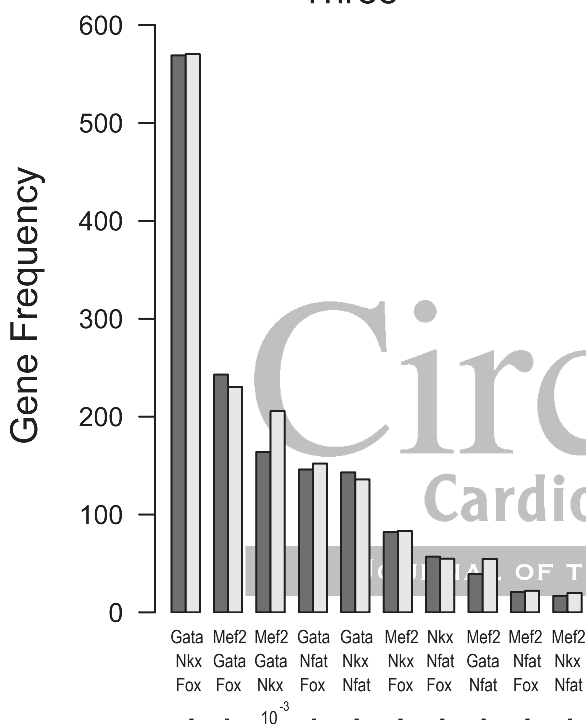
Number of TF families

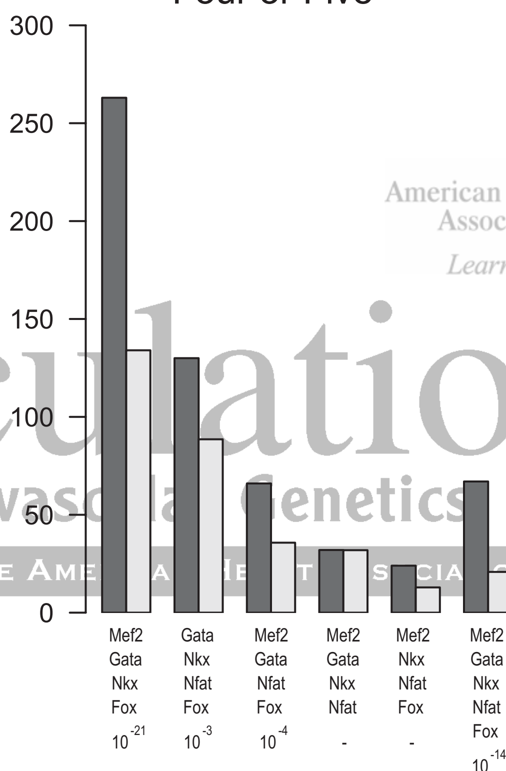
Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 20



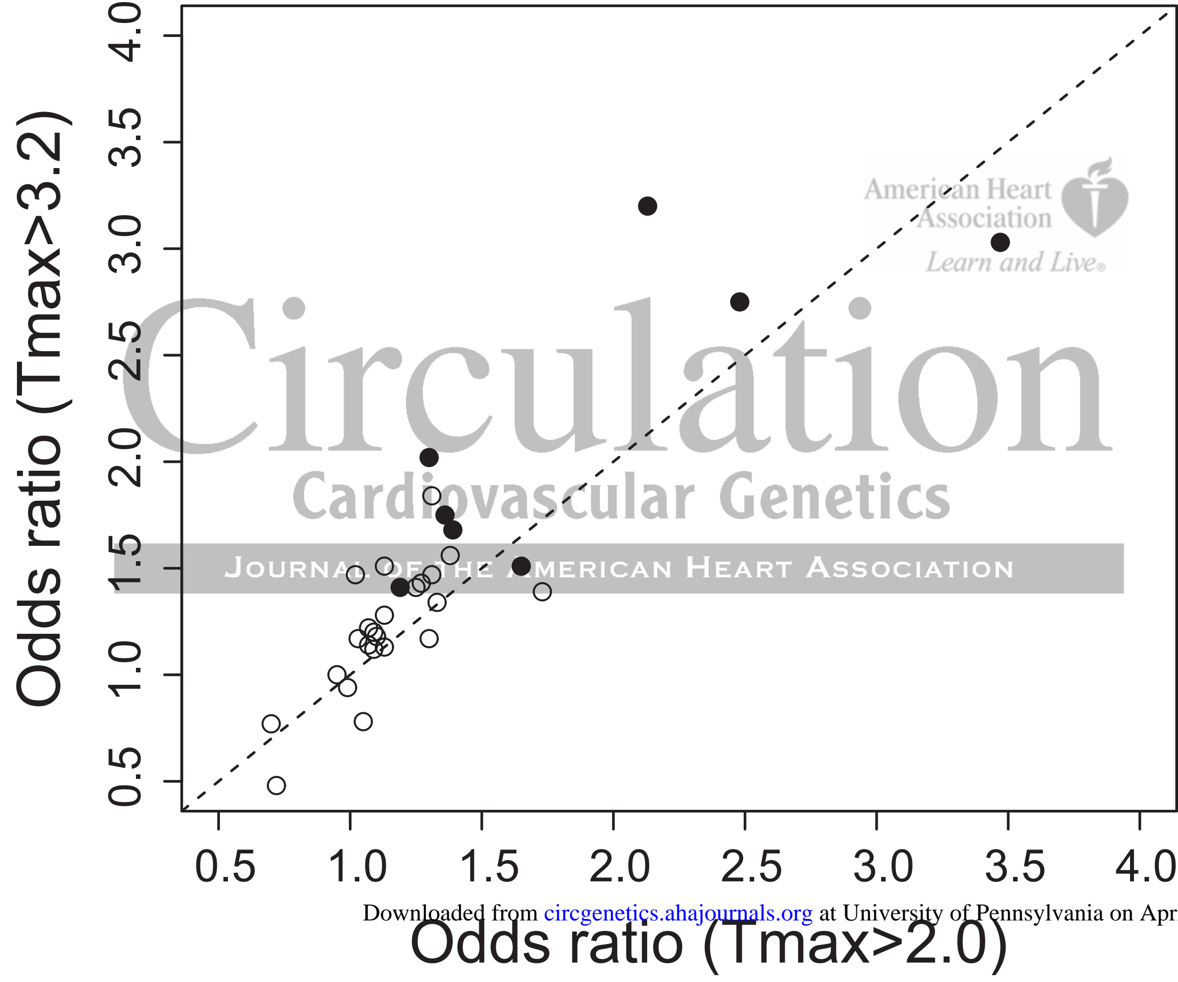






American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
Journal of the American Heart Association

Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on April 1, 2009



American Heart Association
Learn and Live
Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION
Downloaded from circgenetics.ahajournals.org at University of Pennsylvania on Apr