Re-analysis of expression profiles for revealing new potential candidate genes of heart failure

H.B. CHEN, L. WANG¹, J.F. JIANG

Department of Medical Cardiology, Tongji Hospital, Affiliated to Tongji University, Shanghai, China
¹Department of Cardiovascular Surgery, the General Hospital of Ningxia Medical University, Yinchuan, China

Abstract. – BACKGROUND: Chronic heart failure (HF) is characterized by diminished cardiac output and pooling of blood in the venous system.

AIM: We used the GSE1145 microarray data to identify potential genes that related to heart failure to construct a regulation network.

MATERIALS AND METHODS: In the network, some of transcription factors (TFs) and target genes have been proved to be related to heart failure in previous study. The gene MYC, RELA, HIF1A, NFκB1 and SMAD3 are as hub nodes in our transcriptional network and have a close relationship with heart failure.

RESULTS: The study did not address regulation network but researched for the significant pathways related to chronic heart failure. Besides, RELA and NFκB1 interfere with each other in response to HF.

CONCLUSIONS: It is demonstrated that transcriptional network analysis is useful in identification of the candidate genes in heart failure.

Key Words:
Heart failure, Regulation network, Microarray data, Pathway.

Introduction

Chronic heart failure (HF) is characterized by diminished cardiac output and pooling of blood in the venous system. Patients with HF suffer from a range of symptoms, including lack of energy, breathlessness, dry mouth, drowsiness, numbness or tingling in the hands and feet, insomnia, cough, anorexia, anxiety and depression¹. Chronic HF always results from myocardial ischemia, myocardial infarction, hypertension, dilated cardiomyopathy and so on.

Recently, the mechanism of heart failure mainly summarizes as following aspects. A variety of neurohormonal proteins, including norepinephrine, angiotensin II, endothelin, aldosterone, and tumor necrosis factor (TNF), have been implicated as some of the potentially biologically active molecules to contribute to disease progression in the failing heart. Many oxidant signaling was increased in heart failure, such as superoxide, hydrogen peroxide, hydroxyl radical, NADPH. Furthermore, TNF-α exerts strong direct effects on cardiomyocytes, as it induces apoptosis, depression of contractility and down-regulation of sarcomeric proteins in cardiomyocytes. Cardioprotection mechanism was mediated by IL-6-glycoprotein (gp130)-STAT3 (signal transducer and activator of transcription 3) signaling²³.

Dilated cardiomyopathy (DCM) is characterized primarily by left ventricular dilatation and impaired systolic function and is one of the leading causes of heart failure with high morbidity and mortality. In the absence of significant coronary artery disease or other discernible causes, its origin is considered unknown and termed idiopathic DCM (IDC).

Pathogenetic mechanisms in DCM mainly include genetic etiology, viral etiology and autoimmunity. Reduction of the sarcomere or the cellular skeleton protein (such as α-myosin heavy chain and β-actin), is the more direct pathogenetic mechanism⁴⁵. Other relevant proteins in cardiovascular regulatory systems (β-adrenergic receptors, Renin-Angiotensin system, Endothelin system, etc.) have also been associated with the disease severity⁶. Furthermore, proteins associated with remodeling of the extracellular matrix are also abundantly expressed, such as collagen proteins, cytokine gene (TNF-α, IL-1β, and IL-6) in DCM⁷. Viral etiology and autoimmunity represent alternative pathogenetic mechanisms in DCM. A virus causes direct myocardial insult, elicits an autoimmune response which in turn promotes progressive myocardial damage, ventricular dilation and heart failure. The hypothesis of an enteroviral, adenoviral infection as an initial pathogenic cause in DCM is supported by studies demonstrating the presence of enteroviral RNA or adenoviral DNA in the myocardial tissue⁸. Other observations stress the importance of the immunological mechanisms. Increasing of T lymphocytes expression in the circulation could indicate an ongoing immune response,
which leads to HLA-DR and CD40L expression and contributes to myocardial dysfunction in DCM.

DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease. A high-throughput microarray experiment was designed to analyze genetic expression patterns and identify potential genes to target for heart failure. Genomic expression profiling evolves as a useful tool to identify novel pathogenic mechanisms in human cardiac disorders.

The purpose of this paper is to propose the hypothesis that a transcriptional network can be developed by a set of transcription factors. The differently expressed genes induced by DCM can be identified and modulated in response to heart failure. Further analysis indicated that the genes and pathways in the network are related to the heart failure. The study did not address regulation network but researched for the significant pathways related to chronic heart failure.

Materials and Methods

Data Sources

Affymetrix Microarray Data

One transcription profiles of heart failure GSE1145 were obtained from a public functional genomics data repository Gene Expression Omnibus (GEO) which are based on the AffymetrixGPL570 platform data (Affymetrix Human Genome U133A Array). There are totally 26 chips in our study, 15 chips are collected from heart failure patients and 11 chips are for control.

Pathway Data

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals. The PATHWAY database records networks of molecular interactions in the human genome that contain DNA-binding domains, and most of these are presumed to function as transcription factors (TFA). The combinatorial use of a subset of the approximately 2000 human transcription factors easily accounts for the unique regulation of each gene in the human genome during development. These TFs are grouped into 5 super class families based on the presence of conserved DNA-binding domains. The Transcription Factor (TRANSFAC) database contains data on transcription factors, their experimentally-proven binding sites, and regulated genes.

Regulation Data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as transcription factors (TFA). The combinatorial use of a subset of the approximately 2000 human transcription factors easily accounts for the unique regulation of each gene in the human genome during development. These TFs are grouped into 5 super class families based on the presence of conserved DNA-binding domains. The Transcription Factor (TRANSFAC) database contains data on transcription factors, their experimentally-proven binding sites, and regulated genes.

Transcriptional Regulatory Element Database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans-regulatory elements in mammals. TRED annotate for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The annotation is currently focusing on target genes of 36 cancer-related TF families.

774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target genes were collected from TRANSFAC (http://www.gene-regulation.com/pub/databases.html). 5722 pairs of regulatory relationship between 102 TFs and 2920 target genes were collected from TRED (http://rulai.cshl.edu/TRED/).

Combined with the two regulation datasets, total 6328 regulatory relationships between 276TFs and 3002 target genes were collected (Table I).

Methods

Differentially Expressed Genes (DEGs) Analysis

For the GSE1145 dataset, the limma method was used to identify DEGs. The original expres-

Table I. Regulation datasets form TRANSFAC and TRED.

<table>
<thead>
<tr>
<th>Source</th>
<th>Regulation</th>
<th>TFs</th>
<th>Targets</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRED</td>
<td>5722</td>
<td>102</td>
<td>2920</td>
<td><a href="http://rulai.cshl.edu/TRED/">http://rulai.cshl.edu/TRED/</a></td>
</tr>
<tr>
<td>Total</td>
<td>6328</td>
<td>276</td>
<td>3002</td>
<td></td>
</tr>
</tbody>
</table>
sion datasets from all conditions were processed into expression estimates using the RMA (Robust Multichip Average) method with the default settings implemented in Bio-conductor, and then constructed the linear model. The DEGs only with the fold change value larger than 2 and p-value less than 0.05 were selected.

**Co-expression Analysis**

For demonstrating the potential regulatory relationship, the Pearson Correlation Coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The regulatory relationships whose absolute PCC are larger than 0.6 were considered as significance.

**Gene Ontology (GO) Analysis**

The BiNGO (Biological Networks Gene Ontology) analysis\(^\text{21}\) was used to identify over-represented GO categories in biological process.

**Regulation Network Construction**

Using the regulation data that have been collected from TRANSFAC database and TRED database, we matched the relationships between differentially expressed TFs and its differentially expressed target genes.

Based on the above two regulation datasets and the pathway relationships of the target genes, we built the regulation networks by Cytoscape\(^\text{22}\). Based on the significant relationships (PCC (Pearson correlation coefficient) > 0.6 or PCC < -0.6) between TFs and its target genes, 33 putative regulatory relationships were predicted between 7 TFs and 22 target genes.

**Significance Analysis of Pathway**

We not only adopted an influential analysis including the statistical significance of the set of pathway genes but also considered other crucial factors such as the magnitude of each genetic expression change, the topology of the signaling pathway, their interactions, etc\(^\text{23}\). In this model, the Impact Factor (IF) of a pathway Pi is calculated as the sum of two terms:

\[
IF (Pi) = \log \left( \frac{1}{pi} \right) + \frac{\sum_{g \in Pi} |PF(g)|}{|VE| \cdot N_{de}(Pi)}
\]

The first term is a probabilistic term that captures the significance of the given pathway Pi from the perspective of the set of genes contained in it.

It is obtained by using the hyper geometric model in which \(pi\) is the probability of obtaining at least the observed number of differentially expressed gene, Nde, just by chance\(^\text{24,25}\).

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway (i.e., its topology).

The second term sums up the absolute values of the perturbation factors (PFs) for all genes g on the given pathway Pi.

The PF of a gene g is calculated as follows:

\[
PF(g) = VE(g) + \sum_{u \in US_g} \beta_{ug} \cdot \frac{PF(u)}{N_{ds}(u)}
\]

In this equation, the first term \(\beta_{ug}\) captures the quantitative information measured in the gene expression experiment. The factor \(\beta_{ug}\) represents the normalized measured expression change of the gene g. The first term \(\beta_{ug}\) in the above equation is a sum of all PFs of the genes u directly upstream of the target gene g, normalized by the number of downstream genes of each such gene \(N_{ds}(u)\), and weighted by a factor \(\beta_{ug}\), which reflects the type of interaction: \(\beta_{ug} = 1\) for induction, \(\beta_{ug} = -1\) for repression (KEGG supplies this information about the type of interaction of two genes in the description of the pathway topology). \(US_g\) is the set of all such genes upstream of g. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, \(N_{de}(Pi)\). In order to make the IFs as independent as possible from the technology, and also comparable between problems, we also divide the second term in equation 1 by the mean absolute fold change \(E\), calculated across all differentially expressed genes. The result of the significance analysis of pathway was shown in Table II.

**Regulation Network Between TFs and Pathways**

To further investigate the regulatory relationships between TFs and pathways, we mapped DEGs to pathways and got a regulation network between TFs and pathways.
Results

Regulation Network Construction in Idiopathic Divergence

To get pathway-related DEGs of idiopathic divergence, we obtained public available microarray datasets GSE1145 from GEO. After microarray analysis, the DEGs with the fold change value larger than 2 of GSE1145 and p-value less than 0.05 were selected. 487 genes were selected as DEGs from GSE1145. To get the regulatory relationship, the co-expressed value (PCC ≥ 0.6) was chose as the threshold. Finally, we got 25 regulatory relationships between 10 TFs and their 14 differently expressed target genes. By integrating the regulatory relationships above, a regulation network of idiopathic dilated was built between TFs and its target genes (Figure 1). In this network, MYC, RELA, HIF1A, NFκB1, SMAD3 with higher degrees formed a local network which suggesting that these TFs may play an important role in idiopathic divergence. Besides, the NFκB1 and RELA not only regulated the MYC, but also the PTX3 and CCL2 were observed in our network.

GO Analysis of the Regulation Network in Idiopathic Divergence

Several Gene Ontology (GO) categories were enriched among these genes in the regulatory network, including positive regulation of nitrogen compound metabolic process and so on (Table III).

Significant Pathways in Idiopathic Divergence

To identify the relevant pathways changed in heart failure, we used a statistical approach on pathway level. Significant analysis at single gene...
level may suffer from the limited number of samples and experimental noise that can severely limit the accuracy of the chosen statistical test. The pathways can provide an alternative way to relax the significant threshold applied to single genes and may lead to a better biological interpretation. So we adopted a pathway based impact analysis method that contained many factors including the statistical significance of the set of differentially expressed genes in the pathway, the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions and so on. The impact analysis method yields many significant pathways contained Adherens junction, GnRH signaling pathway, Gap junction and so on (Table II).

### Regulation Network Between TFs and Pathways in Idiopathic Divergence

To further investigate the regulatory relationships between TFs and pathways, we mapped DEGs to pathways and got a regulation network between TFs and pathways (Figure 2). In the network, MYC, RELA, NFκB1, SMAD3 and so on were shown as hub nodes linked to lots of idiopathic dilated related pathways. ETS1 (external transcribed spacer region 1), TP53 (tumor protein p53), MYC and other TFs together regulate pathways in cancer, cell cycle, ErbB (human epidermal growth factor receptor 2) signaling pathway and colorectal cancer. Some of TFs, such as TP53, SMAD3 (small mother against decapentaplegic 3) and NME2 (nucleoside diphosphate kinase 2) may active the downstream pathways through regulate the MYC.

### Discussion

From the result of regulation network construction in heart failure, we could find that many TFs and pathways closely related to HF have been linked by our method. The gene MYC, RELA, HIF1A, NFkB1 and SMAD3 are as hub nodes in our transcriptional network and have a close relationship with heart failure. RELA and

### Table II. Pathway significant analysis.

<table>
<thead>
<tr>
<th>Database name</th>
<th>Pathway name</th>
<th>Impact factor</th>
<th>Pathway genes in input (%)</th>
<th>Corrected gamma p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG</td>
<td>Adherens junction</td>
<td>55.357</td>
<td>3.846</td>
<td>5.13E-23</td>
</tr>
<tr>
<td>KEGG</td>
<td>GnRH signaling pathway</td>
<td>34.336</td>
<td>2.913</td>
<td>4.33E-14</td>
</tr>
<tr>
<td>KEGG</td>
<td>Gap junction</td>
<td>29.104</td>
<td>3.125</td>
<td>6.90E-12</td>
</tr>
<tr>
<td>KEGG</td>
<td>Melanoma</td>
<td>20.648</td>
<td>7.042</td>
<td>2.33E-08</td>
</tr>
<tr>
<td>KEGG</td>
<td>Non-small cell lung cancer</td>
<td>17.47</td>
<td>9.259</td>
<td>4.78E-07</td>
</tr>
<tr>
<td>KEGG</td>
<td>Epithelial cell signaling in <em>Helicobacter pylori</em> infection</td>
<td>17.223</td>
<td>1.471</td>
<td>6.04E-07</td>
</tr>
<tr>
<td>KEGG</td>
<td>Focal adhesion</td>
<td>16.07</td>
<td>4.433</td>
<td>1.79E-06</td>
</tr>
<tr>
<td>KEGG</td>
<td>Prostate cancer</td>
<td>15.425</td>
<td>5.556</td>
<td>3.28E-06</td>
</tr>
<tr>
<td>KEGG</td>
<td>ErbB signaling pathway</td>
<td>15.244</td>
<td>9.195</td>
<td>3.89E-06</td>
</tr>
<tr>
<td>KEGG</td>
<td>Glioma</td>
<td>13.991</td>
<td>7.692</td>
<td>1.26E-05</td>
</tr>
</tbody>
</table>

### Table III. GO biological process analysis.

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>Description</th>
<th>Count</th>
<th>p-value</th>
<th>corr p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>51173</td>
<td>Positive regulation of nitrogen compound metabolic process</td>
<td>12</td>
<td>1.84E-10</td>
<td>1.60E-07</td>
</tr>
<tr>
<td>31328</td>
<td>Positive regulation of cellular biosynthetic process</td>
<td>12</td>
<td>3.74E-10</td>
<td>1.60E-07</td>
</tr>
<tr>
<td>9891</td>
<td>Positive regulation of biosynthetic process</td>
<td>12</td>
<td>4.45E-10</td>
<td>1.60E-07</td>
</tr>
<tr>
<td>45935</td>
<td>Positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process</td>
<td>11</td>
<td>2.51E-09</td>
<td>6.75E-07</td>
</tr>
<tr>
<td>31325</td>
<td>Positive regulation of cellular metabolic process</td>
<td>12</td>
<td>1.07E-08</td>
<td>2.29E-06</td>
</tr>
<tr>
<td>9893</td>
<td>Positive regulation of metabolic process</td>
<td>12</td>
<td>1.92E-08</td>
<td>3.44E-06</td>
</tr>
<tr>
<td>10557</td>
<td>Positive regulation of macromolecule biosynthetic process</td>
<td>10</td>
<td>5.51E-08</td>
<td>8.46E-06</td>
</tr>
<tr>
<td>43066</td>
<td>Negative regulation of apoptosis</td>
<td>8</td>
<td>1.08E-07</td>
<td>1.42E-05</td>
</tr>
<tr>
<td>43069</td>
<td>Negative regulation of programmed cell death</td>
<td>8</td>
<td>1.19E-07</td>
<td>1.42E-05</td>
</tr>
<tr>
<td>60548</td>
<td>Negative regulation of cell death</td>
<td>8</td>
<td>1.40E-07</td>
<td>1.50E-05</td>
</tr>
</tbody>
</table>
NFkB1 interfere with each other in response to IDCM. Although the role of HIF1A in heart failure has not been investigated to date, some evidence has suggested that it as a surrogate marker of hypoxia was identified to increase in IDCM.

MYC protein as a transcription factor is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. Bitransgenic mouse inducibly expressing MYC under the control of the cardiomyocyte-specific MHC (major histocompatibility complex) promoter was developed to address the causal relationship between increases in c-Myc (Myc) and cardiomyopathy. The results showed that the induction of Myc expression in cardiomyocytes led to the development of severe hypertrophic cardiomyopathy followed by ventricular dysfunction and ultimately death from congestive heart failure. Myc activation in cardiomyocytes is an important regulator of downstream pathological sequelae. Besides, c-Myc protein over-expression was seen in 4 patients (33.3%) with DCM in the experiment. Furthermore, in six patients with DCM, c-Myc protein was showed to elevate by 71% in the left ventricle, 55% in the right ventricle, 48% in the left atrium and 91% in the right atrium in DCM, respectively.

CCL2 is a cytokine protein structurally related to the CXC subfamily of cytokines and involved in immunoregulatory and inflammatory processes. It is also termed as MCP-1. The cytotoxic ac-

**Figure 2.** Regulation network of TF-pathway.
tion of leukocytes is known to be a probable cause of the cardiac myocyte damage seen in idiopathic dilated cardiomyopathy (IDC). Monocyte chemoattractant protein-1 (MCP-1) is a major signal for the accumulation and activation of leukocyte. Semi-quantitative analysis revealed that the expression of MCP-1 may contribute to the deterioration of left ventricular function39,40.

GADD45B gene is a member of a group of genes whose transcription levels are increased following stressful growth arrest conditions and treatment with DNA-damaging agents. The function of GADD45B protein is involved in the regulation of growth and apoptosis. DNA microarray profiling indicated decreased expression of TNF-α and NF-kB-induced anti-apoptotic genes GADD45B in human IDC, which may be important in the shift in the TNF-α signaling toward enhanced apoptosis31.

PTX3 belongs to the pentraxin family of proteins, but distinct from classic pentraxins (CRP and SAP) sharing no homology in the NH2-terminal part. PTX3 can be induced by IL-1, TNF, etc to involve in inflammation and innate immunity. The plasma PTX3 levels were higher in DCM patients than in healthy subjects32.

NF-κB is always bound to REL proteins to form the complex. One of them, the p50 (NFkB1)/p65 (RELA) heterodimer is the major NF-κB complex. So NF-κB usually refers to a p50-p65 (RelA) dimer. The activity of NF-κB is controlled by inhibitory protein IκB and IKK (IκB kinase). In human myocardium, NF-κB was about 2-fold addition in patients with heart failure due to dilated cardiomyopathy33,34.

Hypoxia-inducible factor-1 (HIF1) is a transcription factor found in mammalian cells cultured under reduced oxygen tension that plays an essential role in cellular and systemic homeostatic responses to hypoxia. HIF1 is a hetero-dimer composed of an alpha subunit and a beta subunit. Ultra-structural analysis showed epicardial microvascular density was reduced in IDC. New blood vessel growth requires up-regulation of several growth factors and cytokines, such as vascular endothelial growth factor (VEGF-A). HIF-1α as a surrogate marker of hypoxia was identified to increase in IDC, suggesting in turn modulates VEGF-A synthesis and secretion35.

Chagas’ disease (CHD) and IDC, this two cardiomyopathies have different aetiologies, but with similar symptoms, such as arrhythmias, inflammatory suggesting some relations between them. Anti-β1-adrenoceptor autoantibodies, first described in sera of patients with CHD, are now well documented in patients with IDC as an autoimmune response. And the same extracellular domain of the β1-adrenergic receptor was recognized by antibodies in CHD and IDC. The anti-β1-adrenoceptor response observed in patients with CHD, originating from an immune cross-reaction between the P2β ribosomal protein (RPLP2) and the epitope on the β1-adrenoceptor might be conceived as a pathological by stander effect of a genuine anti-parasite (T. cruzi) response. Several viruses have been suggested as possible causative factors by activation of the innate immunity, enhancement of the inflammatory response and inducing disruption of the cytoskeleton of myocardial cells leading to DCM. It can be hypothesized that the destruction of myocardial cells during the development of the disease liberates the β1-adrenoceptor as an autoantigen inducing an autoimmune response in IDC. So a similar origin for anti-receptor antibodies in both diseases is predicted36.

IDC is characterized by myocyte cell loss and cell proliferation, which are all associated with cell cycle. On the one hand p53 is involved in the regulation of cell cycle progression in response to DNA damage, causing the cell to delay its entry into S phase until the damage has been repaired. But importantly p53 plays a role in triggering an apoptotic response in instances in which the damage is too severe to repair. Bcl-2 (B cell lymphoma 2) protein is an anti-apoptotic protein representing a possible compensatory mechanism to p5337. The expression of proliferation cell nuclear antigen (PCNA), a nuclear protein necessary for DNA synthesis and cell cycle progression, and the expression of proliferation-associated nuclear antigen Ki-67 antigen have been demonstrated in stressed myocytes (severe ventricular dysfunction and failure). In IDC, cell cycle may be initiated (increased expression of PCNA (proliferating cell nuclear antigen) and Ki-67 antigen) which a negative correlation with apoptosis38.

Conclusions

GSH, the major non proteinthiol in living cells, plays an important role as an antioxidant. It is pivotal in various protective systems such as glutathione peroxidase, glutathione transferase, and free radical reductase. Studies showed whole-blood GSH concentrations of IDC patients
were decreased as compared with the control group. This result may be a consequence of increased GSH depletion due to increased oxidative stress in the IDC patients.

Conflict of Interest
None.

References


Re-analysis of expression profiles for revealing new potential candidate genes of heart failure


