

RESEARCH ARTICLE

Open Access

Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy

Chen-Ching Lin^{1,2}, Jen-Tsung Hsiang³, Chia-Yi Wu⁴, Yen-Jen Oyang², Hsueh-Fen Juan^{2,4*}, Hsuan-Cheng Huang^{1*}

Abstract

Background: Molecular networks represent the backbone of molecular activity within cells and provide opportunities for understanding the mechanism of diseases. While protein-protein interaction data constitute static network maps, integration of condition-specific co-expression information provides clues to the dynamic features of these networks. Dilated cardiomyopathy is a leading cause of heart failure. Although previous studies have identified putative biomarkers or therapeutic targets for heart failure, the underlying molecular mechanism of dilated cardiomyopathy remains unclear.

Results: We developed a network-based comparative analysis approach that integrates protein-protein interactions with gene expression profiles and biological function annotations to reveal dynamic functional modules under different biological states. We found that hub proteins in condition-specific co-expressed protein interaction networks tended to be differentially expressed between biological states. Applying this method to a cohort of heart failure patients, we identified two functional modules that significantly emerged from the interaction networks. The dynamics of these modules between normal and disease states further suggest a potential molecular model of dilated cardiomyopathy.

Conclusions: We propose a novel framework to analyze the interaction networks in different biological states. It successfully reveals network modules closely related to heart failure; more importantly, these network dynamics provide new insights into the cause of dilated cardiomyopathy. The revealed molecular modules might be used as potential drug targets and provide new directions for heart failure therapy.

Background

Protein-protein interactions (PPI) are of central importance for most biological processes, and thus the protein interaction network (PIN) provides a global picture of cellular mechanisms. With the accumulation of interactome and transcriptome data, the integration of gene expression profiles has revealed the dynamics of protein interaction networks. For example, *Han et al.* [1] analyzed the protein interaction network of yeast and uncovered two types of hub proteins: “party” hubs and “date” hubs, which displayed condition- or location-

specific features in the interactome network. *Xue et al.* [2] developed a new analytic method to discover the dynamic modular structure of the human protein interaction network in their aging study. Recently, *Taylor et al.* [3] also proposed another two types of hub proteins: intermodular hubs and intramodular hubs, and identified whether the interactions between proteins were context specific or constitutive in the human protein interaction network. Similar techniques were also applied to reveal disease related genes or modules. *Chuang et al.* [4] improved the prognostic predictive performance of gene expression signatures by incorporating interactome data in breast cancer. *Taylor et al.* [3] used a method analogous to a previous study [1] and revealed that dynamic modularity of the human protein interaction network may be a good indicator of breast cancer prognosis. In the context of heart failure,

* Correspondence: yukijuan@ntu.edu.tw; hsuancheng@ym.edu.tw

¹Institute of Biomedical Informatics, Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei 112, Taiwan

²Graduate Institute of Biomedical Electronics and Bioinformatics, Center for Systems Biology and Bioinformatics, National Taiwan University, Taipei 106, Taiwan

Full list of author information is available at the end of the article

Camargo and Azuaje [5,6] integrated gene expression profiles with the protein interaction network in human dilated cardiomyopathy and efficiently identified potential novel DCM signature genes and drug targets. These studies suggest that the integration of interactome data with transcriptome information may facilitate the identification or discovery of disease biomarkers.

Heart failure is one of the main causes of death in the world and is the consequence of many complex factors including genetics, diet, environment and lifestyle. Heart failure is a physiological state in which the heart cannot provide sufficient output of blood to meet the body's needs. Dilated cardiomyopathy (DCM), the major cause of heart failure, impairs the blood pumping ability of the heart and leads to insufficient blood flow to vital organs due to the enlargement and weakening of the heart [7]. Previous studies [8-10] on gene expression profiles have provided distinct perspectives on the etiology of DCM. Barth et al. [8] pointed out the significant immune response processes involved in end-stage DCM and presented a robust gene expression signature of this disease. Wittchen et al. [10] suggested novel therapeutic targets by gene expression profile analysis of human inflammatory cardiomyopathy. Kabb et al. [9] analyzed microarray dataset of human myocardial tissue to obtain region- and DCM-specific transcription profiles and determined the gene expression fingerprint of DCM. Even though various causes of DCM have been revealed, the underlying molecular mechanism of this disease remains unclear.

Here, we developed a network-based analysis approach to discover DCM or non-DCM related functional subnetworks by integrating DCM related gene expression profiles with the human protein interaction network and Gene Ontology (GO) annotations. A comparative analysis was utilized to extract DCM exclusive subnetworks as heart failure related modules. These modules could be used to classify normal and disease samples. We further investigated the co-expressed protein interaction network structures of each module for DCM and non-DCM and observed dynamic variations of the identified modules between the two states. Our results suggest that the modular changes between DCM and non-DCM could imply plausible molecular mechanisms involved in heart failure progression.

Results

Condition-Specific Co-expressed Protein Interaction Networks

We integrated gene expression profiles with the protein interaction network to construct condition-specific co-expressed protein interaction networks of DCM and non-DCM, respectively. Two interacting proteins are called co-expressed protein-protein interaction (CePPI)

Table 1 Structural information of DCM and non-DCM networks

| | Proteins | PPIs | Co-expressed PPIs |
|---------|----------|-------|-------------------|
| DCM | 3160 | 13928 | 4306 |
| non-DCM | 3302 | 14557 | 4872 |
| Overlap | 2226 | 11092 | 1425 |

under a specified condition, if their encoding genes are expressed correlatively with each other under particular biological conditions, such as DCM or non-DCM in this study. The network that consists of all the CePPIs under the DCM condition is abbreviated as DCM CePIN and the network under the non-DCM condition as non-DCM CePIN. Table 1 summarizes the structural information of DCM and non-DCM networks. Given that the gene expression profiles were derived from the same tissue [8], DCM and non-DCM CePINs shared almost 70% of proteins and nearly 80% of physical protein interactions in common. However, the shared CePPIs were only approximately 30%, indicating a large amount of interaction rewiring between DCM and non-DCM. Hence, CePPIs may reveal the condition-specific dynamic information hidden among otherwise common static interactions. Previous studies [1,11] have shown that by integrating co-expression information into physical protein interaction networks, dynamic features among various networks can be extracted. This preliminary observation of the network structure indicates the limitation of a static protein interaction network, and demonstrates the necessity of applying a condition-specific co-expressed protein interaction network model to discover the dynamic features which are attributed to the changes in gene expression profiles inside the static biological networks.

Hub Proteins in CePINs Tend to Be Differentially Expressed

To further investigate the features of condition-specific CePINs, we examined their topological properties, including degree, betweenness centrality, closeness centrality and clustering coefficient for significantly differentially expressed genes (SDEGs). All four topological properties of SDEGs were significantly higher than those of non-SDEGs in DCM CePIN, as well as non-SDEGs in non-DCM CePIN (Table 2). The results showed that the proteins encoded by differentially expressed genes under a specific biological condition tended to play topologically important roles in the corresponding CePINs. Figure 1A shows that proteins with higher CePPI degrees have larger proportions of SDEGs. However, due to the fact that hub proteins in static PIN had higher CePPI degrees in condition-specific CePINs, we examined high- and low-degree proteins in three types of networks

Table 2 Comparison of key topological properties

| Property | SDEGs | DCM CePIN | |
|------------------------|--------|-----------|----------|
| | | Non-SDEGs | P-value |
| Degree | 3.47 | 2.51 | 9.58E-09 |
| Betweenness centrality | 0.0023 | 0.0012 | 1.79E-07 |
| Closeness centrality | 0.1769 | 0.1734 | 4.50E-03 |
| Clustering coefficient | 0.04 | 0.03 | 2.23E-04 |
| Non-DCM CePIN | | | |
| Degree | 3.72 | 2.6 | 7.64E-15 |
| Betweenness centrality | 0.0021 | 0.0011 | 8.79E-15 |
| Closeness centrality | 0.1837 | 0.1796 | 1.64E-04 |
| Clustering coefficient | 0.04 | 0.03 | 1.81E-07 |

The average of each topological property for SDEGs and non-SDEGs in either DCM or non-DCM CePIN is given in the table together with the corresponding P-value. The P-values were calculated using non-parametric Wilcoxon rank-sum tests.

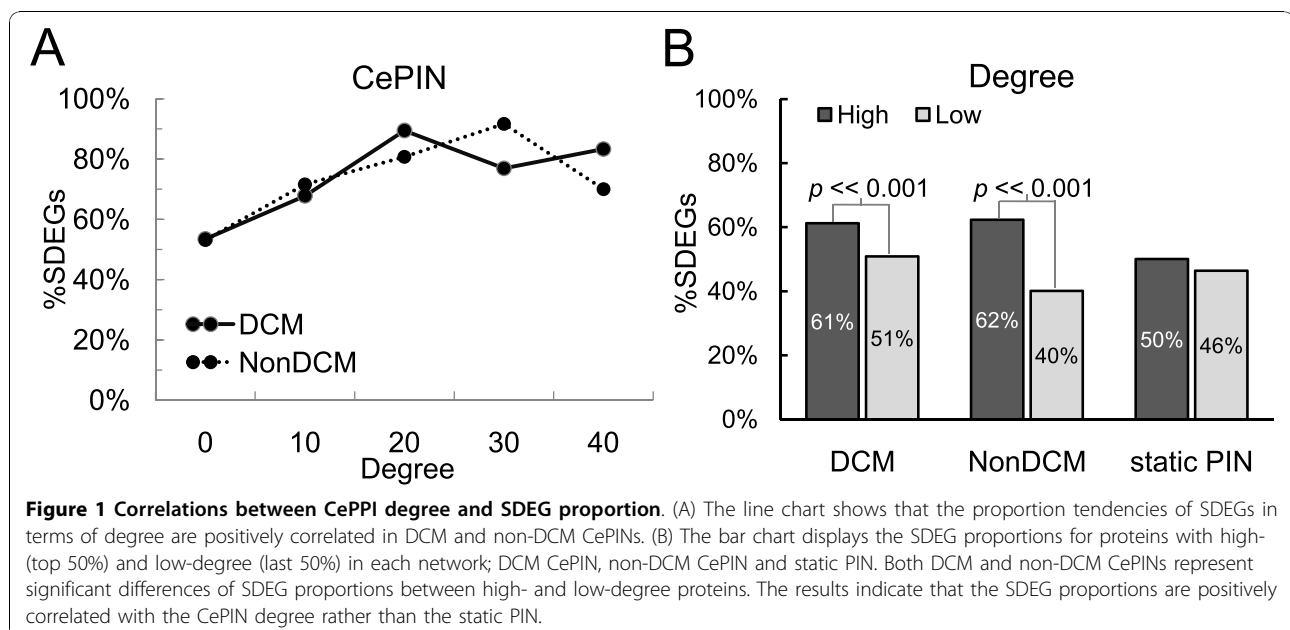
respectively, and compared their proportions of SDEGs (Figure 1B). Although the proportion of SDEGs (50%) for high-degree proteins in static PIN was similar to those with a low-degree (46%), SDEGs were especially enriched for high-degree proteins in both DCM CePIN and non-DCM CePIN (> 60%) in comparison with low-degree proteins (Fisher's exact test, $P < 0.001$). The results implied that hub proteins in CePINs are more likely to be differentially expressed, suggesting modular rewiring of co-expressed interactions between different biological states. Altogether, the condition-specific CePINs could carry more biologically significant network structure that reflects the underlying gene expression dynamics.

Hub proteins of DCM CePIN were also studied in the context of Gene Ontology to identify major biological

Table 3 Top 10 significant level-6 GO annotations of hubs

| GO term | Description | P-value |
|---------|---|----------|
| 16310 | phosphorylation | 2.40E-12 |
| 7167 | enzyme linked receptor protein signaling pathway | 3.01E-10 |
| 6464 | protein modification process | 7.28E-08 |
| 42981 | regulation of apoptosis | 1.41E-05 |
| 6915 | apoptosis | 2.09E-05 |
| 31325 | positive regulation of cellular metabolic process | 2.79E-05 |
| 43068 | positive regulation of programmed cell death | 4.35E-05 |
| 48340 | paraxial mesoderm morphogenesis | 4.95E-05 |
| 48339 | paraxial mesoderm development | 4.95E-05 |
| 45941 | positive regulation of transcription | 9.67E-05 |

roles in each protein category. Table 3 shows the top 10 significantly over-represented functional annotations of DCM hub proteins in GO biological process category (level 6) by BiNGO [12] (Hypergeometric test, Benjamini & Hochberg FDR correction, $P \leq 0.01$). Previously, Camargo and colleagues [6] analyzed the differentially expressed genes of DCM in the context of GO and identified phosphorylation to be significantly over-represented, which is also the most significant functional annotation of hub proteins in DCM CePIN from our study. Kang and colleagues [13] suggested that DCM is highly associated with dysfunction of apoptosis pathways. Consistent with their study, apoptosis-related functions, including regulation of apoptosis, apoptosis, and positive regulation of programmed cell death, came up in our GO identification. These findings were consistent with what has been observed in previous studies and imply that hub



proteins in CePINs are similar to SDEGs with respect of their biological roles.

Identification of Two DCM-Related Functional Modules

Based on the condition-specific CePINs, we propose a new analysis approach to identify condition-specific functional modules. First, for any specific biological condition under study, we selected significantly enriched functions among proteins in condition-specific CePIN. For each enriched function, hypergeometric tests were performed to examine whether the co-expressed interactions between two common-function proteins are also significantly enriched. By applying the proposed analysis to DCM and non-DCM CePINs, two functional modules; muscle contraction and organ morphogenesis, were specifically revealed for DCM, as shown in Figure 2. Table 4 shows the information of the network structure, hypergeometric test results, and classification accuracy of these two modules, respectively. Both modules can be used to discriminate DCM from normal tissue samples with a classification accuracy of 0.79 (muscle contraction) and 0.86 (organ morphogenesis). We further assessed their classification performance using receiver operator characteristic (ROC) curves and tested these with the expression profiles (GSE4172) from another cohort of DCM patients. The expression levels of the genes in each module were averaged to compute a

Table 4 Summary of the two identified DCM-related modules

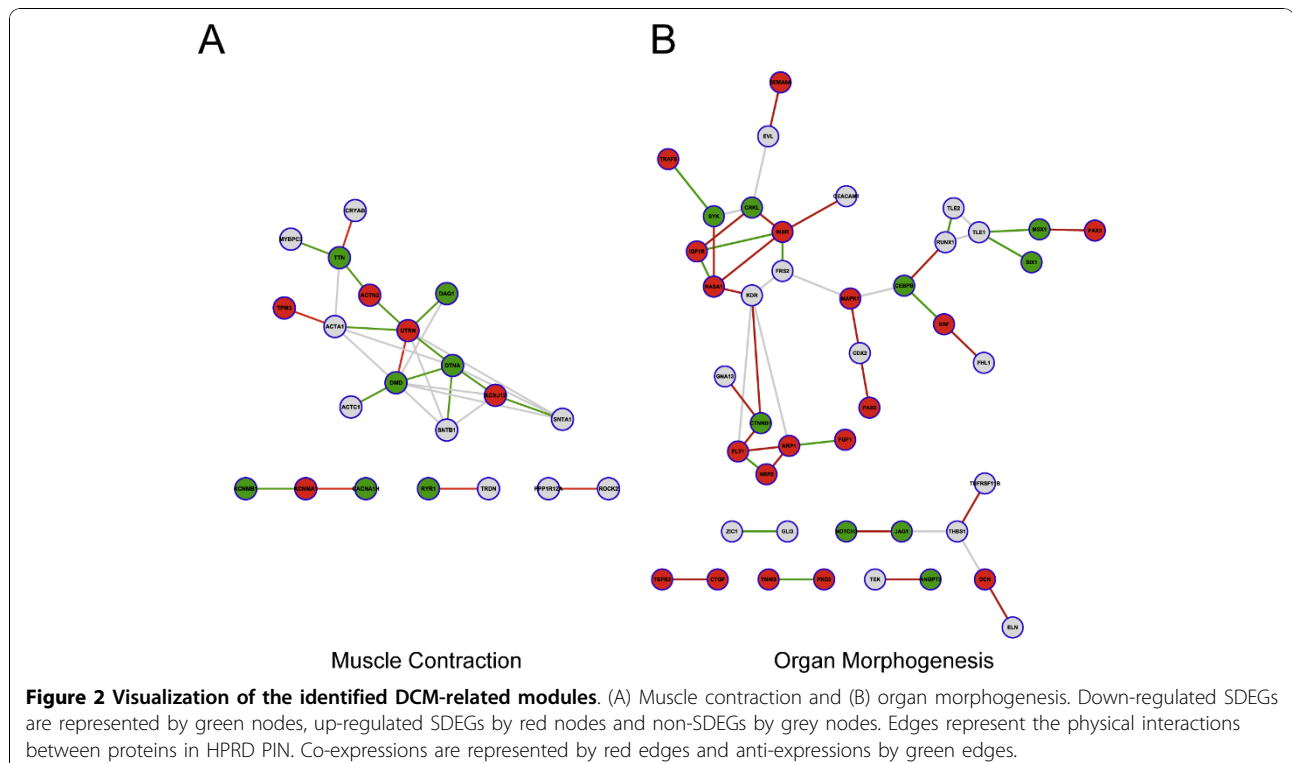
| Module | Node | Ep | Edge | Dp | Accuracy | AUC |
|---------------------|------|-------|------|-------|----------|------|
| Muscle contraction | 21 | 0.016 | 18 | 0.008 | 0.79 | 0.76 |
| Organ morphogenesis | 43 | 0.017 | 34 | 0.007 | 0.86 | 0.89 |

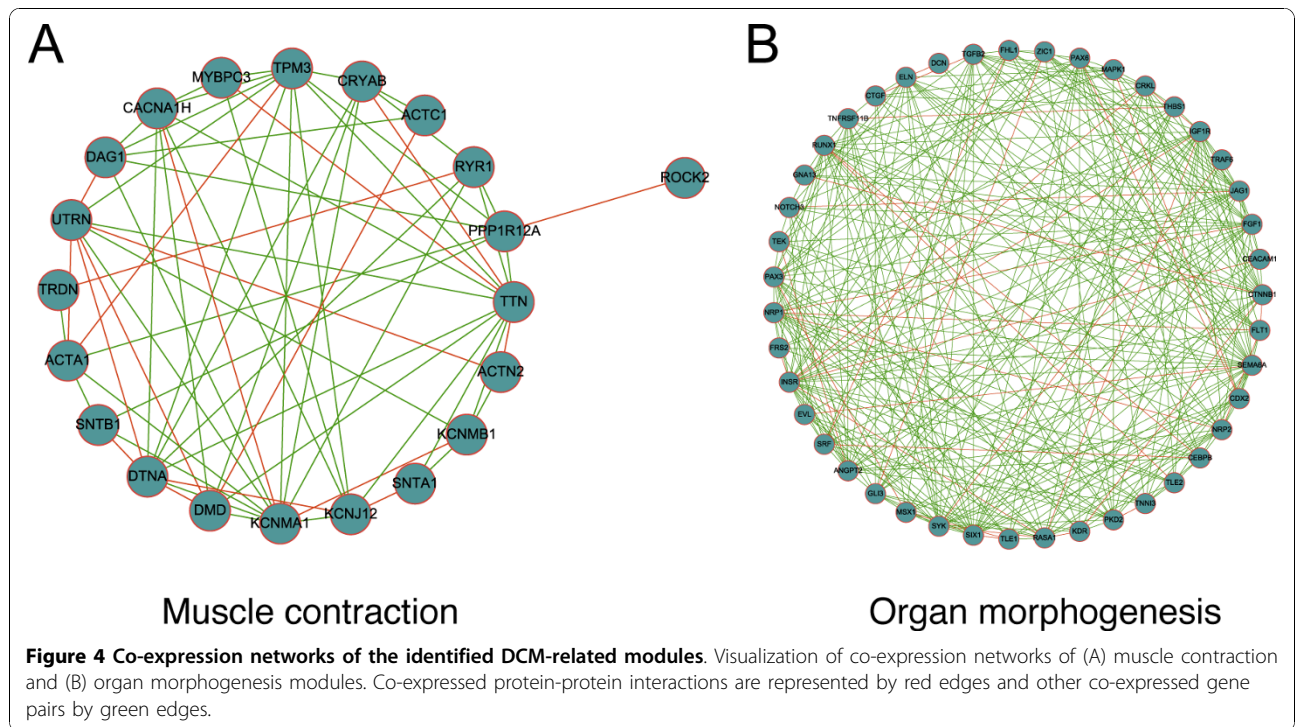
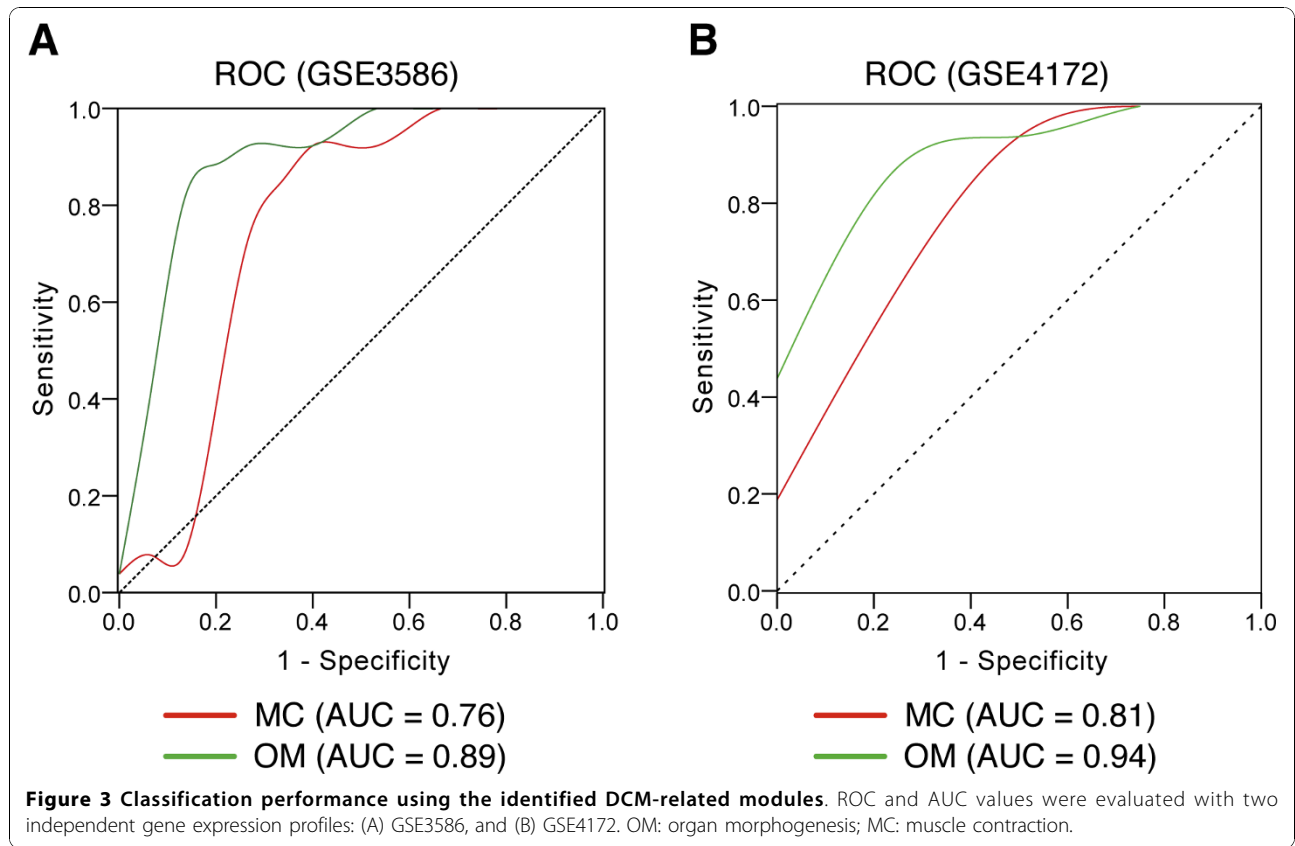
The list includes numbers of nodes and edges, hypergeometric test results (Ep: P-value for functional enrichment; Dp: P-value for functional dyad enrichment) and classification performance (accuracy and AUC).

module activity score. This revealed typical area under curve (AUC) values of 0.75 ~ 0.90 as shown in Figure 3. These results indicate that the identified two modules might serve as potential therapeutic targets of DCM.

To reveal connections between member genes in both modules, their gene co-expression networks were studied and visualized in Figure 4. Both networks were organized with dense structure. The expression profiles of member genes in each module correlated more significantly with each other than those in random subnetworks of equal size (organ morphogenesis: 0.47 versus 0.36 for averaged PCC, $Z = 5.1$, $P < 0.001$; muscle contraction: 0.45 vs. 0.36, $Z = 2.62$, $P = 0.06$). These results suggest that strong connections within each module exist at the transcriptional level.

Subsequently we compared conditional gene expression levels and correlations of gene co-expression between DCM and non-DCM samples to illustrate the





dynamic features of identified functional modules. We found that the average expression levels of member genes in each module changed between two conditions by an amount larger than expected from random sub-networks of equal size (organ morphogenesis: 0.59 vs. 0.39 for average gene expression difference, $Z = 2.58$, $P = 0.01$; muscle contraction: 0.58 vs. 0.39, $Z = 1.7$, $P = 0.06$). Moreover, the average change of member PPI gene expression correlation between conditions were also significantly higher (organ morphogenesis: 0.58 vs. 0.44, $Z = 2.56$, $P = 0.01$; muscle contraction: 0.56 vs. 0.44, $Z = 1.75$, $P = 0.05$).

Figure 5B shows that gene pairs associated with PPIs in the organ morphogenesis module were more strongly correlated in DCM than in non-DCM, implying its activation in DCM. Meanwhile, in the muscle contraction module, gene pairs with PPIs were strongly correlated in both DCM and non-DCM (see Figure 5A, histogram), indicating that it is activated in both DCM and non-DCM. Nonetheless, the CePPIs of muscle contraction module in DCM were quite different from that in non-DCM (see Figure 5A, scatter plot), suggesting that genes in the muscle contraction module changed their associated partners between non-DCM and DCM. Although the muscle contraction module was not significantly differentially expressed and activated in both DCM and non-DCM, it was still discovered by our method due to its diverse CePPIs between DCM and non-DCM. Therefore, the muscle contraction module would have been omitted if the transcriptomic data were not integrated into the interactome data to study the condition-specific PPIs or CePPIs.

We further studied the largest connected components of these two modules with respect to the co-expressed protein interaction difference between DCM and non-DCM to determine their dynamic features. In the muscle contraction module, DTNA, SNTA1, SNTB1, and DMD were shown to be highly correlated with each other in non-DCM CePIN, but not in DCM CePIN (see Figure 6A). Proteins encoded by DTNA, SNTA1, and SNTB1 are components of the cytoplasmic part of dystrophin-associated protein complex (DAP) [14]. In addition, pivot proteins in both non-DCM and DCM, which have relatively more co-expressed interacting partners, were observed to change from DTNA and DMD to UTRN.

Similar dynamic features were also observed in the organ morphogenesis module. We roughly defined this module into two major regions: the upper diamond, formed by INSR, CRKL, IGF1R and RASA1, and the bottom triangle, by FLT1, NRP2 and NRP1 (see Figure 6B). From the results, it is evident that the communication between the diamond and the triangle in DCM CePIN was bridged by KDR and CTNNB1, but

disconnected in non-DCM CePIN. Moreover, the diamond structure in non-DCM CePIN was observed to have collapsed. These changes in the muscle contraction and organ morphogenesis modules may hold some clues to the progression of DCM.

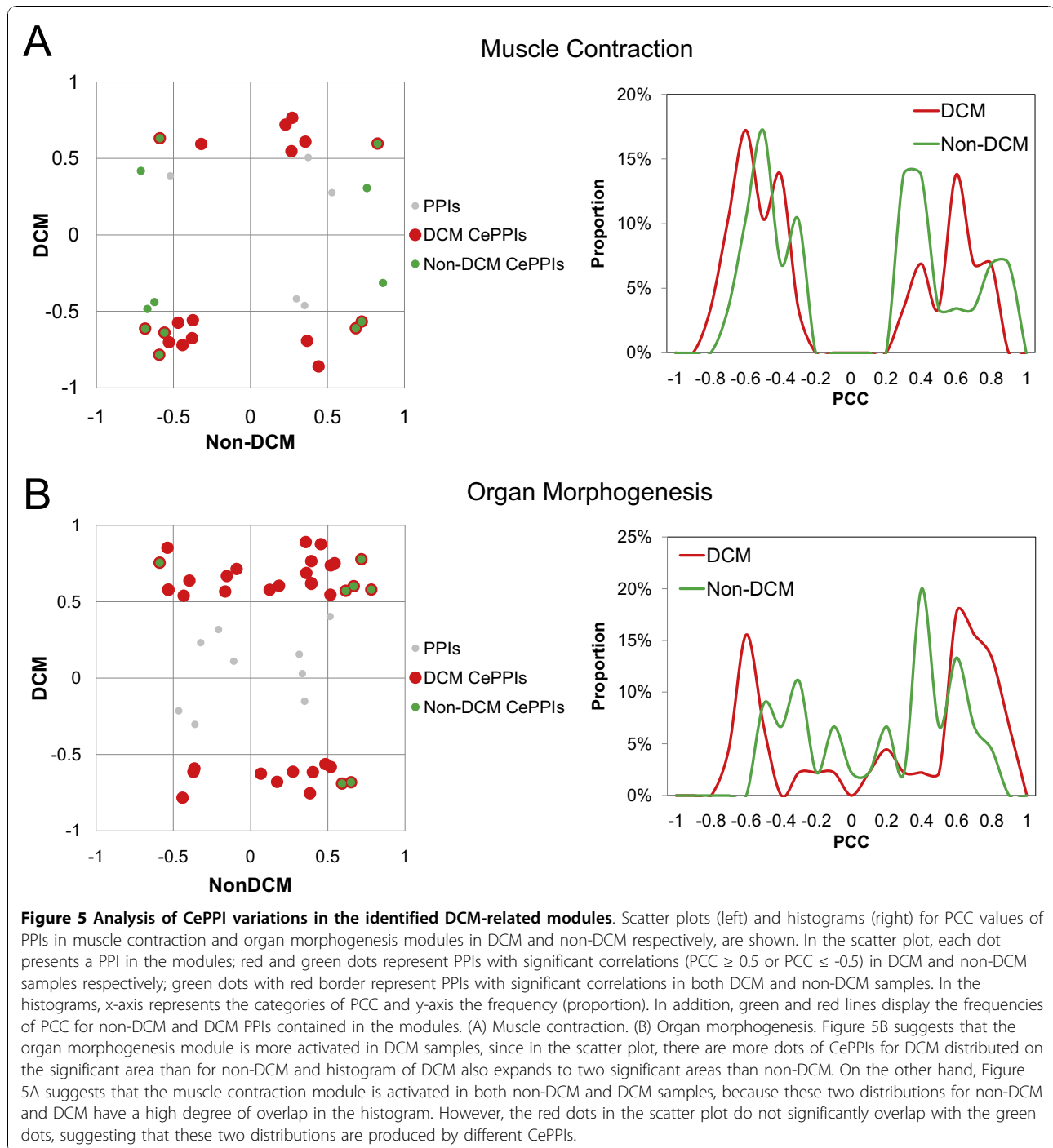
Discussion

Protein-protein interaction networks cover all possible interactions regardless of when or where the interaction takes place. In this sense, they are static. By integrating gene expression profiles of DCM with human protein-protein interaction networks, we successfully extracted two co-expressed protein interaction networks (CePINs), i.e. DCM and non-DCM. Here, we showed that DCM and non-DCM CePINs exhibit substantial differences in co-expressed protein-protein interactions, even though their network structures are similar. The differences may be attributed to gene expression variations and interaction rewiring between DCM and non-DCM conditions. We suggest that CePINs are able to reveal condition-specific interactions and the dynamic features hidden in static protein-protein interaction networks.

Next, we showed that hub proteins in CePINs tended to be SDEGs compared to non-hub proteins. In CePINs, proteins with higher degrees imply that they have more direct interacting partners co-expressed in gene expression levels; therefore any significant modification in their expression levels might influence more interacting partners. This observation suggests that once gene expression of hub proteins is changed, it is expected to cause greater expression variations to its neighboring interaction network in DCM.

Since our analysis relies on PCC calculated from gene expression data to define CePPIs and construct condition-dependent networks, we have carried out some further examinations about its robustness. First, we performed the same analysis with tightened PCC threshold, $P < 0.01$, and obtained consistent results (see Table S1 - S4, Figure S1 - S2 in Additional File 1). Second, Li *et al.* [15] recently revealed that correlations of gene expression levels to disease states could vary a lot with randomly selected subsets of the samples from one single microarray data set. Under this light, we have performed re-sampling of our gene expression data and found that both the recovery rate of CePPIs and the identification rates of the two DCM-related modules decreased as we lowered the sample size (Figure S3 - S4 in Additional File 1). However, the identification rates of the two DCM-related modules were much higher than any other modules, indicating that these two identified modules were robustly related to DCM.

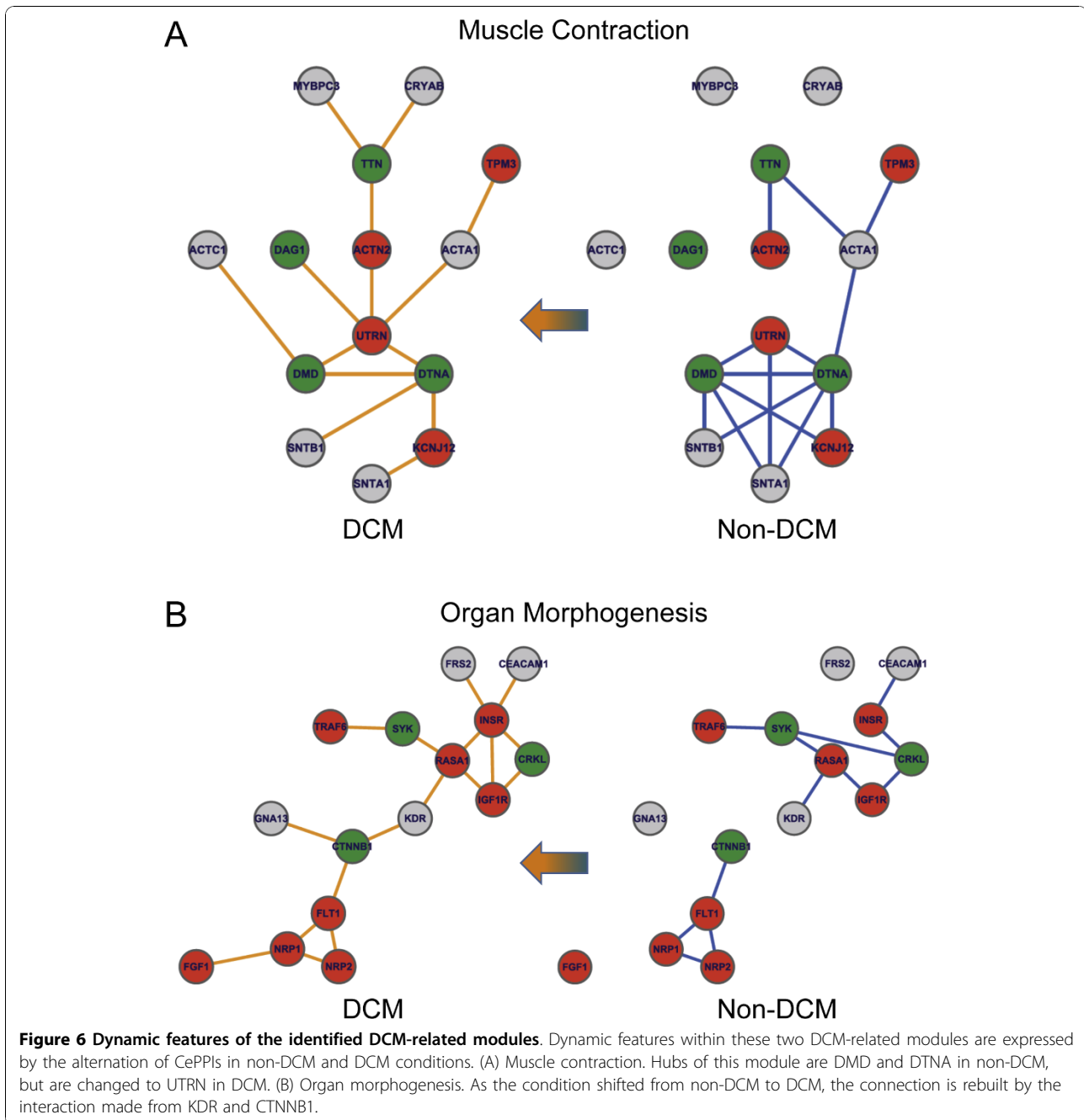
Since the human PPI data are still incomplete and noisy, there are different curated collections of human PPI data available. To examine whether our analysis was



robust against the PIN we used, we performed the same analysis with an expanded PIN integrating the PPI data from HPRD [16] and BioGRID [17] databases and obtained consistent results (see Table S5 - S8, Figure S5 - S6 in Additional File 1).

One of the major symptoms of heart failure is the inability of the heart to sufficiently supply blood flow to

the rest of the body. This is strongly related to heart muscle contraction efficiency. The failing heart undergoes morphological changes and becomes weakened and enlarged in DCM, the most common form of cardiomyopathy. Our findings in DCM-related modules of muscle contraction and organ morphogenesis were consistent with the known symptoms of cardiomyopathy.



Consequently, we further investigated these two modules in relation to the underlying molecular mechanisms of dilated cardiomyopathy.

In the muscle contraction module, three SDEGs, DMD, DTNA and UTRN, form a closed loop, implying topological significance. Dystrophin, encoded by DMD, is a recessive, fatal X-linked disorder. It appears to stabilize the sarcolemma and protects muscle fibers from long-term contraction-induced damage and necrosis [18], though its precise roles at the cellular level are still

to be elucidated. Dystrobrevin-alpha, encoded by DTNA, belongs to the dystrobrevin subfamily of the dystrophin family and is a component of the dystrophin-associated protein complex (DAP) [14]. Disruption of DAP is associated with various forms of muscular dystrophy. Dystrophin binds to the intracellular cytoskeleton by associating with actin filaments at its N-terminus, whereas at its C-terminus dystrophin interacts with members of the DAP, such as β -dystroglycan, which is encoded by DAG1. Dystrophin therefore links the

intracellular microfilament network of actin to a complex series of linking proteins in the cell membrane, and hence to the extracellular matrix [19]. In the muscle contraction module, both DMD and DTNA were significantly down-regulated. The absence of dystrophin was suggested to cause the collapse of the entire DAP and plasma membrane, leading to muscle damage [14]. On the other hand, mutations of DTNA are associated with left ventricular non-compaction with congenital heart defects [20]. These defects might lead to cardiac muscle damage and possibly DCM.

Utrophin encoded by UTRN shares both structural and functional similarities with the dystrophin gene. Mouse studies suggested that utrophin may serve as a functional substitute for the dystrophin gene and can be viewed as a rescue protein in muscular dystrophy caused by abnormal dystrophin expression [21]. In our results, UTRN was significantly up-regulated, which might illustrate the rescue action of utrophin in the absence of dystrophin. However, no protein-protein interactions exist or are found between UTRN and ACTC1, which encodes the actin in cardiac cells, suggesting that progressive heart failure could be due to the failure of UTRN rescue. Based on these results, we proposed a hypothesis to explain how the muscle contraction module affects DCM progression (Figure 7). Recently, *Bostick* and colleagues [22] reported that utrophin up-regulation alone (without dystrophin expression) leads to DCM in mice, which provides strong evidence in support of our hypothesis.

On the other hand, in the organ morphogenesis module, we noticed that the largest connected component

contains two major clusters. The two major clusters consisted of insulin pathway-related genes, including IGF1R (insulin-like growth factor 1 receptor) and INSR (insulin receptor) and vascular endothelial growth factor (VEGF) pathway-related genes, including FLT1 (VEGFR1), NRP1, NRP2 and KDR (VEGFR2). These two pathways have both been reported to be important in cardiac remodeling [23]. Proper IGF1R and INSR signaling plays an essential role in cardiac function, and the disruption of this signaling induces the onset of DCM in knockout mice [24,25]; while the VEGF pathway is crucial in vasculogenesis and angiogenesis, which was reported to be altered in DCM [26]. The significant up-regulation of these two clusters of genes can therefore signify autosomal repair for damage caused by hypoxia induced by early DCM symptoms. Malfunction of these activated pathways are possible reasons for the disease progression.

With focus on these pathways, we compared the sub-networks in DCM and non-DCM patients and found several notable points. First, these two clusters were not independent from each other, but were linked by a string of genes: RASA1, KDR, CTNNB1, and FLT1. RASA1, encoding p120-RasGAP which activates RAS GTPase, is best known for its negative regulation of the RAS/MAPK pathway downstream of several growth factor pathways responsible for cell proliferation, including the IGF-1, insulin and VEGF pathway. Proper activation of the RAS-dependent pathway is important for the functions of these pathways. The up-regulation and the linkage between RASA1 and INSR may infer a possible negative regulation of the Ras-dependent pathway. We

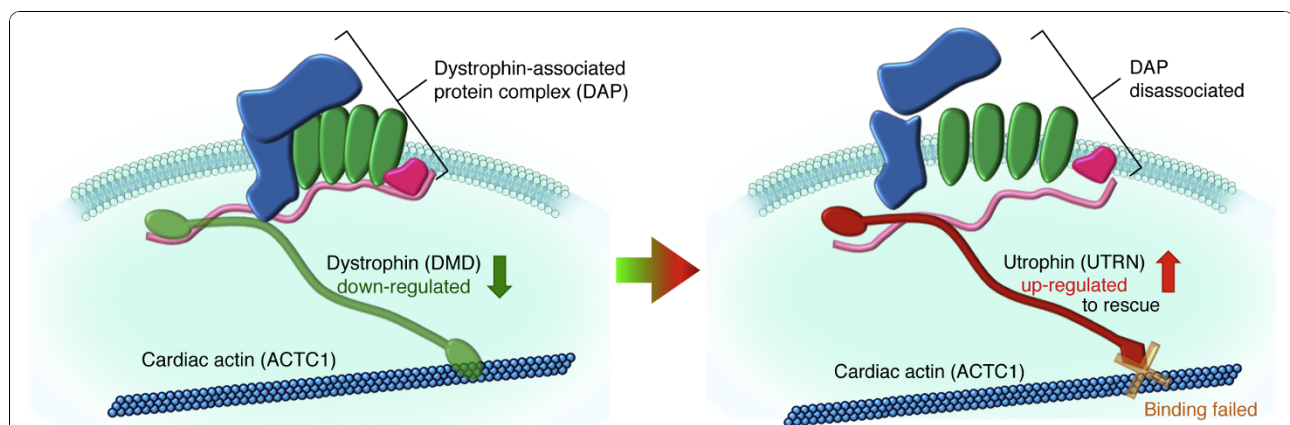


Figure 7 Hypothesis of the underlying molecular mechanism of DCM. In a normal myocardium cell, the dystrophin encoded by DMD builds a link between the intracellular microfilament network of actin and the dystrophin-associated protein complex (DAP) to protect the myocardium from long-term contraction-induced damage and necrosis. In the left panel, down-regulation of dystrophin (green rod-shaped protein) causes the collapse of the DAP and plasma membrane, leading to myocardium damage and possibly heart failure. To rescue this defect, utrophin (red rod-shaped protein) encoded by UTRN, which is the functional substitute for dystrophin, is up-regulated, as shown in the right panel. However, utrophin (UTRN) could not bind to the actin ACTC1 in cardiac cells. Hence, the connection between the actin and DAP is broken, leading to the disassociation of the DAP. As a result, cardiac cell membranes become permeable and gradually lyse, resulting in tissue destruction and heart failure.

also found CrkL, a protein that mediates Ras-dependent activation, to be significantly down-regulated in DCM patients. These observations imply the negative regulation of growth factor signaling involving insulin and insulin-like growth factor. Second, the down-regulation of CTNNB, which encodes beta-catenin in VE-cadherin essential for contact inhibition of VEGF-induced proliferation [26,27], infers a malfunction in the control of VEGF-induced vasculogenesis. The failure of beta-catenin regulation and defective vascularization have been observed in idiopathic DCM [26].

Our findings regarding the organ morphogenesis module successfully revealed possible integration of two important pathways in DCM and the crucial role that RASA1 up-regulation and CTNNB1 down-regulation might play in the etiology of DCM.

Conclusions

Altogether we have developed a network-based comparative analysis approach that integrates protein-protein interactions with gene expression profiles and biological function annotations to reveal dynamic functional modules under different biological states. Application to DCM reveals two functional modules with dynamic features accounting for the underlying disease mechanisms. The revealed molecular modules might be used as potential drug targets and provide new directions for heart failure therapy.

Methods

Protein Interaction Network and Expression Data

The human protein interaction network (PIN) was downloaded from Human Protein Reference Database (HPRD) [16], and only the largest connected component, containing 9,059 proteins and 34,869 interactions, was studied.

The gene expression data of DCM were retrieved from Gene Expression Omnibus (GEO), accession number GSE3586 [8], containing 37,530 genes and 28 samples (13: DCM, 15: non-DCM) in total, with 6,475 genes involved in HPRD PIN. Another set of expression profiles of DCM was retrieved from GEO accession number GSE4172 [10] to evaluate the classification performance of identified modules.

Co-expressed Protein Interaction Networks

To obtain the condition-specific information, an auxiliary network, called the co-expressed protein interaction network (CePIN) was constructed. This procedure involved several data integration and search steps for extracting CePINs from HPRD PIN (or called "static" PIN as opposed to condition-specific CePIN), as outlined in

Figure 8. Correlation of gene expression profiles between each pair of interacting proteins in PIN is evaluated by Pearson correlation coefficient (PCC).

PCC of paired genes (X and Y), which encode interacting proteins in PIN, is defined as:

$$PCC(X, Y) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{Exp(X, i) - \overline{Exp}(X)}{\sigma(X)} \right) \cdot \left(\frac{Exp(Y, i) - \overline{Exp}(Y)}{\sigma(Y)} \right)$$

where n is the number of condition-specific samples; $Exp(X, i)$ ($Exp(Y, i)$) is the expression level of gene X (Y) in the sample i under a specific condition (DCM or non-DCM); $\overline{Exp}(X)$ ($\overline{Exp}(Y)$) represents the average expression level of gene X (Y) and $\sigma(X)$ ($\sigma(Y)$) represents the standard deviation of expression level of gene X (Y). Larger absolute values of PCC indicate higher correlation between evaluated gene pairs. Those with a P -value of less than or equal to 0.05 were considered as significantly correlated. Protein-protein interactions between proteins encoded by significantly correlated gene pairs are defined as co-expressed protein-protein interactions (CePPIs). Based on this definition, the co-expressed protein interaction network (CePIN) is defined as the set of CePPIs. Note that we defined CePPI as co-expression in gene expression levels instead of protein concentrations since we were lack of the corresponding proteome data. Although a gene expression level cannot always represent its protein concentration, previous studies have observed notable correlations between them [28]. If proteome data are available, the same analysis procedures described here can be applied with replacement of $Exp(X)$ and $Exp(Y)$ by protein concentrations.

Significantly Differentially Expressed Genes and Hub Proteins

Significantly differentially expressed genes (SDEGs) were determined by Wilcoxon rank sum test ($P \leq 0.05$, DCM against non-DCM). Up-regulated and down-regulated SDGEs were defined as genes expressed significantly higher or lower respectively in DCM than in non-DCM. DCM hub proteins were defined as nodes involved with more than 23 DCM CePPIs, since these proteins were among the top 1% of the CePPI degree distribution of the DCM CePIN.

Topological Analysis of Networks

Four topological measures were used in this analysis: degree, betweenness centrality, closeness centrality and clustering coefficient [29]. The degree is the number of observed interactions of a given protein. Betweenness centrality (BC) measures the importance of a specific

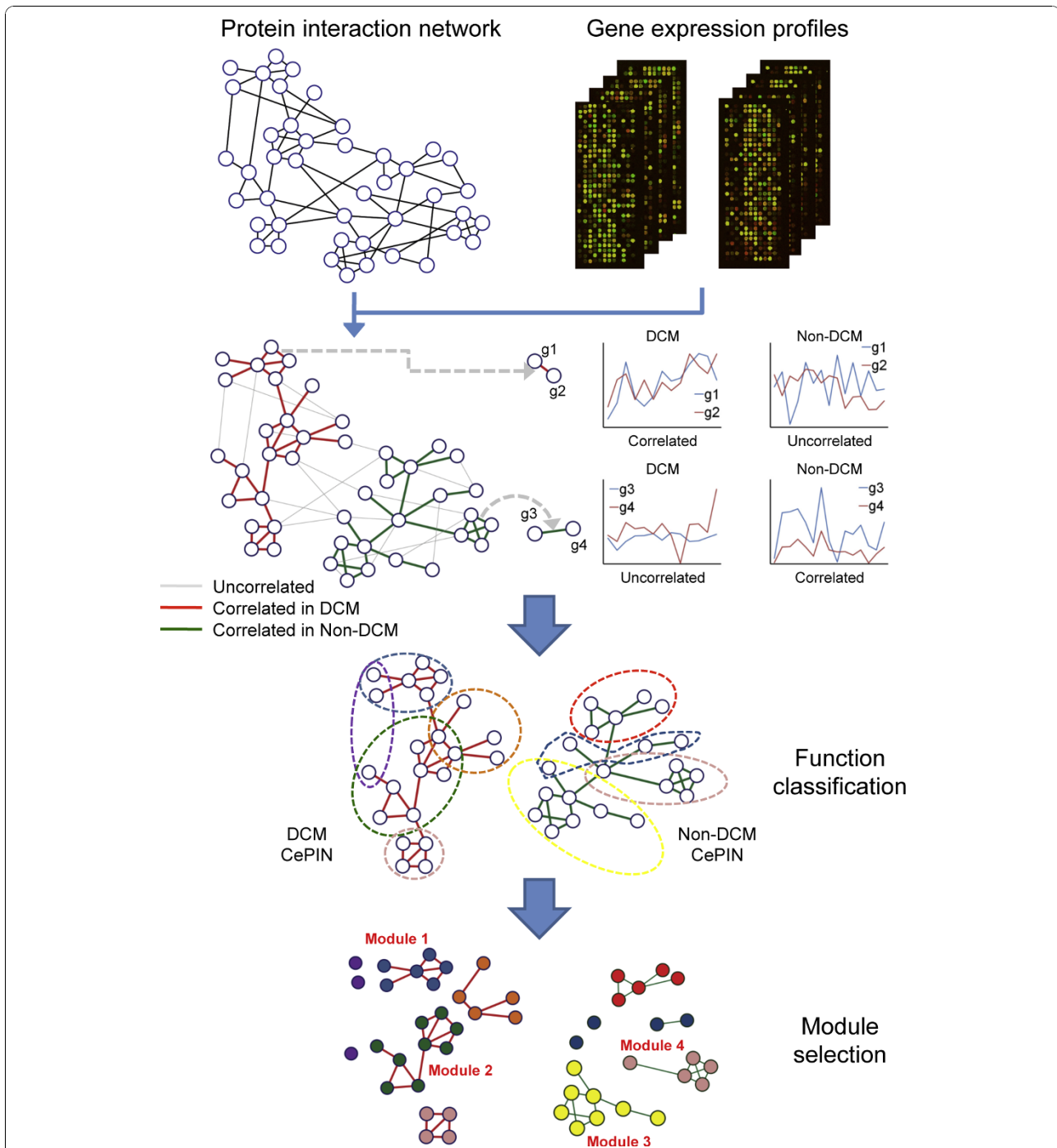


Figure 8 Flow diagram of the proposed integrative network analysis procedure. Flow diagram of discovery DCM related modules from static PIN by integrating gene expression profiles. In static PIN, each PPI was assigned a correlation derived from Pearson correlation coefficient between paired genes engaged in a PPI. Then, CePPIs and PPIs with significant correlations were utilized to construct the CePINs which were extracted from the static PIN. Next, gene sets which are provided with significant functional enrichment in GO annotation categories were grouped and isolated from CePINs. Finally, functional subnetworks which were only involved in DCM CePIN were treated as candidate DCM related modules. In addition, only candidate DCM related modules with GO annotations of levels ≥ 5 were considered as DCM related modules.

protein i in relating any paired proteins in PIN, and is defined by:

$$BC_i = \frac{SP_i}{C_2^N},$$

where SP_i is the number of shortest path pass through protein i and N represents the number of proteins in PIN. Closeness centrality (CC) represents how close an evaluated protein is relative to all the others and is defined as the reciprocal of the mean of the shortest path lengths (\overline{SPL}_i) from protein i to the rest of proteins in PIN,

$$CC_i = \frac{1}{\overline{SPL}_i}.$$

Higher closeness centrality implies shorter distance between the evaluated protein and all the other proteins in PIN. The clustering coefficient (C) of a protein determines how frequently its interacting partners interact with each other and is defined by:

$$C_i = \frac{e_{NB_i}}{C_2^{|NB_i|}},$$

in which e_{NB} is the number of observed interactions between interacting partners of protein i , and $|NB_i|$ represents the number of its interacting partners. $C_2^{|NB_i|}$ gives the number of all possible interactions among its interacting partners. In this study, only the largest connected component of each CePIN was considered.

Identification of Condition-Specific Functional Modules

To identify heart failure related modules, we applied a comparative analysis of CePINs. This analysis included several steps of functional module discovery and selection, as illustrated in Figure 8. First, GO annotation was utilized to choose gene sets, which are involved in DCM (or non-DCM) CePIN, with significant enrichment of functional categories in biological process ontology ($P \leq 0.05$). The P -value of functional enrichment is determined by a hypergeometric test. The hypergeometric distribution is described as follows:

$$P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}.$$

X denotes the evaluated functional category in GO. N represents the number of GO annotated genes, which

appeared in our microarray data as well as in HPRD PIN, while m represents that in DCM or non-DCM CePIN. n represents the number of genes which are annotated as the evaluated GO functional category in HPRD PIN. Thus, this formula calculated the probability of the evaluated functional category that had k genes in CePIN. The calculated P -value is then adjusted by applying the Benjamini and Hochberg multiple testing procedure [30] to control the false discovery rate (FDR) at significance level of 0.05.

To further retain significantly denser functional subnetworks, we tested the functional homogeneity of each significant functionally enriched gene sets in respect to CePPIs. For this purpose, we defined functional dyads as paired proteins that formed CePPIs and shared common biological process as annotated by GO. The significance of the functional subnetwork is determined by P -value (≤ 0.05) derived from a modified hypergeometric test:

$$P_e(X = k_e) = \frac{\binom{m_e}{k_e} \binom{N_e - m_e}{n_e - k_e}}{\binom{N_e}{n_e}},$$

e is the abbreviation of the functional dyad. Each symbol represents the same meaning with the previous one in the original hypergeometric test, but the counting targets are changed from functional genes to functional dyads. The same multiple testing procedure is performed to adjust the P -value.

After these two tests, only subnetworks with significantly high functional homogeneity in both genes and interactions would be kept for follow-up comparative analysis in order to reveal candidate DCM related modules. We defined $F_{DCM_candidate}$ as the set including all GO annotations of significantly denser functional subnetworks in DCM and $F_{non-DCM_candidate}$ in non-DCM. Then, $F_{DCM_exclusive}$ could be defined as the difference between $F_{non-DCM_candidate}$ and $F_{DCM_candidate}$:

$$F_{DCM_exclusive} = F_{DCM_candidate} - F_{non-DCM_candidate}$$

Functional subnetworks mapped in $F_{DCM_exclusive}$ will be designated as candidate DCM related modules. For functional specificity, only candidate DCM related modules with GO annotations greater than or equal to level 5 were considered as DCM related modules.

Evaluation of Classification Accuracy

Hierarchical clustering was used to group samples into two categories (i.e. DCM or non-DCM) according to the gene expression levels of members in each module. Gene expression distance between different samples was

calculated by Euclidean distance. The classification tree produced by hierarchical clustering was separated at the root into two groups (sub-trees). The group with more DCM samples is defined as positive and the other group is negative. The number of DCM samples in the positive group is defined as true positive (*TP*) and the number of non-DCM samples clustered into this group is false positive (*FP*). The number of non-DCM samples in the negative group is defined as true negative (*TN*) and the number of the rest DCM samples is false negative (*FN*).

They were used to evaluate sensitivity ($\frac{TP}{TP + FN}$) and specificity ($\frac{TN}{FP + TN}$). Then, accuracy was calculated by

$\frac{\text{Sensitivity} + \text{Specificity}}{\text{Number of samples}}$ and was used to measure the

classification ability of each module. The receiver operating characteristic curve (ROC) was obtained according to the module activity score of each sample, which was defined as the average expression level of all member genes in the module.

Additional material

Additional file 1: Supplementary information. This file contains the robustness analysis of our results in different thresholds of PCC, sample sizes of gene expression profiles, and protein interaction network.

Acknowledgements

This work was supported by National Science Council of Taiwan; National Health Research Institutes, Taiwan (NHRI-EX98-9819PI); and National Taiwan University Frontier and Innovative Research Projects.

Author details

¹Institute of Biomedical Informatics, Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei 112, Taiwan. ²Graduate Institute of Biomedical Electronics and Bioinformatics, Center for Systems Biology and Bioinformatics, National Taiwan University, Taipei 106, Taiwan. ³Department of Physics, National Dong Hwa University, Hualien 974, Taiwan. ⁴Department of Life Science, Institute of Molecular and Cellular Biology, National Taiwan University, Taipei 106, Taiwan.

Authors' contributions

CCL implemented the computational method and carried out the analysis. JTH, CYW, and YJO helped to perform the analysis. CCL, JTH, and CYW drafted the manuscript. CCL and HCH conceived the study. HFJ and HCH participated in the design and coordination of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

Received: 29 June 2010 Accepted: 15 October 2010

Published: 15 October 2010

References

- Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M: **Evidence for dynamically organized modularity in the yeast protein-protein interaction network.** *Nature* 2004, **430**:88-93.
- Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, Hou L, Zhang Q, Zhang Y, Han JD: **A modular network model of aging.** *Mol Syst Biol* 2007, **3**:147.
- Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: **Dynamic modularity in protein interaction networks predicts breast cancer outcome.** *Nat Biotechnol* 2009, **27**:199-204.
- Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
- Camargo A, Azuaje F: **Linking gene expression and functional network data in human heart failure.** *PLoS ONE* 2007, **2**:e1347.
- Camargo A, Azuaje F: **Identification of dilated cardiomyopathy signature genes through gene expression and network data integration.** *Genomics* 2008, **92**:404-413.
- McMurray JJ, Pfeffer MA: **Heart failure.** *Lancet* 2005, **365**:1877-1889.
- Barth AS, Kuner R, Bunes A, Ruschhaupt M, Merk S, Zwermann L, Kaab S, Kreuzer E, Steinbeck G, Mansmann U, et al: **Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies.** *J Am Coll Cardiol* 2006, **48**:1610-1617.
- Kaab S, Barth AS, Margerie D, Dugas M, Gebauer M, Zwermann L, Merk S, Pfeufer A, Steinmeyer K, Bleich M, et al: **Global gene expression in human myocardium-oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure.** *J Mol Med* 2004, **82**:308-316.
- Wittchen F, Suckau L, Witt H, Skurk C, Lassner D, Fechner H, Sipo I, Ungethüm U, Ruiz P, Pauschinger M, et al: **Genomic expression profiling of human inflammatory cardiomyopathy (DCMI) suggests novel therapeutic targets.** *J Mol Med* 2007, **85**:257-271.
- Ekman D, Light S, Bjorklund AK, Elofsson A: **What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*?** *Genome Biol* 2006, **7**:R45.
- Maere S, Heymans K, Kuiper M: **BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics* 2005, **21**:3448-3449.
- Kang PM, Izumo S: **Apoptosis and heart failure: A critical review of the literature.** *Circ Res* 2000, **86**:1107-1113.
- Ervasti JM, Ohlendieck K, Kahl SD, Gaver MG, Campbell KP: **Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle.** *Nature* 1990, **345**:315-319.
- Li J, Lenferink AEG, Deng Y, Collins C, Cui Q, Purisima EO, O'connor-Mccourt MD, Wang E: **Identification of high-quality cancer prognostic markers and metastasis network modules.** *Nature Communications* 2010, **1**:34.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al: **Human Protein Reference Database-2009 update.** *Nucleic Acids Res* 2009, **37**:D767-772.
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucleic Acids Res* 2006, **34**:D535-539.
- Davies KE, Nowak KJ: **Molecular mechanisms of muscular dystrophies: old and new players.** *Nat Rev Mol Cell Biol* 2006, **7**:762-773.
- Ervasti JM, Campbell KP: **Membrane organization of the dystrophin-glycoprotein complex.** *Cell* 1991, **66**:1121-1131.
- Ichida F, Tsubata S, Bowles KR, Haneda N, Uese K, Miyawaki T, Dreyer WJ, Messina J, Li H, Bowles NE, Towbin JA: **Novel gene mutations in patients with left ventricular noncompaction or Barth syndrome.** *Circulation* 2001, **103**:1256-1263.
- Burton EA, Tinsley JM, Holzfeind PJ, Rodrigues NR, Davies KE: **A second promoter provides an alternative target for therapeutic up-regulation of utrophin in Duchenne muscular dystrophy.** *Proc Natl Acad Sci USA* 1999, **96**:14025-14030.
- Bostick B, Yue Y, Long C, Duan D: **Prevention of dystrophin-deficient cardiomyopathy in twenty-one-month-old carrier mice by mosaic dystrophin expression or complementary dystrophin/utrophin expression.** *Circ Res* 2008, **102**:121-130.
- Shiojima I, Sato K, Izumiya Y, Schiekofer S, Ito M, Liao R, Colucci WS, Walsh K: **Disruption of coordinated cardiac hypertrophy and angiogenesis contributes to the transition to heart failure.** *J Clin Invest* 2005, **115**:2108-2118.
- Laustsen PG, Russell SJ, Cui L, Entingh-Pearsall A, Holzenberger M, Liao R, Kahn CR: **Essential role of insulin and insulin-like growth factor 1 receptor signaling in cardiac development and function.** *Mol Cell Biol* 2007, **27**:1649-1664.

25. Dube PE, Forse CL, Bahrami J, Brubaker PL: **The essential role of insulin-like growth factor-1 in the intestinal tropic effects of glucagon-like peptide-2 in mice.** *Gastroenterology* 2006, **131**:589-605.
26. Roura S, Planas F, Prat-Vidal C, Leta R, Soler-Botija C, Carreras F, Llach A, Hove-Madsen L, Pons Llado G, Farre J, et al: **Idiopathic dilated cardiomyopathy exhibits defective vascularization and vessel formation.** *Eur J Heart Fail* 2007, **9**:995-1002.
27. Grazia Lampugnani M, Zanetti A, Corada M, Takahashi T, Balconi G, Breviario F, Orsenigo F, Cattelino A, Kemler R, Daniel TO, Dejana E: **Contact inhibition of VEGF-induced proliferation requires vascular endothelial cadherin, beta-catenin, and the phosphatase DEP-1/CD148.** *J Cell Biol* 2003, **161**:793-804.
28. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS: **Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise.** *Nature* 2006, **441**:840-846.
29. Wasserman S, Faust K: *Social Network Analysis: Methods and Applications* Cambridge University Press 1994.
30. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency.** *Ann Statist* 2001, **29**:1165-1188.

doi:10.1186/1752-0509-4-138

Cite this article as: Lin et al.: Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy. *BMC Systems Biology* 2010 **4**:138.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

