

# SCIENTIFIC REPORTS



OPEN

## Functional Analysis and Characterization of Differential Coexpression Networks

Received: 27 February 2015

Accepted: 27 July 2015

Published: 18 August 2015

Chia-Lang Hsu<sup>1</sup>, Hsueh-Fen Juan<sup>1,2,3</sup> & Hsuan-Cheng Huang<sup>4</sup>

Differential coexpression analysis is emerging as a complement to conventional differential gene expression analysis. The identified differential coexpression links can be assembled into a differential coexpression network (DCEN) in response to environmental stresses or genetic changes. Differential coexpression analyses have been successfully used to identify condition-specific modules; however, the structural properties and biological significance of general DCENs have not been well investigated. Here, we analyzed two independent *Saccharomyces cerevisiae* DCENs constructed from large-scale time-course gene expression profiles in response to different situations. Topological analyses show that DCENs are tree-like networks possessing scale-free characteristics, but not small-world. Functional analyses indicate that differentially coexpressed gene pairs in DCEN tend to link different biological processes, achieving complementary or synergistic effects. Furthermore, the gene pairs lacking common transcription factors are sensitive to perturbation and hence lead to differential coexpression. Based on these observations, we integrated transcriptional regulatory information into DCEN and identified transcription factors that might cause differential coexpression by gain or loss of activation in response to different situations. Collectively, our results not only uncover the unique structural characteristics of DCEN but also provide new insights into interpretation of DCEN to reveal its biological significance and infer the underlying gene regulatory dynamics.

In biological systems, distinct groups of molecules that are functionally coordinated, physically interacting or co-regulated, drive complex biological processes. To dissect the complexity of biological systems, a complete map of intermolecular interactions is required. Therefore, numerous large networks have been measured systematically for humans and many other model species. These networks include physical attachments underlying protein-protein interactions, kinase-substrate interactions, protein-DNA interactions and metabolic reactions, as well as functional associations such as epistasis, synthetic lethality relationships and correlated expression between genes<sup>1,2</sup>. These various molecular networks have been successfully applied to address different biological questions, such as identification of disease genes<sup>1</sup> and drug discovery<sup>3,4</sup>.

Molecular interactions can change dramatically in response to different conditions, such as environmental stresses and genetic changes. In other words, a molecule interaction can be present in some conditions, but absent in others. However, most large-scale networks to date have been measured under a single static condition, usually standard laboratory growth media. Understanding of network dynamics has been achieved to some extent by integrating static networks with gene expression profiles<sup>5</sup>. However, these approaches are typically unable to identify new interactions that are condition-specific. To completely understand the cellular dynamics, various differential network analyses have been proposed by experimental mapping of networks across multiple conditions<sup>6</sup>. Analogous to differential gene

<sup>1</sup>Department of Life Science, National Taiwan University, Taipei 10617, Taiwan. <sup>2</sup>Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan. <sup>3</sup>Institute of Molecular and Cellular Biology, National Taiwan University, Taipei 10617, Taiwan. <sup>4</sup>Institute of Biomedical Informatics, Center of Systems and Synthetic Biology, National Yang-Ming University, Taipei 11221, Taiwan. Correspondence and requests for materials should be addressed to H.F.J. (email: yukijuan@ntu.edu.tw) or H.C.H. (email: hsuancheng@ym.edu.tw)

expression analyses, differential network analyses involve pairwise subtraction of interactions that have been mapped in differential experimental conditions. The extracted interactions are differentially present, absent or modified and relevant to the studied condition or phenotype. Several differential network mappings have revealed that the architecture of a molecular network can be massively re-wired during a cellular response and demonstrated the power of differential network analyses for elucidating biological mechanisms<sup>7,8</sup>.

Coexpression networks are typically constructed from gene expression data using correlation-based inference methods. These networks have been commonly used to reveal gene functions and investigate gene regulatory systems<sup>9,10</sup>. However, similar to other static molecular networks, gene coexpression networks only disclose the gene regulatory interactions under specific conditions. To understand the dynamics of cellular regulation, differential coexpression analysis incorporating regulatory changes between different conditions is emerging. Differential coexpression analysis investigates the differences among gene interconnections by calculating the expression correlation change of each gene pair between conditions. A large variety of differential coexpression analysis methods have been developed, such as the Log Ratio of Connections<sup>11</sup>, Average Specific Connection<sup>12</sup>, Weighted Gene Coexpression Network Analysis (WGCNA)<sup>13</sup>, Differential Coexpression profile (DCp)<sup>14</sup>, Differential Coexpression enrichment (DCe)<sup>14</sup>, Differential Correlation in Expression for meta-module Recovery (DICER)<sup>15</sup>, and DiffCoEx<sup>16</sup>. For a review on the methods for differential coexpression networking, see Ref. 17. Differential coexpression analysis has been applied to successfully identify the differential coexpression modules, a group of genes strongly correlated under one condition but not the other<sup>17–20</sup>, and differentially expressed genes. Moreover, differential coexpression may indicate the rewiring of transcriptional networks in response to disease or adaptation to different environments. In brief, differential coexpression network (DCEN) can provide a more informative picture of the dynamic changes in gene regulatory networks.

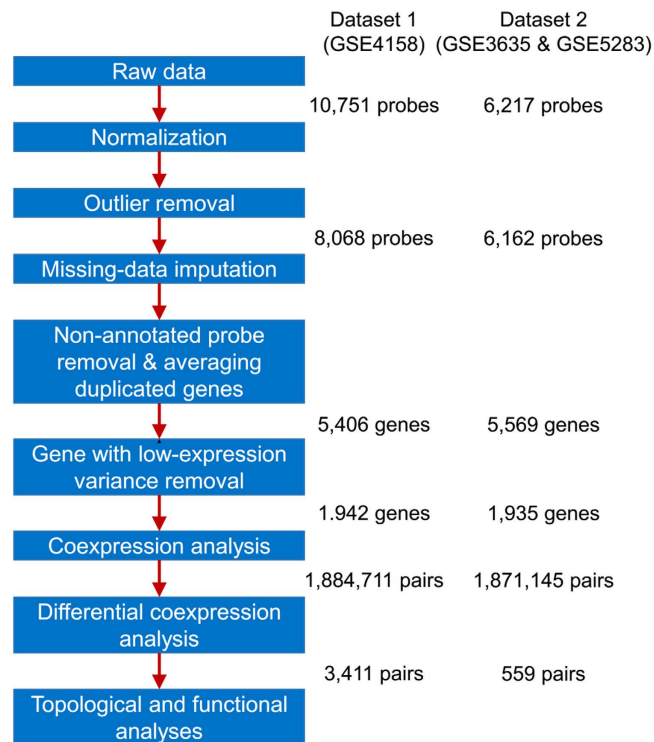
Previous studies with differential coexpression analyses mainly emphasized looking for differentially expressed genes and differential coexpression modules rather than whole networks<sup>17–20</sup>. In this study, we investigated the structural characteristics and biological significance of DCEN and used time-course gene expression data to construct coexpression networks and obtain DCEN by comparing the networks between two biological conditions. Several network structural features were quantified to investigate the common and unique properties between DCEN and other differential networks. Then, we incorporated other information to interpret the biological significance of DCENs. Finally, we offered a computational method to identify differential activation of transcription factors inferred from DCENs.

## Results and Discussion

**Construction of differential coexpression networks.** We constructed differential coexpression networks (DCENs) by using time-course gene expression data. Since biological systems are dynamic, temporal profiles of gene expression during a given biological process can often provide more insight about how genes are dependent on each other in a given biological process. To find the common properties of DCENs, we employed two distinct large-scale time-course gene expression datasets from Gene Expression Omnibus (GEO)<sup>21</sup>. The first dataset (Dataset 1, GEO accession GSE4158) was designed for understanding the dynamics of transcriptional response to changing environments by administering two different pulses (0.2 g/l and 2.0 g/l) of glucose on steady-state cultures of *Saccharomyces cerevisiae*<sup>22</sup>. The gene expression profiles over 12 (including 2, 4, 6, 8, 10, 15, 20, 30, 45, 90, 120, and 150 minutes) and 14 time points (including 3, 5, 7, 10, 15, 20, 30, 45, 90, 120, 150, 180, 210 and 240 minutes) were measured after addition of 0.2 g/l and 2.0 g/l glucose into cells, respectively. For the second data set (Dataset 2, GEO accession GSE3635 and GSE5283), the gene expression profiling of wild-type and strains with deleted YOX1 and YHP1 was performed to understand the regulation of transcription factor YOX1 and YHP1 during the cell cycle of *Saccharomyces cerevisiae*<sup>23</sup>. The wild-type and mutant cells were collected at 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, and 120 minutes after synchronization with alpha factor.

The gene expression data were pre-processed using the following steps: normalization, outlier removal, missing-data imputation, non-annotated probe removal and averaging duplicated genes, as well as removal of genes with low-expression variance (Fig. 1). These steps not only discarded non-informative genes but also reduced the computing time of coexpression analysis. Subsequently, we applied the method proposed by Wally *et al.*<sup>20</sup> with slight modification to deduce differentially coexpressed links (DCELs), which were further assembled into DCENs. The abbreviations DCEN1 and DCEN2 were used to denote the DCEN constructed using Dataset 1 and Dataset 2, respectively. DCEN1 represents the coexpression changes of cells that were treated with low (0.2 g/l) to high (2.0 g/l) glucose concentration, and DCEN2 reveals coexpression changes in cells with and without YOX and YHP1 mutations. DCEN1 consisted of 1425 genes and 3411 DCELs with 1795 positive and 1616 negative links (Table S1), whereas DCEN2 consisted of 611 genes and 559 DCELs with 284 positive and 275 negative links (Table S2). These two DCENs were used for further analyses.

Additionally, gene pairs with highly correlated expression patterns (Spearman correlation coefficient >0.95) under both conditions were considered as constitutively coexpressed links (CCEL), and the network assembled from CCELs was termed the constitutive coexpression network (CCEN). The CCEN from Dataset 1 (CCEN1) contained 569 genes and 1545 links, and that from Dataset 2 (CCEN2) contained 262 genes and 570 links.



**Figure 1. Overview of differential coexpression analysis pipeline.** The order of analysis is shown for the data pre-processing and differential coexpression analysis. The values depicted to the right are the number of available data points remaining after a given procedure.

**Structural characteristics of differential coexpression networks.** To unravel whether the structural properties of a DCEN are similar to those of other biological networks, the following quantities were measured in this study: average degree ( $\langle k \rangle$ ), maximum degree (Max.  $k$ ), exponent of the degree distribution ( $\gamma$ ), average clustering coefficient ( $C$ ), diameter ( $D$ ), and average shortest path length ( $L$ ). To clarify whether the observed structural properties are unique for DCENs or common for differential networks, topological analysis on other types of differential networks is necessary. Therefore, we collected a differential genetic interaction network (DGIN) that arises when cells are challenged by DNA damage<sup>8</sup> for comparison with the DCENs. Moreover, because the DCEN1 and DCEN2 could be broken down into 22 and 94 connected components, respectively (Figure S1), the corresponding largest components were used for the topology measurements.

The structural features of all differential networks are summarized in Table 1 and Fig. 2. Although the edge and node counts of DCEN1 are larger than those of DCEN2, the number of positive and negative differential links in each network is balanced. The degree distributions of differential networks suggest that they possess the scale-free properties<sup>24</sup> with their exponent of degree distribution ( $\gamma$ ) in the range between 2 and 4 (Fig. 2A and Table 1). The DGIN might have an intrinsic hierarchical structure as the average clustering coefficient of node decreases when the node degree increases<sup>24</sup> (Fig. 2B). However, the average clustering coefficients of DCENs are quite low. To clarify whether the low average clustering coefficient is due to their specific degree distributions, we generated the background distribution of average clustering coefficients from  $10^5$  degree-preserving random networks. Surprisingly, the average clustering coefficient of DCEN was significantly lower than the background distribution ( $P$ -value  $< 10^{-4}$  for both DCEN1 and DCEN2) in contrast to that of DGIN (Fig. 2C). Moreover, the average shortest path length of DCEN was greater than that of DGIN (Fig. 2D).

The comparison of these structural properties between differential networks indicates that DCENs differ substantially from the DGIN and other well-known biological networks, such as protein-protein interaction networks<sup>24</sup>. Although DCEN possesses the scale-free properties, DCEN is a tree-like network due to its low average clustering coefficient and high average shortest path length. These unique properties of DCEN might be inherited from those of coexpression networks. Coexpression network can be considered as a signed network that consists of positive (correlation) and negative (anti-correlation) links. Since correlations have transitive characteristics, two genes with common coexpressed and/or anti-coexpressed gene are expected to express simultaneously<sup>25</sup>. In terms of triad, which consists of three mutually linked nodes and is the smallest unit of a complete graph, a signed network generally contains four types of triads<sup>25</sup>. However, the transitive property of coexpression links indicates that only two types of triads can be observed in the coexpression network, that is, triads with an odd number of positive

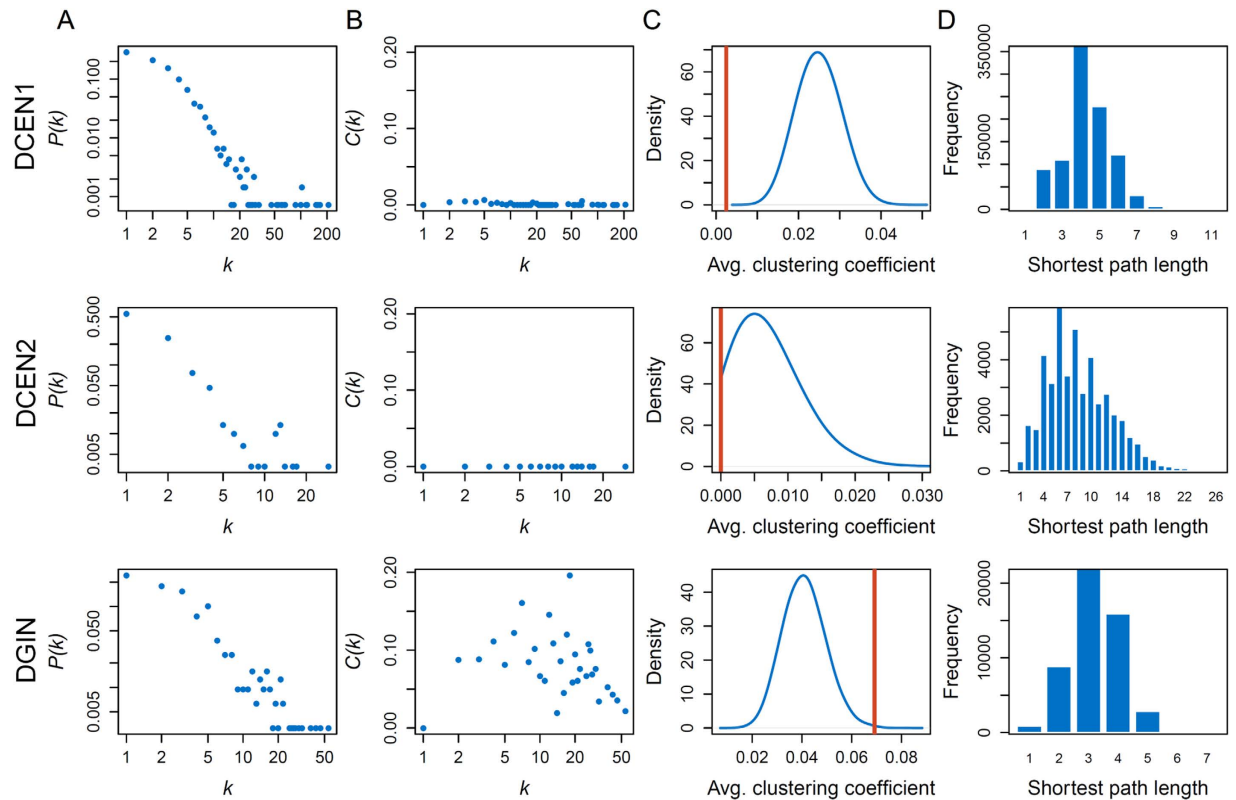
Quantity	DCEN 1: 0.2 g/l glucose → 2.0 g/l glucose	DCEN 2: Wild-type → deletions of YOX1 & YHP1	DGIN: No treatment → MMS treatment
Nodes	1,380 (1,425)	300 (611)	318
Edges	3,387 (3,411)	341 (559)	873
Positive differential	1,779 (1,795)	184 (284)	494
Negative differential	1,608 (1,616)	157 (275)	379
$\langle k \rangle$	4.91	2.27	5.49
Max. $k$	208	29	54
$\gamma$	2.43	2.44	3.4
$C$	0.0025	0	0.069
$D$	11	26	7
$L$	4.31	8.50	3.23

**Table 1. Characteristics of differential networks.** The numbers in parentheses are the corresponding quantity values of the whole network.  $\langle k \rangle$ , average degree; Max.  $k$ , maximum degree;  $C$ , average clustering coefficient;  $\gamma$ , exponent of degree distribution;  $D$ , diameter of a network;  $L$ , average shortest path length.

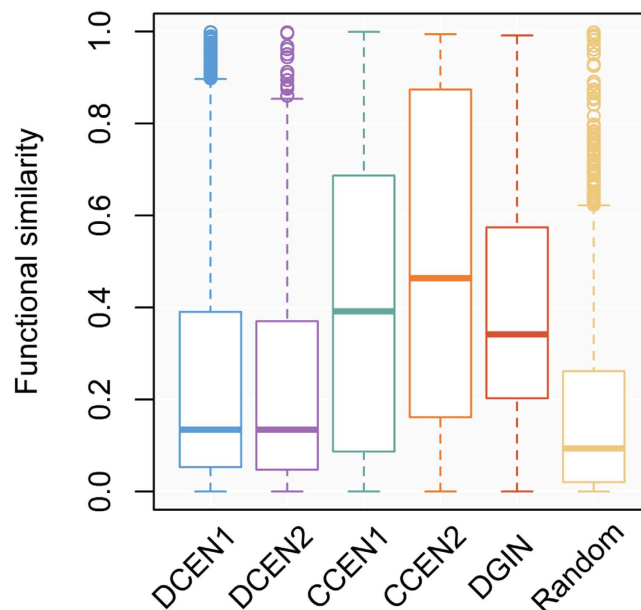
links<sup>25</sup>. Therefore, in a comparison between two coexpression networks, only two out of three links in a triad will reveal significant differences (Figure S2), and this property could result in the structure of DCEN as a tree-like network. To further confirm this characteristic of DCEN, we performed differential coexpression analysis on randomized gene expression data between conditions and counted the number of triads in the resultant random DCEN. Surprisingly, the proportion of observed triads in real DCEN is significantly lower than that in random DCEN ( $P$ -value  $< 10^{-5}$ , Figure S3). We also found that strong coexpression triads tend to be coexpressed in another condition (Figure S4). That is, genes rarely form triads in DCENs. Additionally, we used a different method, DCE<sup>14</sup>, to identify differentially coexpressed links and re-analyze the topological properties of the constructed DCEN. As shown in Figure S5, the networks constructed by DCE also have very small average clustering coefficients and high average shortest path length, consistent with those observed by our method. These results supported our speculation that the tree-like structure of DCEN might be caused by the intrinsic characteristics of coexpression networks. Here, we mainly focused on the observed tree structure of DCEN and would uncover its corresponding biological significance and functional interpretation. In the future, it will be interesting to systematically investigate additional topological properties and further compare DCEN structure with various types of networks<sup>26,27</sup>.

**Interpretations of differential coexpression networks.** Biological networks have been comprehensively used to investigate the molecular mechanisms underlying a specific biological process. Therefore, we questioned whether DCEN could reveal the molecular mechanisms in response to the change of biological conditions. We performed functional enrichment analysis on the components of DCEN, and found that the significantly enriched functions are indeed associated with the experimental observations. As DCEN1 was deduced from the expression data in response to changes in glucose concentrations, the functions of DCEN1 components were related to several metabolic processes (Table S3). DCEN2 was derived from the gene expression profiles of a yeast strain with mutations on YOX1 and YHP1, which are important transcription factors in the regulation of the cell cycle. Accordingly, many components of DCEN2 were found to be cell cycle-related proteins (Table S4). These results imply that the differential coexpression network can help reveal the underlying mechanisms during the change of biological conditions.

Next, we would like to interpret the biological significance of DCEs. Because many biological processes are achieved via protein-protein interactions (PPIs), we first examined these DCEs in a PPI database. After PPI searching in the BioGRID database<sup>28</sup>, surprisingly, only 10 links of DCEN1 and no links of DCEN2 were supported by PPI. This result reveals a limited relationship between DCEs and PPIs. We then examined whether two genes with differential coexpression are involved in similar biological processes. Gene Ontology (GO) semantic similarity was used to quantify the functional association of linked genes. Because coexpression patterns and genetic interactions have been used as evidence for gene function annotation by GO, the annotations with evidence codes of “inferred from expression pattern (IEP)” and “inferred from genetic interaction (IGI)” were discarded to avoid annotation bias. Unsurprisingly, links in CCEN tend to have highly functional relationships. However, most of the links in the DCEN and DGIN were not functional relationships (Fig. 3). Based on this observation, we presumed that the corresponding genes of a differential co-expressed link might not directly or indirectly interact in the same biological pathway, but contribute to different biological processes. This assumption is similar to the functional interpretations of genetic interactions in previous studies<sup>28,29</sup>. Genetic interactions can typically be interpreted by the between-pathway mode, in which the genetic interaction of bride genes

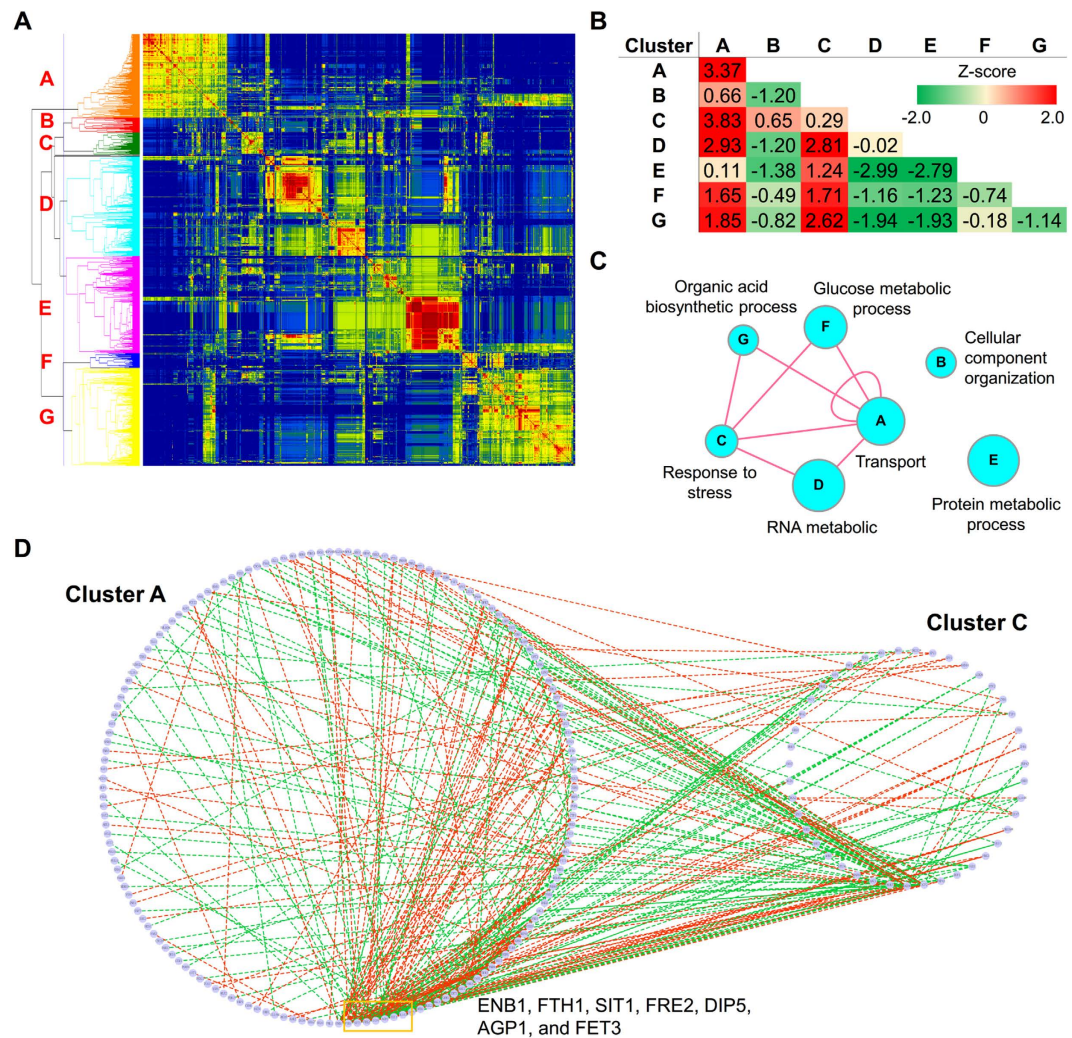


**Figure 2. Topological properties of differential networks.** (A) Degree distributions of differential networks.  $P(k)$  is the number of genes with  $k$  links. (B) Clustering coefficient versus degree for differential networks.  $C(k)$  denotes the average clustering coefficient of all genes with  $k$  links. (C) Distributions of average clustering coefficient for random networks (blue line). The red line denotes the average clustering coefficient of the given differential networks. (D) Distributions of shortest path lengths for differential networks.



**Figure 3. Distributions of functional similarities between genes.** The functional similarities between genes were calculated based on GO semantic similarity. The CCEN and DGIN were used as positive controls because coexpression and genetic interaction were supported evidences in GO annotation. The random sets (Random) were negative controls. DCEN, differential coexpression network; CCEN, constitutive coexpression network; DGIN, differential genetic interaction network.



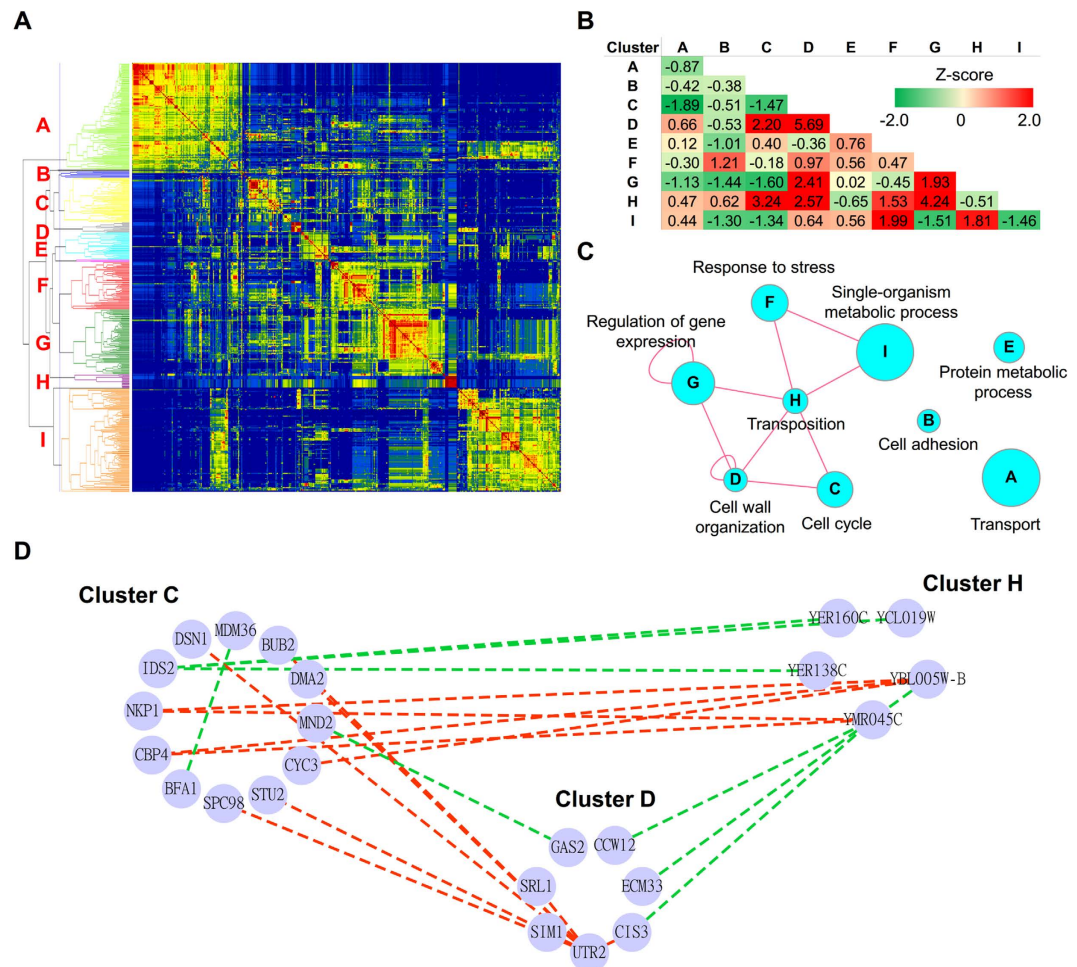


**Figure 4. Functional clusters and their relationships of differential coexpression networks in response to glucose stress.** (A) The heat map presents the functional similarity matrix among the components of a given DCEN. The red and blue colors represent high and low functional similarity, respectively. Based on the hierarchical tree, the components were divided into seven clusters (cluster A–G). (B) The table presents the z-transformed linking densities within- and between-clusters. (C) The interaction map of clusters was generated according to (B). The circle node denotes the cluster, and the edge represents the z-transformed linking density of linked clusters that was equal to or greater than 1.5. (D) The subnetwork of the given DCEN contains the interactions between cluster A and C. The circle node denotes the genes and the edge indicates the linked genes are differentially coexpressed. The red and green edges represent the positive and negative differential coexpression, respectively. Nodes in the given clusters are not shown if no edge linked to them.

operate in two pathways, or the within-pathway mode, in which the genetic interaction occurs between proteins within a single pathway<sup>29,30</sup>, and a high proportion of genetic interactions are associated with the between-pathway mode<sup>31</sup>.

To address the between-pathway characteristics, we proposed a computational procedure to construct the relationships within and between pathways. The components of DCENs were classified into functional clusters, which may be considered as specific biological processes or pathways, and the connections within and between clusters were assessed. These functional clusters can assist us in interpreting the biological significance of a DCEN.

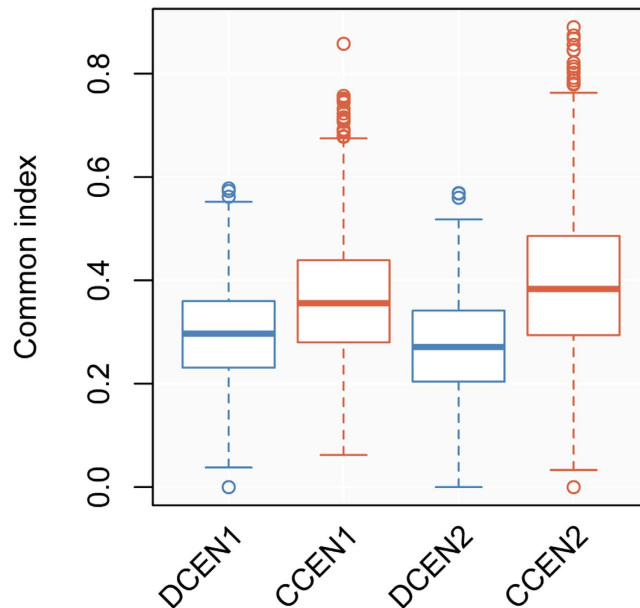
In DCEN1, 1,269 components contained GO annotations and were used to generate a functional similarity profile. After hierarchical clustering, the components of DCEN1 were classified into seven clusters (Fig. 4A), and all clusters were significantly enriched for particular biological processes (Table S5). Based on these clusters, 2,300 and 510 DCEs belonged to between- and within-cluster types, respectively, and this ratio was significantly higher than that expected by chance ( $P$ -value = 0.0017). Then, the cluster relationships were examined, and we subsequently observed that genes in cluster A not only



**Figure 5. Functional clusters and their relationships of differential coexpression networks in response to deletions of YOX1 and YHP1.** (A) The heat map presents the functional similarity matrix among components of a given DCEN. The red and blue colors represent high and low functional similarity, respectively. Based on the hierarchical tree, the components were divided into nine clusters (cluster A-I). (B) The table shows the z-transformed linking densities within- and between-clusters. (C) The interaction map of clusters was generated according to (B). The circle node denotes the cluster, and the edge represents the z-transformed linking density of linked clusters that was equal to or greater than 1.5. (D) The subnetwork of the given DCEN contains the interactions between cluster A and C. The circle node denotes the genes and the edge denotes the linked genes are differentially coexpressed. The red and green edges represent the positive and negative differential coexpression, respectively. The nodes in the given clusters are not shown if no edge linked to them.

had high intra-interactions, but also strongly interacted with genes in other clusters, especially cluster C (Fig. 4B,C). Additionally, cluster C tended to interact with other clusters. Genes in cluster C were involved in the response to stress, which is consistent with previous studies that a change in the carbon source causes dramatic stress to yeast<sup>22,32</sup>. However, we were interested in the interaction between cluster A and C. Most genes in cluster A were related to molecular transport processes, and a subset of genes in cluster A connected with most genes in cluster C (Fig. 4D). We found these genes with high connections, including ENB1, FTH1, SIT1, FRE2 and FET3, are involved in the iron uptake pathway<sup>33-35</sup>. This observation may imply that iron uptake is required in response to carbon stress. Indeed, previous studies have indicated that a change in the carbon source can simultaneously induce the stress response and iron uptake pathways<sup>33,34</sup>.

In DCEN2, 518 components were classified into nine clusters (Fig. 5A), and a high proportion of DCEs belonged to the between-cluster mode (362 and 66 for between- and within-cluster modes, respectively,  $P$ -value = 0.0025). All clusters had significantly enriched biological processes (Table S6). Several clusters densely interacted with other clusters (Fig. 5B,C). Because the absence of YOX1 and YHP1 results in dysregulation of the cell cycle process, we focused on the interactions with cluster C, whose components were related to the cell cycle. Cluster C had significant connections with cluster D



**Figure 6. The distribution of common transcription factors binding to differentially or constitutively coexpressed genes.** The common index is the arithmetic mean of the Jaccard and overlap coefficients. Suppose that we are given two gene lists that are bound by the set of transcription factors A and B, respectively. The Jaccard coefficient is defined as  $|A \cap B| / |A \cup B|$  and the overlap coefficient is defined as  $|A \cap B| / \max(|A|, |B|)$ . DCEN, differential coexpression network; CCEN, constitutive coexpression network.

and H, whose components are related to cell wall organization and transposition, respectively. Previous studies indicated that YOX1 and YHP1 bind the promoters of genes involved in cell wall synthesis<sup>36</sup>, and many effectors of the cell wall integrity signaling pathway are active through the cell cycle<sup>37,38</sup>. Interestingly, the components of cluster H were all gag-pol fusion proteins that can regulate their own translation. However, the relationships between gag-pol proteins and cell cycle and cell wall organization are still unclear.

These results suggest that the differential coexpressed links typically span multiple pathways instead of occurring within a single pathway. Moreover, the linked pathways have clear interdependent functional relationships. One reason for the preference towards between-pathway types may be that these mechanisms increase the efficiency in response to environmental or genetic changes, or provide synergistic effects.

**Inferring differential activation of transcription factors.** Finally, we would like to determine what mechanisms cause the differential coexpression phenomena. Transcription factors (TFs) are one of the major regulators in transcriptional control of gene expressions. TFs may coordinate by forming a complex, compete for promoter occupancy, or play antagonistic regulatory roles<sup>39</sup>. First, we examined the relationship between TFs and DCEs. Counting the number of common TFs between any pair of genes showed that CCEs tend to be regulated by more common TFs (Fig. 6), but DCEs do not. This suggests that gene pairs with more common regulators are robust to perturbations and tend to be coexpressed across various conditions, whereas gene pairs with less common regulators easily reveal differential coexpression when one of regulators are dysfunctional. Based on this observation, we presumed that the activated or deactivated transcription factors induced by environmental stress or genetic change could be identified from DCENs. A similar idea appeared in another study<sup>40</sup>.

We have taken a computational approach to infer differential activation of TFs inferred from DCENs. Our approach is based on the assumption that a TF is likely to be activated or deactivated across two conditions if significantly more differential coexpressed genes bound by this TF are observed than expected by chance. A total of 31 and 24 TFs revealing differential activation ( $P$ -value  $< 0.05$ ) were identified from DCEN1 and DCEN2, respectively (Tables 2 and 3). Interrogating the expression correlation among targets of each TF, we found that the distributions of most TFs differed significantly between two conditions (Figure S6 and Figure S7). This might be alternative evidence to demonstrate that these TFs gain or lose activity under different perturbations. Moreover, many TFs have been known to be related to processes in response to the corresponding stress, and we discuss this in detail in the following two paragraphs.

Some TFs derived from DCNE1 are activators or repressors in the carbon-source metabolic pathways (Table 2). For example, ADR1 regulates genes involved in utilization of non-fermentable carbon sources<sup>41</sup> whereas CAT8 regulates gluconeogenic genes<sup>42</sup>. However, both ADR1 and CAT8 are known to synergistically act for strong derepression of target genes, such as ADH2 and ACS1<sup>43</sup>, which are also



Transcription factor	# co-occurrence	P-value	Transcription factor	# co-occurrence	P-value
PDR1	641	<0.00001	MAC1	106	0.00539
SFP1	2624	<0.00001	AFT2	100	0.00675
AFT1	566	0.00007	OPI1*	73	0.00859
STB1	73	0.0001	UME6	215	0.00896
PDR3	384	0.00011	GZF3	45	0.01054
SWI5*	875	0.0003	MIG3*	475	0.01172
ADR1*	255	0.00049	CAT8*	24	0.01176
HAP3	47	0.00078	GCN4	1405	0.01394
MSN4	1034	0.00086	YAP1	1309	0.02051
ZAP1	548	0.00182	RIM101	190	0.02827
RTG3	180	0.00205	RME1	54	0.03382
DAL81	62	0.00308	HSF1	614	0.04566
ROX1*	248	0.00316	SMP1	26	0.04606
SOK2*	1033	0.00342	OAF1*	198	0.04678
ACE2*	2638	0.00413	GAT1	18	0.04729
MET4	582	0.00454			

**Table 2. Differential activation of transcription factors induced by glucose stress.** \*TFs are also the components of differential coexpression network.

Transcription factor	# co-occurrence	P-value	Transcription factor	# co-occurrence	P-value
YOX1	86	0.00001	SWI4	43	0.0128
STE12	288	0.00011	ROX1	33	0.01608
RIM101	39	0.00015	FKH1*	25	0.01667
CIN5	138	0.00019	FKH2*	17	0.01867
SPT23	98	0.00077	SOK2*	128	0.02385
YHP1*	64	0.00145	GCN4	195	0.02526
IXR1*	82	0.00179	KAR4*	25	0.03046
TEC1	262	0.00346	INO2	12	0.04265
SWI6*	14	0.00498	SUM1	15	0.04679
MCM1	82	0.00856	FLO8	18	0.04835
PHD1	26	0.01135	SPT3	42	0.04891
CRZ1	15	0.0116	MBP1	18	0.04923

**Table 3. Differential activation of transcription factors induced by deletions of YOX1 and YHP1.** \*TFs are also the components of differential coexpression network.

the components of DCEN1. YAP1 is a transcription activator involved in oxidative stress response and accumulates in the nucleus in response to carbon stress<sup>44</sup>. The function of MIG3 is still unclear, but a recent study showed that MIG3 may be related to the glucose-signaling network<sup>45</sup>. Additionally, several TFs are associated with metal cation uptake, which is consistent with the observation in the functional cluster analysis (Fig. 4C). AFT1 and AFT2 are the iron responsive transcriptional activators that regulate a series of genes involved in cell surface iron uptake (FET3, FRE1, FRE2, and FRE3), siderophore uptake (ARN1, FIT1, and FIT3), and iron transport across the vacuole membrane (FET5 and FTH1)<sup>46–48</sup>. Moreover, AFT1 can interact with other TFs to regulate transport of other metals. For example, AFT1 and MAC1 might corporately regulate transcription of CTR2 to mediate the mobilization of vacuolar copper stores in yeast<sup>49</sup>. Other TFs, such as HSF1, regulate genes in response to stress or heat shock<sup>50</sup>.

A third of the differential activation of TFs inferred from DCEN2 is yeast cell-cycle transcription factors (Table 3). Among these TFs, YOX1 and YHP1 were respectively ranked as the top 1 and 5 TFs according to their *P*-values. This result may also demonstrate the ability of our approach for inferring differential activation of TFs. Moreover, MCM1, SWI4, SWI6, MBP1, FKH1, and FKH2 are involved in activating gene expression during the cell cycle in yeast. YOX1 and YHP1 act in concert with MCM1 to

confer M/G1-specificity to the early cell cycle box (ECB) elements<sup>23</sup>. SBF, a heterodimer of SWI4 and SWI6, and MBF, a heterodimer of MBP1 and SWI6, are sequence-specific transcription factors that activate gene expression during the G1/S transition of the cell cycle in yeast. SBF binds to the so-called SCB (Swi4,6-dependent cell cycle box) promoter elements found upstream of the cyclin genes and cell wall biosynthetic genes, while MBF binds to a distinct element called the MCB (MluI cell cycle box) found mostly upstream of DNA replication and repair genes<sup>51–54</sup>. FKH1 and FKH2, forkhead transcription factors, assemble into ternary complexes with MCM1 to control transcription required for M-phase<sup>55</sup>. However, YOX1 competes with FKH2 for binding to MCM1 through protein-protein interactions at promoters of a subset of MCM1-regulated genes<sup>56</sup>.

Identifying the regulators that are relevant or even causative to a phenotypic change is a challenging goal. This problem cannot be solved using traditional differential expression analysis because a causal regulator is not necessarily differentially expressed. However, the causal regulators might be captured from DCENs. For example, a previous study that used a differential wiring analysis of expression data succeeded in identifying the gene containing the causal mutation<sup>57</sup>. Our method successfully identified several transcription factors, which have been demonstrated to be associated with corresponding phenotypes. We also found that many TFs are not the components of DCENs (Tables 2 and 3). This indicates that the expression of these TFs is not dramatically perturbed across the conditions, but the activity of these TFs might still be influenced. Interestingly, the functions in which TFs are involved are consistent with the results of functional clustering analysis (Figs 4C and 5C). In brief, we demonstrated the ability of DCENs in understanding the dynamics and regulatory mechanisms of cellular systems.

## Conclusions

In this work, we uncovered the unique structural characteristics and biological significance of differential coexpression networks (DCENs). DCENs resemble a tree-like network due to the intrinsic properties of coexpression networks. Although they possess the scale-free property, DCENs have a lower average clustering coefficient and higher average shortest path length in contrast to other biological networks. Furthermore, we proposed new approaches to the interpretation of DCEN. Our analysis found that differentially coexpressed genes tend to participate in different pathways. Pathways linked by differentially coexpressed genes may play complementary functions or have synergistic effects. We integrated transcription factor information with DCEN to reveal the regulatory mechanisms inducing differential coexpression. In brief, these results demonstrate that DCENs provide insights into the gene regulatory dynamics in response to various stresses. Additionally, the computational procedures proposed in this study can be applied to other coexpression networks, such as disease mechanism studies.

## Methods

**Gene expression data pre-processing.** The raw expression data were obtained from Gene Expression Omnibus (GEO)<sup>21</sup> and processed using the following steps. First, the gene expression data were normalized using a quantile normalization method to ensure a similar empirical distribution of each array. Second, probes with more than 20% of missing time points for a condition were discarded. Third, because the missing data were not allowed in the correlation analysis, the remaining missing data were imputed using the  $k$ -nearest neighbors (KNN) technique. In this step, we used the KNN algorithm implemented in the R package “impute” with  $k = 10$  to impute the missing values. Fourth, the probes without annotation were removed from further analysis, and if the probes were derived from the same gene, the expression intensity of this gene was represented by the average intensity of these probes. Finally, the genes with low variance across all time points were ignored for further analysis. Because these genes were constitutively expressed during all time points and were not perturbed under any condition, they might be non-informative for a coexpression network. We calculated the standard deviation of each gene to represent the perturbation of this gene under a condition, and if the standard deviation of a gene in both conditions were not in the top 25% of all genes, the gene was removed.

**Construction of differential coexpression network.** We modified the method proposed by Wally *et al.*<sup>20</sup> to identify differentially coexpressed links (DCELs). The difference of coexpression of gene  $u$  and  $v$  between two conditions  $a$  and  $b$  was quantified by the following formula,

$$D_{a \rightarrow b}^{u,v} = (R_b^{u,v} - R_a^{u,v}) * \max(|R_a^{u,v}|, |R_b^{u,v}|) \quad (1)$$

where  $R_a^{u,v}$  is the correlation between gene  $u$  and  $v$  at condition  $a$ . The latter term of this equation (i.e.,  $\max(|R_a^{u,v}|, |R_b^{u,v}|)$ ) was used to distinguish the importance when the values of the former term were identical. For example, the correlation of two pairs of genes between two conditions ranges from 1.0 to 0.5 and from 0.5 to 0.0. Although the correlation difference of these pairs is  $-0.5$ , the former case might be more informative than the latter. Under the definition of equation (1), positive aggregated scores mark the cases in which the correlation in condition  $b$  is higher than in condition  $a$ ; we refer to this situation as a “positively differentially coexpressed link”. Negative aggregated scores indicate a lower correlation between the pair in condition  $b$ , referred to as a “negatively differentially coexpressed link”. We calculated the correlation between genes using the Spearman correlation coefficient.

Significant differences were evaluated using permutation testing with different resample schemes chosen according to the dependency of the two samples. After 100,000 randomizations, empirical  $P$ -values were computed as the proportion of the difference observed in the permuted data set that was equal to or greater than that observed in the original data. Then, empirical  $P$ -values were adjusted using the Benjamini-Hochberg (BH) procedure to account for multiple hypothesis testing. The gene pairs with  $P$ -values  $< 10^{-4}$  were considered as DCEs, and then these DCEs were assembled into a DCEN.

**Differential genetic interaction network.** To compare the differences between various differential networks, we obtained a differential genetic interaction network (DGIN), derived from the comparison of two genetic interaction networks with and without perturbation by the DNA-damaging agent methyl methanesulfonate (MMS)<sup>8</sup>. This study identified 78,841 genetic interactions and 873 differential genetic interactions with  $P$ -value  $< 0.001$  covering 318 genes. A total of 379 interactions were “negatively differential”, which indicates DNA damage-induced lethality or sickness, whereas 494 were “positively differential”, which indicates inducible epistasis or suppression.

**Topological properties.** To investigate the structural properties of differential networks, the following network topologies were quantified.

- (1) The degree  $k_i$  of a vertex indicates the number of other vertices connected with this vertex. The average degree  $\langle k \rangle$  is the overall mean  $k_i$  of network vertices. The degree distribution ( $P(k)$ ) is defined as the fraction of vertices in the network having degree  $k$ . Most of the real networks display a power law shaped degree distribution  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  is a constant usually between 1 and 3. We estimated the exponent  $\gamma$  of each network using the Python package powerlaw<sup>58</sup>.
- (2) The shortest path length  $d_i$  of two vertices represents the number of edges along the shortest path connecting them. The average shortest path length ( $L$ ) and diameter ( $D$ ) are, respectively, defined as the mean and maximum  $d_i$  across all vertex pairs in the network.
- (3) The clustering coefficient  $c_i$  of a vertex is defined as the ratio between the number of connections existing between its neighbors and the maximal number of edges that can exist between them. The average clustering coefficient ( $C$ ) and the degree-dependent clustering coefficient ( $C(k)$ ) are the mean  $c_i$  across all network vertices and the overall vertices with degree  $k$ , respectively.

**Gene Ontology enrichment analysis.** The GO terms were obtained from the GO website (<http://www.geneontology.org/>), and the GO annotations for all yeast genes were downloaded from NCBI Entrez Gene. The enrichment analysis was performed using a hypergeometric test. Suppose that we are given a test set with  $n$  genes of which  $k$  genes belong to a certain GO term  $g$ , and a reference set with  $N$  genes of which  $M$  genes belong to  $g$ . The probability of  $g$  can be calculated according to the following formula.

$$P = \sum_{y=k}^n \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}} \quad (2)$$

Then, the adjustment for multiple comparisons was undertaken using the BH procedure. We only considered the “Biological Process” ontology in this work.

**Functional semantic similarity between genes.** The functional similarity between genes was measured by the semantic similarity between sets of GO terms with which they were annotated. We applied the method proposed by Schlicker *et al.*<sup>59</sup> to quantify the functional similarity. Schlicker’s measurement method combines Resnik’s method, which uses the concept of “information content” to define a semantic similarity<sup>60</sup>, and Lin’s method, which defines the similarity between two terms as the ratio of the commonality of the terms and the information needed to fully describe the terms<sup>61</sup>.

The first step in the comparison of two genes is the pairwise comparison of their GO mappings. Considering two genes A and B annotated with the sets of GO terms with sizes  $N$  and  $M$ , respectively, a similarity matrix  $S$  is calculated. This matrix contains all pairwise semantic similarity values between GO terms of gene A and GO terms of gene B. For the  $i$ -th GO term gene A,  $GO_i^A$ , and the  $j$ -th GO term of gene B,  $GO_j^B$ , the semantic similarity score  $s_{ij}$  of Schlicker’s method is defined as:

$$s_{ij} = \max_{c \in (GO_i^A, GO_j^B)} \left( \frac{2 \log p(c)}{\log p(GO_i^A) + \log p(GO_j^B)} \times (1 - p(c)) \right) \quad (3)$$

where  $c$  is the set of common ancestors of term  $GO_i^A$  and  $GO_j^B$ , and  $p(c)$  denotes the probability of the term  $c$  that is equal to its frequency in the annotations. The functional similarity of gene A ( $G^A$ ) and B ( $G^B$ ),  $funSim$ , is calculated using the best-match average (BMA) of the matrix  $S$ :

$$\text{funSim}(G^A, G^B) = \frac{\frac{1}{N} \sum_{i=1}^N \max_{1 \leq j \leq M} (s_{ij}) + \frac{1}{M} \sum_{j=1}^M \max_{1 \leq i \leq N} (s_{ij})}{2} \quad (4)$$

**Identification of functional clusters and densities among clusters.** Genes in DCEN were clustered based on their functional similarity profiles. In brief, the functional similarity between any given pair of genes was computed using GO semantic similarity as mentioned previously and a functional similarity matrix was formed. Hierarchical agglomerative average-linkage clustering with the Pearson correlation coefficient as the distance metric was applied to the functional similarity matrix. Generalized Association Plots (GAP)<sup>62</sup> were used to perform hierarchical clustering analysis and to view the results. The distance thresholds were determined based on the visual inspection of the heatmap and hierarchical trees.

The connectivity densities of DCEs ( $D_{ij}$ ) within and between clusters were calculated based on following formula:

$$D_{ij} = \begin{cases} \frac{I}{C_2^{n_i}}, & \text{if } i = j \\ \frac{I}{n_i * n_j}, & \text{if } i \neq j \end{cases} \quad (7)$$

where  $I$  is the number of DCEs within a single cluster of genes (if  $i = j$ ) or connecting two clusters (if  $i \neq j$ ), and  $n_i$  is the number of genes in the cluster  $i$ . Because the density is difficult to interpret, the density was transformed into a z-score. To calculate the z-score, the background distribution of connectivity density was estimated by randomizing the gene-cluster association and re-calculating the connectivity density. After 100,000 randomizations, the mean ( $\mu$ ) and standard deviation ( $\rho$ ) of the background distribution were obtained, and the z-score of density between cluster  $i$  and  $j$  ( $z_{ij}$ ) is given by the formula:

$$z_{ij} = \frac{D_{ij} - \mu_{ij}}{\rho_{ij}} \quad (8)$$

As the z-score  $> 1.5$ , it indicates that the connections within a given cluster or between the given clusters is significantly greater than that expected by chance.

We also used the same randomization procedure to estimate the significance of the ratio of between-cluster to within-cluster DCEs. The empirical  $P$ -value was calculated by counting the number of permutations in which the ratio of between-cluster to within-cluster DCEs was greater than or equal to the observed ratio, and then divided that number by the total number of permutations.

**Identification of differential activation of transcriptional factors.** To investigate the regulators of differentially coexpressed genes, the relationships between transcriptional factors (TFs) and regulated genes in budding yeast were obtained from YEASTRACT<sup>63</sup>. A total of 191,902 TF-gene regulations were available, including 315 transcriptional factors and 5,963 genes.

We proposed a computational approach to identify differential activation of transcription factors from DCENs. For each TF, the number of DCEs in which both genes were regulated by the given TF was counted. Next, the background distribution of the co-occurrence count of each TF was estimated by a randomization procedure as follows. For each component of a DCEN, the number of TFs regulating the given gene was counted, and a gene that was regulated by the same or a similar number of TFs was randomly selected from all yeast genes and used to replace the given gene. Once all components of a DCEN were randomized, the number of co-occurrence on both genes for all TFs was counted. This randomization step was repeated 100,000 times. The empirical  $P$ -value of a TF was calculated by counting the fraction of permutations in which the random co-occurrence count was higher than or equal to the observed value, and then corrected using the BH method to control the false discovery rate. The TFs with  $P$ -value  $< 0.05$  were considered as differentially activated.

## References

- Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–98 (2011).
- Mitra, K., Carvunis, A. R., Ramesh, S. K. & Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* **14**, 719–32 (2013).
- Harrold, J. M., Ramanathan, M. & Mager, D. E. Network-based approaches in drug discovery and early development. *Clin Pharmacol Ther* **94**, 651–8 (2013).
- Robin, X. *et al.* Personalized network-based treatments in oncology. *Clin Pharmacol Ther* **94**, 646–50 (2013).
- Chen, B., Fan, W., Liu, J. & Wu, F. X. Identifying protein complexes and functional modules—from static PPI networks to dynamic PPI networks. *Brief Bioinform* **15**, 177–94 (2014).
- Ideker, T. & Krogan, N. J. Differential network biology. *Mol Syst Biol* **8**, 565 (2012).
- Bisson, N. *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol* **29**, 653–8 (2011).



8. Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385–9 (2010).
9. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–55 (2003).
10. Wyrick, J. J. & Young, R. A. Deciphering gene expression regulatory networks. *Curr Opin Genet Dev* **12**, 130–6 (2002).
11. Reverter, A. *et al.* Simultaneous identification of differential gene expression and connectivity in inflammation, adipogenesis and cancer. *Bioinformatics* **22**, 2396–404 (2006).
12. Choi, J. K., Yu, U., Yoo, O. J. & Kim, S. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* **21**, 4348–55 (2005).
13. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).
14. Yu, H. *et al.* Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC Bioinformatics* **12**, 315 (2011).
15. Amar, D., Safer, H. & Shamir, R. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol* **9**, e1002955 (2013).
16. Tesson, B. M., Breitling, R. & Jansen, R. C. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* **11**, 497 (2010).
17. de la Fuente, A. From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends Genet* **26**, 326–33 (2010).
18. Southworth, L. K., Owen, A. B. & Kim, S. K. Aging mice show a decreasing correlation of gene expression within genetic modules. *PLoS Genet* **5**, e1000776 (2009).
19. van Nas, A. *et al.* Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* **150**, 1235–49 (2009).
20. Walley, A. J. *et al.* Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. *Int J Obes (Lond)* **36**, 137–47 (2012).
21. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–10 (2002).
22. Ronen, M. & Botstein, D. Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. *Proc Natl Acad Sci USA* **103**, 389–94 (2006).
23. Pramila, T., Miles, S., GuhaThakurta, D., Jemiolo, D. & Breeden, L. L. Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. *Genes Dev* **16**, 3034–45 (2002).
24. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–13 (2004).
25. Lin, C. C., Lee, C. H., Fuh, C. S., Juan, H. F. & Huang, H. C. Link clustering reveals structural characteristics and biological contexts in signed molecular networks. *PLoS One* **8**, e67089 (2013).
26. Filkov, V., Saul, Z. M., Roy, S., D'Souza, R. M. & Devanbu, P. T. Modeling and verifying a broad array of network properties. *Europhys Lett* **86**, 28003 (2009).
27. Roy, S. & Filkov, V. Strong associations between microbe phenotypes and their network architecture. *Phys Rev E* **80**, 040902 (R) (2009).
28. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535–9 (2006).
29. Tucker, C. L. & Fields, S. Lethal combinations. *Nat Genet* **35**, 204–5 (2003).
30. Costanzo, M., Baryshnikova, A., Myers, C. L., Andrews, B. & Boone, C. Charting the genetic interaction map of a cell. *Curr Opin Biotechnol* **22**, 66–74 (2011).
31. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nat Biotechnol* **23**, 561–6 (2005).
32. Jona, G., Choder, M. & Gileadi, O. Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochim Biophys Acta* **1491**, 37–48 (2000).
33. Haurie, V., Boucherie, H. & Sagliocco, F. The Snf1 protein kinase controls the induction of genes of the iron uptake pathway at the diauxic shift in *Saccharomyces cerevisiae*. *J Biol Chem* **278**, 45391–6 (2003).
34. Rutherford, J. C. & Bird, A. J. Metal-responsive transcription factors that regulate iron, zinc, and copper homeostasis in eukaryotic cells. *Eukaryot Cell* **3**, 1–13 (2004).
35. Philpott, C. C., Protchenko, O., Kim, Y. W., Boretzky, Y. & Shakoury-Elizeh, M. The response to iron deprivation in *Saccharomyces cerevisiae*: expression of siderophore-based systems of iron uptake. *Biochem Soc Trans* **30**, 698–702 (2002).
36. Horak, C. E. *et al.* Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev* **16**, 3017–33 (2002).
37. Levin, D. E. Regulation of cell wall biogenesis in *Saccharomyces cerevisiae*: the cell wall integrity signaling pathway. *Genetics* **189**, 1145–75 (2011).
38. Levin, D. E. Cell wall integrity signaling in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **69**, 262–91 (2005).
39. Thiel, G., Lietz, M. & Hohl, M. How mammalian transcriptional repressors work. *Eur J Biochem* **271**, 2855–62 (2004).
40. Yang, J. *et al.* DCGL v2.0: an R package for unveiling differential regulation from differential co-expression. *PLoS One* **8**, e79729 (2013).
41. Young, E. T., Dombek, K. M., Tachibana, C. & Ideker, T. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J Biol Chem* **278**, 26146–58 (2003).
42. Haurie, V. *et al.* The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in *Saccharomyces cerevisiae*. *J Biol Chem* **276**, 76–85 (2001).
43. Kratzer, S. & Schuller, H. J. Transcriptional control of the yeast acetyl-CoA synthetase gene, ACS1, by the positive regulators CAT8 and ADR1 and the pleiotropic repressor UME6. *Mol Microbiol* **26**, 631–41 (1997).
44. Wiatrowski, H. A. & Carlson, M. Yap1 accumulates in the nucleus in response to carbon stress in *Saccharomyces cerevisiae*. *Eukaryot Cell* **2**, 19–26 (2003).
45. Lewis, J. A. & Gasch, A. P. Natural variation in the yeast glucose-signaling network reveals a new role for the Mig3p transcription factor. *G3 (Bethesda)* **2**, 1607–12 (2012).
46. Rutherford, J. C., Jaron, S. & Winge, D. R. Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements. *J Biol Chem* **278**, 27636–43 (2003).
47. Stadler, J. A. & Schweyen, R. J. The yeast iron regulon is induced upon cobalt stress and crucial for cobalt tolerance. *J Biol Chem* **277**, 39649–54 (2002).
48. Protchenko, O. *et al.* Three cell wall mannoproteins facilitate the uptake of iron in *Saccharomyces cerevisiae*. *J Biol Chem* **276**, 49244–50 (2001).
49. Qi, J., Han, A., Yang, Z. & Li, C. Metal-sensing transcription factors Mac1p and Aft1p coordinately regulate vacuolar copper transporter CTR2 in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* **423**, 424–8 (2012).
50. Hou, J., Osterlund, T., Liu, Z., Petranovic, D. & Nielsen, J. Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **97**, 3559–68 (2013).

51. Bean, J. M., Siggia, E. D. & Cross, F. R. High functional overlap between MluI cell-cycle box binding factor and Swi4/6 cell-cycle box binding factor in the G1/S transcriptional program in *Saccharomyces cerevisiae*. *Genetics* **171**, 49–61 (2005).
52. Koch, C., Moll, T., Neuberg, M., Ahorn, H. & Nasmyth, K. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* **261**, 1551–7 (1993).
53. Bastajian, N., Friesen, H. & Andrews, B. J. Bck2 acts through the MADS box protein Mcm1 to activate cell-cycle-regulated genes in budding yeast. *PLoS Genet* **9**, e1003507 (2013).
54. Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–8 (2001).
55. Kumar, R. *et al.* Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. *Curr Biol* **10**, 896–906 (2000).
56. Darieva, Z. *et al.* A competitive transcription factor binding mechanism determines the timing of late cell cycle-dependent gene expression. *Mol Cell* **38**, 29–40 (2010).
57. Hudson, N. J., Reverter, A. & Dalrymple, B. P. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* **5**, e1000382 (2009).
58. Alstott, J., Bullmore, E. & Plenz, D. powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *PLoS One* **9**, e85777 (2014).
59. Schlicker, A., Domingues, F. S., Rahnenfuhrer, J. & Lengauer, T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* **7**, 302 (2006).
60. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1* 448–453 (Morgan Kaufmann Publishers Inc., Montreal, Quebec, Canada, 1995).
61. Lin, D. An Information-Theoretic Definition of Similarity. in *Proceedings of the Fifteenth International Conference on Machine Learning* 296–304 (Morgan Kaufmann Publishers Inc., 1998).
62. Wu, H.-M., Tien, Y.-J. & Chen, C.-h. GAP: A graphical environment for matrix visualization and cluster analysis. *Comput Stat Data An* **54**, 767–778 (2010).
63. Teixeira, M. C. *et al.* The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **42**, D161–6 (2014).

## Acknowledgements

The authors would like to thank James Winkler (Texas A&M University) for English editing on the manuscript. This work was supported by Ministry of Science and Technology, Taiwan (102-2628-B-002-041-MY3, 102-2627-B-002-002, and 103-2320-B-010-031-MY3) and the National Taiwan University Cutting-Edge Steering Research Project (104R7602C3).

## Author Contributions

C.L.H. and H.C.H. conceived and designed the study. C.L.H. implemented the method and performed data analysis. C.L.H., H.C.H. and H.F.J. interpreted the data and wrote the manuscript. H.C.H. and H.F.J. supervised the study. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Hsu, C.-L. *et al.* Functional Analysis and Characterization of Differential Coexpression Networks. *Sci. Rep.* **5**, 13295; doi: 10.1038/srep13295 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>