

Link Clustering Reveals Structural Characteristics and Biological Contexts in Signed Molecular Networks

Chen-Ching Lin^{1,2}, Chia-Hsien Lee², Chiou-Shann Fuh², Hsueh-Fen Juan^{2,3*}, Hsuan-Cheng Huang^{1*}

1 Institute of Biomedical Informatics, Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei, Taiwan, **2** Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan, **3** Department of Life Science, Institute of Molecular and Cellular Biology, Center for Systems Biology, National Taiwan University, Taipei, Taiwan

Abstract

Many biological networks are signed molecular networks which consist of positive and negative links. To reveal the distinct features between links with different signs, we proposed signed link-clustering coefficients that assess the similarity of interaction profiles between linked molecules. We found that positive links tended to cluster together, while negative links usually behaved like bridges between positive clusters. Positive links with higher adhesiveness tended to share protein domains, be associated with protein-protein interactions and make intra-connections within protein complexes. Negative links that were more bridge-like tended to make interconnections between protein complexes. Utilizing the proposed measures to group positive links, we observed hierarchical modules that could be well characterized by functional annotations or known protein complexes. Our results imply that the proposed sign-specific measures can help reveal the network structural characteristics and the embedded biological contexts of signed links, as well as the functional organization of signed molecular networks.

Citation: Lin C-C, Lee C-H, Fuh C-S, Juan H-F, Huang H-C (2013) Link Clustering Reveals Structural Characteristics and Biological Contexts in Signed Molecular Networks. PLoS ONE 8(6): e67089. doi:10.1371/journal.pone.0067089

Editor: Eshel Ben-Jacob, Tel Aviv University, Israel

Received: March 10, 2013; **Accepted:** May 16, 2013; **Published:** June 24, 2013

Copyright: © 2013 Lin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the National Science Council of Taiwan (99-2621-B-010-001-MY3 and 99-2621-B-002-005-MY3), National Taiwan University Frontier an Innovative Research Projects (99R70437), and National Health Research Institutes (NHRI-EX101-9819PI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yukijuan@ntu.edu.tw (HFJ); hsuancheng@ym.edu.tw (HCH)

Introduction

Biological processes in living cells are usually accomplished by numerous interactions between biological molecules (genes, proteins and other cell components) at various scales. Therefore, molecular networks, which are comprised of biological molecules and interactions between them, can provide a comprehensive interpretation of complicated biological systems in living cells and have become a key approach to understanding biological systems [1–4]. Investigation of the network structure has been used to reveal biological contexts embedded in molecular and cellular networks [5–7]. For example, Lin *et al.* studied the complete graphs in protein-protein interaction networks, and identified the essential cores in protein networks of *Escherichia coli* and *Saccharomyces cerevisiae* [5]; Roth *et al.* used the minimum spanning trees to extract the most relevant information contained in the gene network of *Bacillus subtilis* [6]; Madi *et al.* also analyzed the minimum spanning trees in immune networks, and found different conservative level between mothers' and newborns' networks [7]. Link clustering denotes the overlap between neighboring links and has been used to identify communities in molecular and social networks [8,9]. Essentiality of a protein in the interaction network was found to be highly associated with the link clustering level of the interactions connecting it [10]. Moreover, Solava *et al.* utilized link clustering to predict new pathogen-interacting proteins which possibly play the role of drug target candidates [11].

Many molecular networks, such as the genetic interaction network (GIN) and the gene coexpression network (CEN), are

signed undirected networks that consist of positive and negative links (genetic interactions or gene coexpression). Genetic interactions (GIs) describe that double mutants confer a significant deviation of phenotype from the expected value [12]. This expected value of phenotype change is referred to as the combination effect of two single mutations [13]. Positive GIs are when the phenotypic changes of double mutants are equivalent to or less severe than expected, such as synthetic suppression or rescue. In contrast, negative GIs are when double mutants display a more severe phenotype than expected, such as synthetic lethality or sickness [14,15]. Genes with positive GIs have been referred to as alleviating or epistatic interactions, while those with negative GIs are usually thought to participate in parallel biological pathways. Thus, single mutants are compatible with continued viability, while the double mutants damage viability [14]. Previous studies have reported that genes with similar patterns of GI profiles tended to participate in the same biological pathways or processes [16,17]. Gene coexpressions (CEs) are measured by expression correlations between genes, usually measured by the Pearson correlation coefficient (PCC) or other metrics. The CEN collates correlated genes under well-designed experimental states. In CENs, simultaneously expressed gene-pairs form positive CEs, while inversely expressed pairs form negative CEs.

Herein, considering the essential differences between positive and negative links, we proposed four measures of link-clustering coefficients (LCs), which were used to evaluate the proportions of common interacting partners, also called neighbors, between linked molecules. By applying LCs to study the network structure

of a CEN, we found that positive links were more adhesive and tended to cluster together, while negative links were more dispersive and usually behaved like bridges between positive clusters. Interestingly, a similar network structure was also observed in the GIN. Additionally, the proposed LC could be further used to reveal hidden biological contexts of signed links and to uncover the network modules that are well characterized by functional annotations or known protein complexes.

Results

Coexpression Network (CEN)

Network structure of the CEN. Coexpression networks consist of gene pairs with similar or opposite gene expression profiles. Here, we defined coexpression as a positive link and anti-coexpression as a negative link, following the sign of the correlation coefficient between expression profiles. Since correlations had transmission characteristics (Figure S1), two genes with common coexpressed and/or anti-coexpressed genes in the CEN are expected to express simultaneously. It could lead CEN to possess specific network structural properties, such as distribution of triads – the smallest units of the complete graph. There are four possible types of triads according to the combinatorial patterns of the three interconnected signed links, denoted T_1 – T_4 in Figure 1A. The frequencies of each type of triad were assessed by the ratio of the observed number for each triad-type to the corresponding expected value from random shuffling of the signs of links (more details in Text S1). As expected, we observed that T_1 (+++) and T_2 (+--) were significantly over-represented, while T_3 (-++) and T_4 (---) were totally absent (Figure 1A and Table S1). In other words, positive CEs tend to cluster with co-positive or co-negative CE neighbors, while negative CEs tend to cluster with hybrid ones. This observation suggested that positive and negative CEs should have distinct clustering features. Thus, we applied an LC that measured the proportion of common neighbors between two linked nodes to assess the aggregation characteristics of links [9]. We first disregarded the signs of the interconnected links by the conventional LC definition, and found that the LC distribution of negative CEs was similar to that of positive CEs (Figure 1B). It suggested that both types of CEs could cluster with other CEs, but the difference between the clustering properties of positive and negative CEs were indistinguishable using unsigned LCs. To differentiate the clustering characteristics of positive and negative links, we took the signs of clustering links into consideration, dividing unsigned LCs into two sign-specific groups: Same (SLC), which considers only the neighboring links of the same signs and Hybrid (HLC), which considers neighboring links of opposite signs. We found that SLC of positive links, SLC(+), remained similar to unsigned LC(+) while that of negative links, SLC(-), were all zero (Figure 1C). On the other hand, HLC of negative links, HLC(-), remained similar to unsigned LC(-) while HLC(+) were all zero (Figure 1D). Apparently, clustering properties of positive and negative CEs can be distinguished by our proposed sign-specific LCs. According to the signs of paired links connecting to their common neighbors, SLC can be further divided into two subtypes, PLC (LCs with two positive signs) and NLC (LCs with two negative signs). Both of their distributions for positive links were similar to SLC(+) (Figure S2a). These results were consistent with the expected network structural characteristics of CEN. Additionally, we found that negative CEs with higher HLC tended to recruit common neighbors with higher PLC(+) and HLC(-) (Figure 1E, F). This suggested that positive CEs linking to the common neighbors that contributed to HLC(-) tended to form positive clusters and negative ones tended to connect to (other

positive cluster(s). In other words, we can infer that positive links are more adhesive and tend to cluster together while negative links are more dispersive and usually behave like bridges between positive clusters. Altogether, above results suggested that the proposed signed LC was capable of reflecting and even highlighting the structural characteristics of CEN.

Biological contexts in the CEN. The CEN was constructed by discretizing the correlations between expression profiles of gene pairs. Each individual link in the CEN preserved only the binary information of whether the two linked genes were coexpressed (for those with a significant positive correlation coefficient above a certain threshold) or anti-coexpressed (with a significant negative correlation coefficient). Although such a network representation seemingly diminishes the quantitative information of individual links, the quantitative correlation information was, in fact, embedded in the network structure and could be recovered to a certain extent. A pair of genes with highly correlated expression profiles was expected to share a larger amount of commonly linked genes, resulting in a higher SLC, as well as PLC and NLC. On the other hand, those with highly anti-correlated expression profiles were expected to share genes with opposite types of links, resulting in a higher HLC. Indeed, we observed a strong correlation between SLC and PCC for gene pairs with positive links (Figure 2A; see also Figure S2b for similar characteristics of PLC and NLC), and a strong anti-correlation between HLC and PCC for negative links (Figure 2B). Furthermore, the proportion of coregulated gene pairs increased along with SLC for positive links, but not for HLC of negative links (Figure 2C; see also Figure S2c for similar characteristics of PLC and NLC). In other words, the coexpressed gene pairs sharing more common coexpressed or anti-coexpressed partners tended to be regulated by the same transcription factors. Therefore, it suggested that sign-specific LCs could reveal the embedded quantitative magnitude of coexpression, as well as the biological contexts involved in the CEN.

Next, we applied the predefined similarity measure, which was derived from the summation of two same-sign LC subtypes, PLC and NLC, to cluster positive links for identification of potential functional modules (see Materials and Methods). Among 34 identified modules (size ≥ 3), we focused on the two largest modules, which covered 263 and 245 genes, respectively. Links inside these two modules are all positive, but those between them are all negative, which is consistent with the observed structural characteristics of the CEN. Positive links inside these two modules tended to have higher PLC and NLC (≥ 0.5 , Figure S2d,e), while negative links between modules tended to have a higher HLC (≥ 0.5 , Figure S2f). Again, positive CEs with high PLC or NLC are modular, while negative CEs with a high HLC are bridge-like. The gene expression profiles of these two modules displayed similar patterns inside modules, but were opposite to each other between modules under different conditions of nutrition sources, i.e., 1% ethanol and 2% glucose (Figure 2D). Notably, we chose these two modules only according to the proposed signed LC and their size. The enriched biological functions of these two modules were ribosome biogenesis and energy-production-related functions, respectively (Table S2 and S3). We noted that the large and small subunits of ribosome and 90S preribosome were involved in the largest module and that ATP synthase, cytochrome c oxidase, cytochrome c reductase and succinate dehydrogenase were involved in the second largest module (Figure 2E and Figure S2g). These well-known protein complexes are directly associated with ribosome biogenesis or energy production. Additionally, we observed that the largest module was activated by 2% glucose and repressed by 1% ethanol, in contrast to the second largest module,

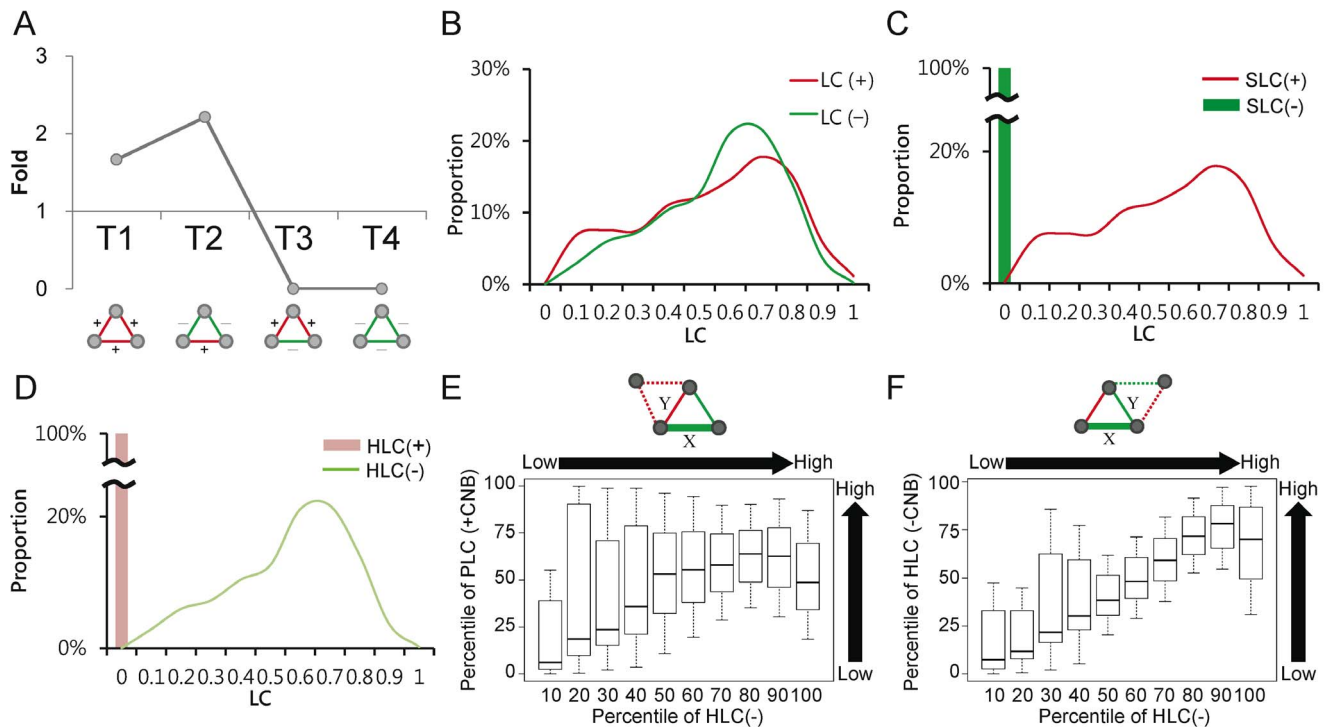


Figure 1. Structural properties of the CEN. (A) Frequency of signed triads in CEN. According to combinatorial patterns of signed links, four types of triads are listed. Fold is the ratio of observed number of triads to the average number of random triads. (B)–(D) LC, SLC, and HLC distributions of positive/negative links in the CEN. The values shown on the x-axis are the upper bounds of the corresponding LC intervals. (E) Median of PLC of positive CEs linking to the common neighbor (CNB) that contributed to the HLC of the observed negative CE with increasing HLC(-). The Pearson's correlation between PLC of positive CEs linking to the common neighbor and HLC(-) is 0.52 ($P < 2.2 \times 10^{-16}$). (F) Median of HLC of negative CEs linking to the common neighbor (CNB) that contributed to the HLC of the observed negative CE with increasing HLC(-). The Pearson's correlation between HLC of negative CEs linking to the common neighbor and HLC(-) is 0.64 ($P < 2.2 \times 10^{-16}$). doi:10.1371/journal.pone.0067089.g001

which behaved in the opposite manner (Figure 2D). It was reported that glucose can transcriptionally repress TCA cycle genes, decrease respiratory activity and activate ribosome protein genes as sufficient amounts of glucose are available to support cell growth [18]. On the other hand, under ethanol stress, yeast initially struggles to maintain energy production by increasing expression of genes associated with energy-generating activities and decreasing expression-rates of genes associated with energy-demanding processes, such as growth [19]. In summary, these results indicate that the proposed LC has potential to reveal the biological contexts of signed links and the functional modules, such as protein complexes, in signed molecular networks.

Genetic Interaction Network (GIN)

Network structure of the GIN. Unlike CEs, GIs didn't possess transitive property. In the GIN of *Saccharomyces cerevisiae*, we observed that four types of triads were present and only T₁ (+++) was significantly over-represented (Figure 3A and Table S1). This resembles the characteristics of CEN—i.e., that positive links tend to cluster with positive link neighbors—although the triads involving negative links behaved differently in the GIN. The unsigned LC distributions of positive GIs also showed a higher tail than negative ones (Figure 3B). To resolve what kind of aggregation forms positive GIs prefer, we analyzed the signed LC of GIs. SLC(+) distributed toward higher coefficients than the other three types of signed LC, HLC(+), SLC(-), and HLC(-) (Figure 3C). Additionally, PLC(+) and NLC(+) also distributed with higher tails than PLC(-) and NLC(-), respectively (Figure 3D). These results suggest that the following: (1) positive

GIs tend to form clusters with co-positive or co-negative GI neighbors rather than with hybrid GI neighbors; (2) compared with positive GIs, negative GIs disfavor clustering. To clarify the characteristics of negative links in GIN, we interrogated the clustering characteristics of the hybrid links contributing to HLC(-). We found that positive GIs of hybrid links involved in HLC(-) tended to form positive clusters (Figure 3E). Furthermore, HLC of negative GIs in HLC(-) hybrid links positively correlated with the observed HLC(-) (Figure 3F). These findings suggested that the negative GIs with high HLC tended to act as bridges between positive clusters. Although the triad and LC distributions of the GIN differ from the CEN, they share similar features, i.e., that positive GIs tend to cluster together and negative GIs usually behave like bridges between positive clusters.

Biological contexts embedded in genetic interaction links. Genetic interactions are measured by the phenotypic change of perturbed living cells, and hence are thought to make functional connections within and/or between biological processes [3,17,20–22]. Previous studies have reported that positive GIs tend to appear between gene pairs with protein-protein interactions (PPIs) or participate in the same protein complex, while negative GIs tend to be interconnections between different protein complexes [23–27]. Herein, we observed that positive GIs with higher SLCs tended to be PPIs or intra-connections within the same protein complex (Figure 4A, B), and negative ones with higher HLCs tended to be interconnections between different protein complexes (Figure 4C). We also found that proteins encoded by genes that formed positive GIs with higher SLCs tended to share the same protein domains (Figure 4D). These

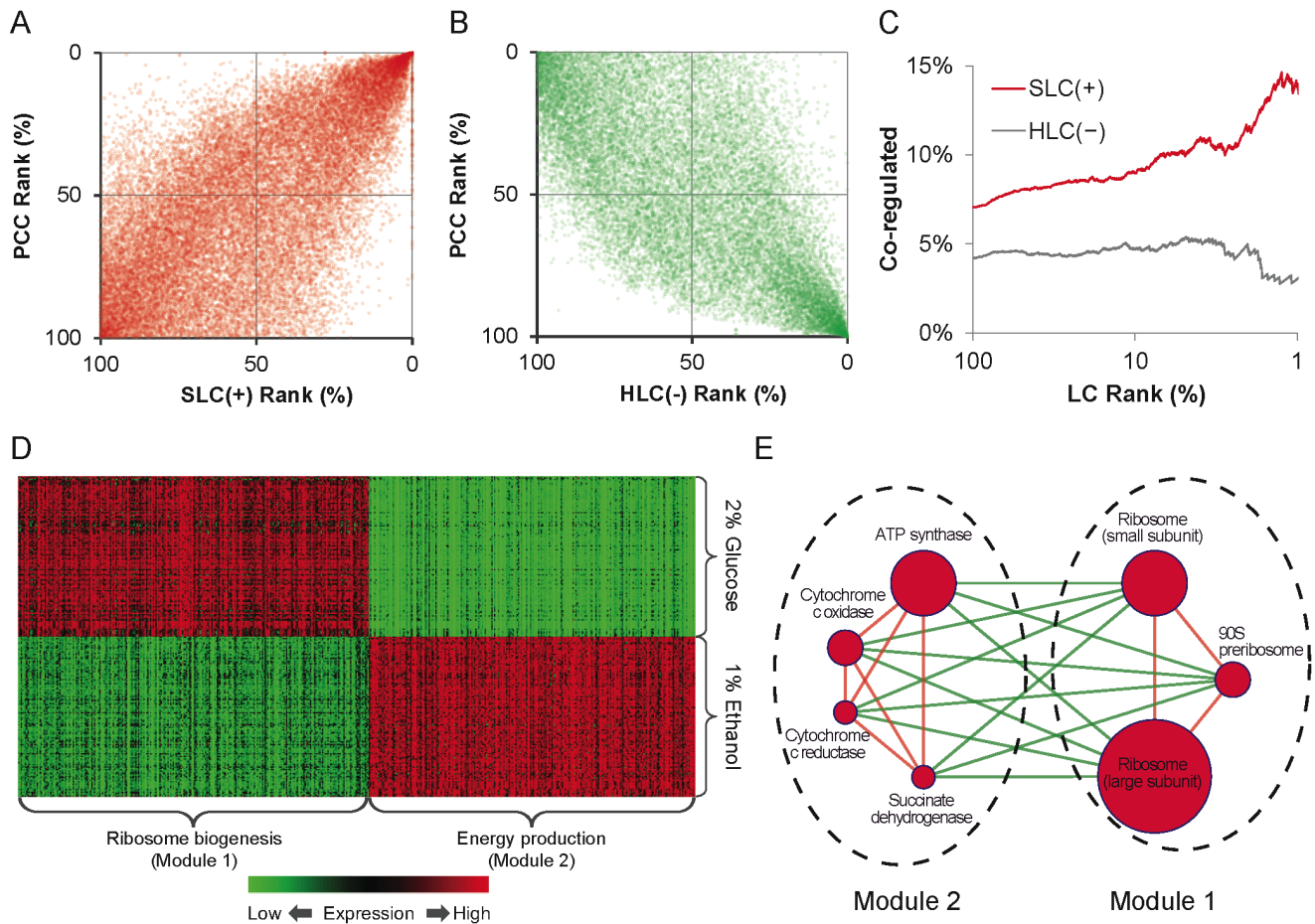


Figure 2. Biological contexts embedded in the CEN. Rank 0% and 100% represent the highest and lowest value of corresponding measurement, respectively. (A)(B) Two positive (negative) CE genes with higher PLC (HLC) tended to coexpress (anti-coexpress) with each other more. (C) Two positive CE genes with higher PLC or NLC tended to be regulated by the same transcription factors. (D) Expression profiles of the two largest functional modules. (E) Well-known protein complex inside selected two largest modules. Node size represents the number of genes covered by the corresponding sub-module. Node color represents the density of positive CEs involved in the sub-module. Red (green) links indicate that CEs between two sub-modules are all positive (negative). doi:10.1371/journal.pone.0067089.g002

observations suggest that positive GIs with higher SLCs could imply a stronger functional relationship or homogeneity between genetic interacting genes. SLC(+) and HLC(-) not only reflect the network topological properties—i.e., the intramodularity of positive GIs and bridgeness of negative GIs—but also help reveal the biomolecular complex structure and organization involved. For example, we found that several protein complexes were enriched by the positive GIs with the top 1% highest SLC (Figure 4F; $p < < 0.0001$, Fisher's exact test); 96% of the negative GIs among these complex subunit genes made interconnections between different complexes and were enriched in the top 1% highest HLC ($p < < 0.0001$, Fisher's exact test). Notably, all the PPIs among them were from positive GIs with the highest 1% SLC and intra-connections within the complex. Their shared protein domains were mostly related to proteasome subunits and prefoldin. On the other hand, previous studies have reported that negative GIs possibly reflect the evolutionary relationship between two genetic interacting genes [28–31]. Interestingly, we found that only gene pairs of negative GI with higher SLCs tended to be duplicated genes (Figure 4E). Since higher SLC implies potentially higher functional homogeneity, this observation might result from the

functional compensatory relationship between negative genetic interacting genes [31].

Genetic Interaction Modules

After investigating the structure and biological contexts of GIN, we noticed that positive GIs tended to form functionally homogeneous modules. To discover these modules inside the GIN, we applied single-linkage hierarchical clustering with the LC-based similarity score of positive GIs, ranked in descending order, and utilized partition density [8] to determine the similarity score cut-off of the optimal modular structure. As the cut-off of positive GIs increased, partition density was first elevated to a maximal value and then decreased (Figure S3a), implying that the positive GIN did contain a local modular structure. However, partition density was only decreased when the similarity score cut-off of negative GIs increased (Figure S3b), which implied that the negative GIN consisted of no local denser subnetworks. When the similarity score of positive GI that corresponded to the maximal partition density was applied, 33 positive modules that consisted of more than three genes were discovered (Figure 5). Indeed, over 90% of modules possessed highly intraconnected positive GIs (positive density ≥ 0.5) and almost 80% of them contained no

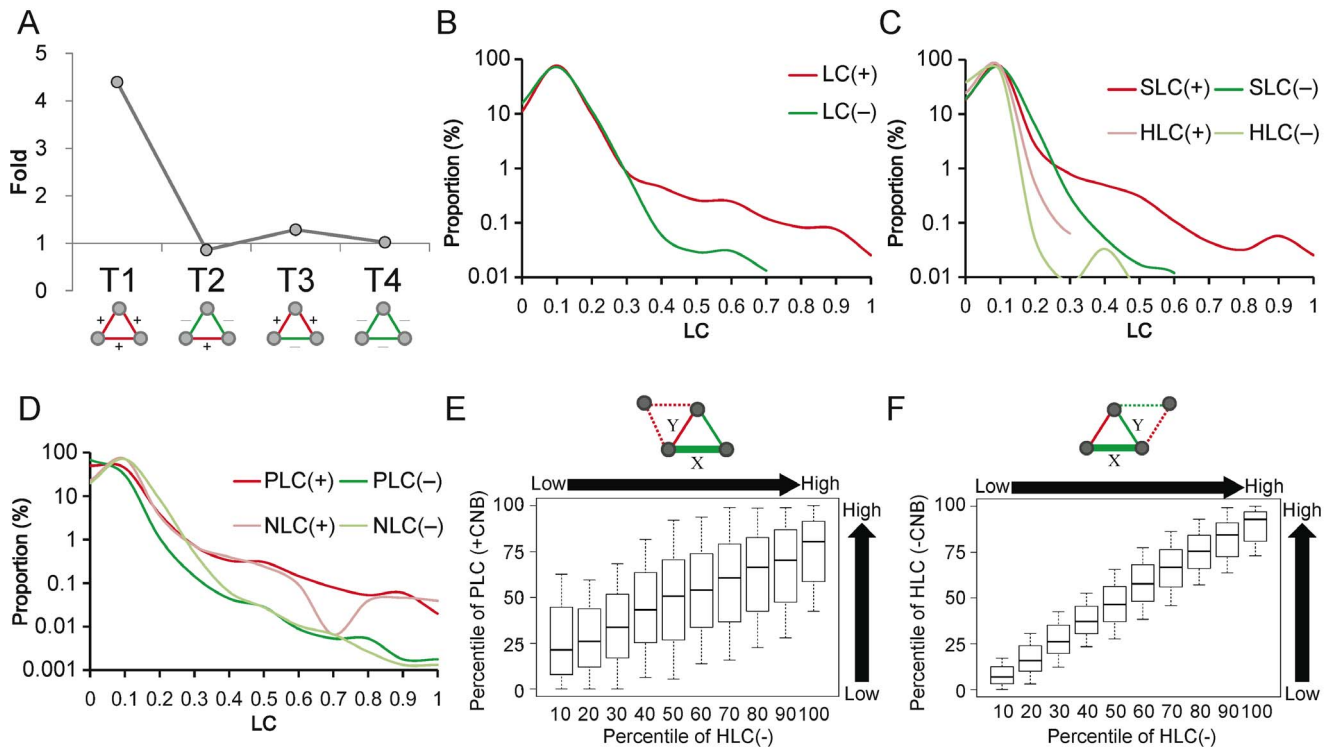


Figure 3. Structural properties of the GIN. (A) Frequency of signed triads in GIN. (B)–(D) LC, SLC, HLC, PLC, and NLC distributions of positive/negative links in GIN. (E) Pearson's correlation between PLC of positive GIs linking to the common neighbor (CNB) and HLC(-) is 0.57 ($P < 2.2 \times 10^{-16}$). (F) Pearson's correlation between HLC of negative GIs linking to the common neighbor (CNB) and HLC(-) is 0.88 ($P < 2.2 \times 10^{-16}$). doi:10.1371/journal.pone.0067089.g003

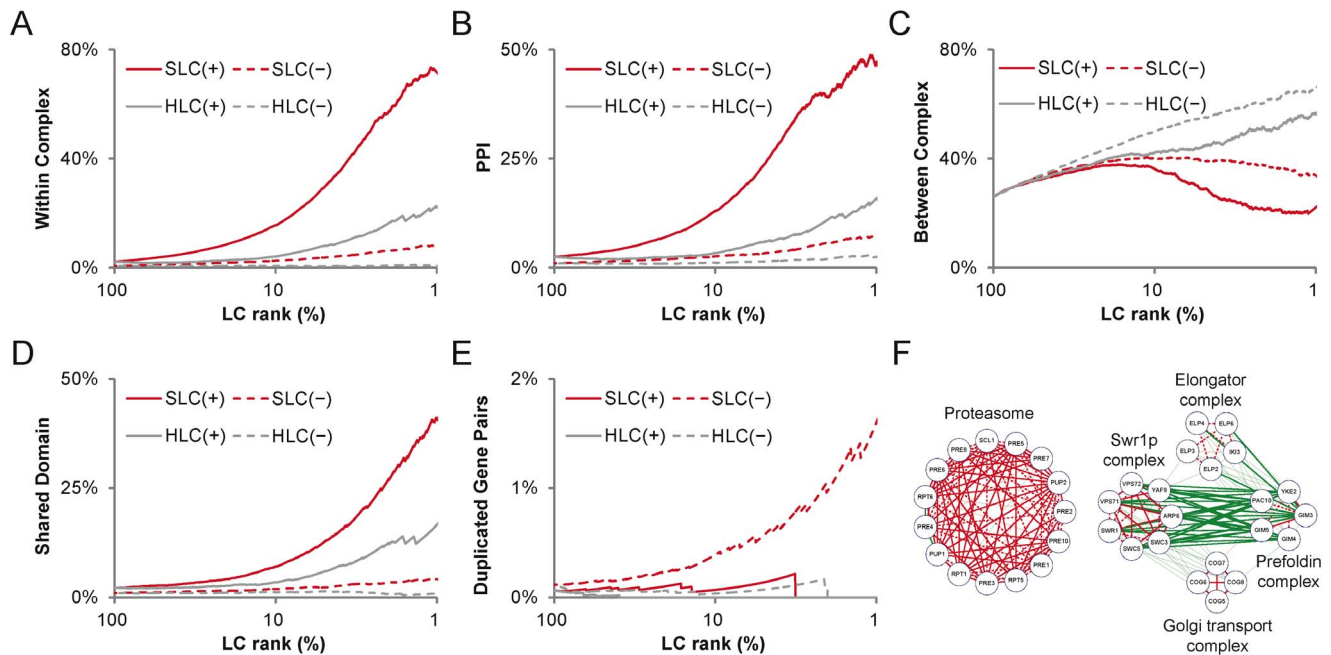


Figure 4. Biological contexts embedded in GIs. (A)–(E) The correlations between GI and biological contexts. The x-axis represents the percentage of ranked LC value, and the value 1 means the top 1% highest LC value. The y-axis represents accumulated proportions of GIs with corresponding biological contexts. (F) Example complexes formed by positive GIs with the top 1% highest SLC. Positive and negative GIs are represented by red and green links, respectively. Dashed links are PPIs. Bold links are GIs with the top 1% highest SLC or HLC, and thin ones are the other GIs. doi:10.1371/journal.pone.0067089.g004

negative GI (Figure S3c). 70% of link sets between modules only contained negative GIs (Figure S3d). Additionally, these modules could be well characterized by known protein complexes or biological processes (Figure 5). Some of the modules, such as “response to DNA damage stimulus” and “double-strand break repair”, have been reported to be synthetic lethal with each other [30]. More importantly, this implies that the gene-based GIN can be summarized as a module-based network by applying the LC-based similarity score to cluster positive GIs.

Discussion

In this study, we applied a rigorous threshold to define the coexpression links in a CEN, $|\text{PCC}| \geq 0.9$. Because of the transmission property of correlation, only type 1 and type 2 triads are allowed, while type 3 and type 4 are not. In GIN, all four types of triads were observed, which might imply that the GIN possessed a triad-enriched network structure. On the other hand, GIs measure the phenotypic relevance between genes and the changes of phenotypes often relate to complicated and numerous biological processes. Consequently, GIs are usually thought to be subtle and to underlie diverse biological contexts. Therefore, triple GIs that formed triads in GIN might easily be derived from different biological contexts, and thus they might not follow the transitory information.

According to the *structural balance theory* – proposed by Heider in the 1940s [32] and formulated by Cartwright and Harary in graph theory [33], type 1 and type 2 triads are balanced and type 3 and type 4 are unbalanced. Therefore, CEN is structurally balanced and follows the two structure theorems [33,34] summarized by Hummon and Doreian [35]: *A network is balanced if and only if the network can be divided into two or more subnetworks, wherein links in the same subnetwork are all positive and between different subnetworks are negative.* In the GIN, four types of triads were present, while only type 1 was significantly over-represented (Figure 1A and Table S1). This suggests that the GIN was weakly structurally balanced [34] and, thus, abates the requirement of T_2 over-representation. In summary, the signed molecular network is (weakly) structurally balanced and T_1 (three mutually positively linked genes) is significantly over-represented relative to chance.

In the proposed module map, one notable interaction is between “double-strand break repair” and “Swr1p complex”. In the double-strand break repair module, XRS2 and RAD50 are parts of the MRE11-RAD50-XRS2 (or MRX) complex, which plays a vital role in both homologous recombination (HR) repair and non-homologous end-joining (NHEJ) repair [36]. Additionally, RAD51, RAD52, RAD54, and RAD55 participate in the primary repair process [37]. The TOP3-RMI1-SGS1 complex is required to resolve the DNA intermediate structure, which is produced in the final steps of HR [38]. Genes in the “Swr1p

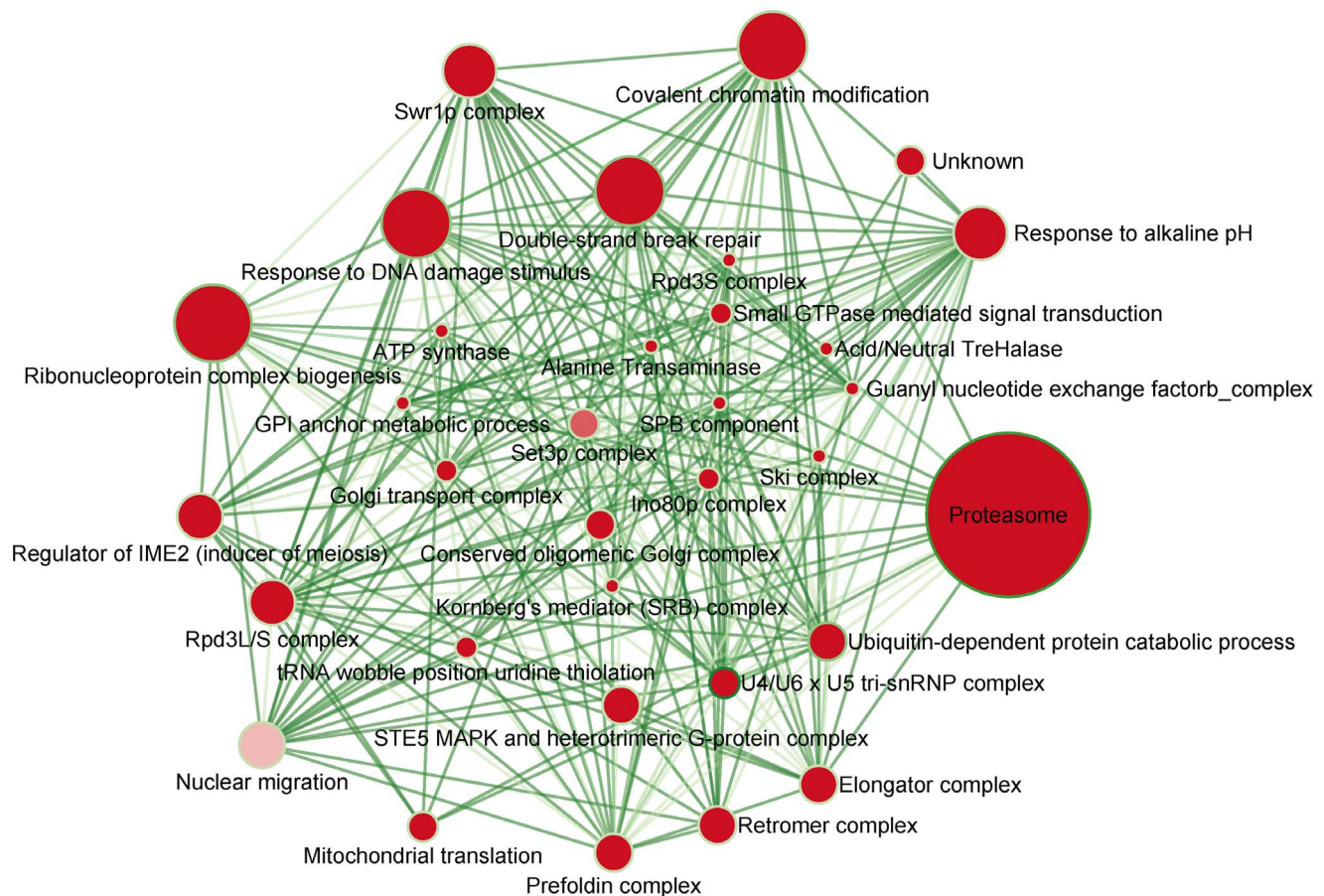


Figure 5. Map of genetic interaction modules. Each node represents a module clustered by positive GIs, and each edge represents a bunch of negative GIs between different modules. Node size indicates the number of genes in each module and node color intensity indicates the density of positive GI. Edge width indicates the number of GIs between modules and edge color intensity indicates the proportion of negative GI. Color intensity of node border indicates the density of negative GI in each module. doi:10.1371/journal.pone.0067089.g005

complex” module are part of the histone post-modification pathway [39]. In this pathway, H2BK123 is ubiquitinated by the Rad6-Bre1 complex [40]. The ubiquitination requires the presence of the Paf1 complex, which contains two subunits, RTF1 and CDC73, in this module [39]. After the ubiquitination of H2BK123, H3K4 is trimethylated by the Set1 complex, which contains four subunits, SWD1, SWD3, SDC1 and BRE2, in this module [41]. The H3K4 trimethylation is related to the NHEJ repair pathway [40]. In agreement with the balance structure of the signed molecular network, the density of positive GIs in these two modules are 0.5 and 0.7, respectively, and links between these two modules are almost completely negative (96%). As described above, genes in the double-strand break repair module are part of the HR repair pathway, and genes in the Swr1p complex module participate in NHEJ-related histone post-translational modifications. HR and NHEJ are two major DNA double-strand repair pathways of the yeast cell [42]. This suggests that these two modules participate in two different pathways with the same or similar output and, therefore, they should be able to complement each other.

In this study, we applied the signed LC to study the network structure of the signed molecular network and successfully revealed the differences of clustering characteristics between positive and negative links. The results showed that positive links tend to cluster together, while negative links are more dispersive and usually make interconnections between positive clusters. Furthermore, the signed LC facilitated the discovery of the diverse biological contexts covered by signed links and the functional modules within signed molecular networks.

Materials and Methods

Coexpression and Genetic Interaction Networks

To construct the CEN, we downloaded the expression profiles of yeast genes from Gene Expression Omnibus (GEO), accession number GSE9376 [43], containing 6,253 genes and 246 samples in various nutrition sources. The correlations between genes were evaluated by the Pearson correlation coefficient (PCC). To stress the correlations between genes, paired genes with $PCC \geq 0.9$ were defined as positive coexpression and those with $PCC \leq -0.9$ as negative. The studied CEN consisted of 1,240 genes and 48,497 coexpression links (28,651 positive and 19,846 negative).

The yeast genetic interactions were downloaded from BioGRID 3.1.72 [44]. We retrieved “synthetic rescue” and “positive genetic” relationships between genes as positive GIs, and “synthetic lethality” and “negative genetic” ones as negative GIs. In addition, 448 ambiguous GIs were removed from this dataset. After this filtration, 5,084 genes and 91,743 GIs (15,821 positive and 75,922 negative) were included in the yeast GIN.

We applied the algorithm proposed by Lin *et al.* [5] to identify and count the number of triads in the CEN and GIN.

Link-clustering Coefficient

Given a network composed of nodes and links connecting paired nodes (edges), the link-clustering coefficient (LC) measures the proportions of shared neighbors (common linking partners) between linked molecular pairs, and is defined as:

$$LC_{e(i,j)} = \frac{|n(i) \cap n(j)|}{|n(i) \cup n(j)|}$$

where $LC_{e(i,j)}$ is the LC of the link e formed by node i and j . Note that $n(i)$ ($n(j)$) is the excess neighbors of node i (j) excluding node j (i). Previous studies have noted that biological molecules would be

likely to share similar functions with their neighbors [45,46]. Thus, a higher LC means a larger proportion of shared neighbors and implies higher functional similarity between two interacting molecules. Herein, LC was calculated for positive and negative links in signed molecular network separately. $LC(+)$ and $LC(-)$ denoted the LCs of positive links and the LCs of negative links, respectively. Further, according to the signs of paired links connecting to the common neighbors, LC can be classified into two subtypes, same (SLC, $+/+$ or $-/-$) and hybrid (HLC, $-/+$ or $+/-$), which are defined as:

$$SLC_{e(i,j)} = \frac{|(n_+(i) \cap n_+(j)) \cup (n_-(i) \cap n_-(j))|}{|n(i) \cup n(j)|};$$

$$HLC_{e(i,j)} = \frac{|(n_+(i) \cap n_-(j)) \cup (n_-(i) \cap n_+(j))|}{|n(i) \cup n(j)|}$$

where $n_{+/-}(i)$ ($n_{+/-}(j)$) is the excess positive/negative neighbors of node i (j) excluding node j (i). Based on its definition, SLC can be further categorized into two subtypes, positive (PLC) and negative (NLC), which are defined as:

$$PLC_{e(i,j)} = \frac{|n_+(i) \cap n_+(j)|}{|n_+(i) \cup n_+(j)|}; \quad NLC_{e(i,j)} = \frac{|n_-(i) \cap n_-(j)|}{|n_-(i) \cup n_-(j)|}.$$

Revealing Biological Contexts and Communities in Signed Networks

Herein, several biological relationships between genes–PPI, within/between protein complex, shared protein domain and duplicated genes–were used to discover the embedded biological contexts of GIs (more details in Text S1) [47–51]. The proportions of biological contexts covered by positive/negative GIs were calculated and referred to as the relevance of biological contexts to GIs.

To discover the biological communities, single-linkage hierarchical clustering was applied with the similarity score defined as:

$$Sim_{LC} = \begin{cases} PLC(+), & \text{if } \exists PLC(+)\text{ and } \neg \exists NLC(+), \\ NLC(+), & \text{if } \neg \exists PLC(+)\text{ and } \exists NLC(+), \\ \frac{PLC(+) + NLC(+)}{2}, & \text{if } \exists PLC(+)\text{ and } \exists NLC(+), \\ NA, & \text{if } \neg \exists PLC(+)\text{ and } \neg \exists NLC(+). \end{cases}$$

The threshold for cutting this dendrogram to yield communities was determined by maximum partition density, which was introduced by Ahn *et al.* [8]. The potential biological processes of each community were investigated by functional enrichment analysis (more details in Text S1).

Supporting Information

Figure S1 Correlation transmission. Correlation transmission via common (a) co-expressed and/or (b) anti-expressed neighbors. (ZIP)

Figure S2 Biological context revealed by LC of CE. (a) PLC and NLC distributions of positive/negative links in CEN. The values shown on the x-axis are the upper bounds of the

corresponding LC intervals. (b) Two positive CE genes with higher PLC or NLC tended to have higher rates of coexpression with each other. Red points: PLC(+); Green points: NLC(+). (c) Two coexpressed genes that shared more common coexpressed (PLC) or anti-expressed (NLC) partners tended to be regulated by the same transcription factors. (d) – (f) LC distributions of the two largest modules. (g) Coexpression subnetworks of seven well-known protein complexes involved in the two largest modules. (ZIP)

Figure S3 LC-score vs. partition density of GIN and GI density of discovered modules. (a) LC-score vs. partition density of positive GIN. (b) LC-score vs. partition density of negative GIN. (c) Distributions of positive/negative GI density of discovered modules. (d) Distributions of negative GI proportion of meta-links between modules. (ZIP)

Table S1 Number of triads in CEN and GIN.

References

- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nature reviews Genetics* 5: 101–113.
- Charbonnier S, Gallego O, Gavin AC (2008) The social network of a cell: recent advances in interactome mapping. *Biotechnology annual review* 14: 1–28.
- Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C (2009) Systematic mapping of genetic interaction networks. *Annual review of genetics* 43: 601–625.
- Zhao W, Langfelder P, Fuller T, Dong J, Li A, et al. (2010) Weighted gene coexpression network analysis: state of the art. *Journal of biopharmaceutical statistics* 20: 281–300.
- Lin CC, Juan HF, Hsiang JT, Hwang YC, Mori H, et al. (2009) Essential core of protein-protein interaction network in *Escherichia coli*. *Journal of proteome research* 8: 1925–1931.
- Roth D, Madi A, Kenett DY, Ben-Jacob E (2011) Gene Network Holography of the Soil Bacterium *Bacillus subtilis*. In: Witzany G, editor. *Biocommunication in Soil Microorganisms*.
- Madi A, Kenett DY, Bransburg-Zabary S, Merbl Y, Quintana FJ, et al. (2011) Network theory analysis of antibody-antigen reactivity data: the immune trees at birth and adulthood. *PLoS One* 6: e17445.
- Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466: 761–764.
- Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D (2004) Defining and identifying communities in networks. *Proc Natl Acad Sci U S A* 101: 2658–2663.
- Wang J, Li M, Wang H, Pan Y (2012) Identification of essential proteins based on edge clustering coefficient. *IEEE/ACM Trans Comput Biol Bioinform* 9: 1070–1080.
- Solava RW, Michaels RP, Milenkovic T (2012) Graphlet-based edge clustering reveals pathogen-interacting proteins. *Bioinformatics* 28: i480–i486.
- Bateson W, Saunders ER, Punnett RC, CC H (1905) Reports to the Evolution Committee of the Royal Society. Report II. London: Harrison and Sons.
- Mani R, St Onge RP, Hartman JLT, Giaevar G, Roth FP (2008) Defining genetic interaction. *Proceedings of the National Academy of Sciences of the United States of America* 105: 3461–3466.
- Guarente L (1993) Synthetic enhancement in gene interaction: a genetic tool come of age. *Trends in genetics*: TIG 9: 362–366.
- Dobzhansky T (1946) Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of *Drosophila Pseudoobscura*. *Genetics* 31: 269–290.
- Tong AH, Lesage G, Bader GD, Ding H, Xu H, et al. (2004) Global mapping of the yeast genetic interaction network. *Science* 303: 808–813.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. (2010) The genetic landscape of a cell. *Science* 327: 425–431.
- Yin Z, Wilson S, Hauser NC, Tournu H, Hoheisel JD, et al. (2003) Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs. *Mol Microbiol* 48: 713–724.
- Stanley D, Bandara A, Fraser S, Chambers PJ, Stanley GA (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. *J Appl Microbiol* 109: 13–24.
- Magtanong L, Ho CH, Barker SL, Jiao W, Baryshnikova A, et al. (2011) Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nat Biotechnol* 29: 505–511.
- Sharifpoor S, van Dyk D, Costanzo M, Baryshnikova A, Friesen H, et al. (2012) Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. *Genome Res* 22: 791–801.
- Costanzo M, Baryshnikova A, Myers CL, Andrews B, Boone C (2011) Charting the genetic interaction map of a cell. *Current opinion in biotechnology* 22: 66–74.
- Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123: 507–519.
- St Onge RP, Mani R, Oh J, Proctor M, Fung E, et al. (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet* 39: 199–206.
- Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446: 806–810.
- Ulitsky I, Shlomi T, Kupiec M, Shamir R (2008) From E-MAPs to module maps: dissecting quantitative genetic interactions using physical interactions. *Mol Syst Biol* 4: 209.
- Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, et al. (2009) Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* 136: 952–963.
- Gurley KE, Kemp CJ (2001) Synthetic lethality between mutation in *Atm* and *DNA-PK(cs)* during murine embryogenesis. *Curr Biol* 11: 191–194.
- Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, et al. (2005) Targeting the DNA repair defect in *BRCA* mutant cells as a therapeutic strategy. *Nature* 434: 917–921.
- Pan X, Ye P, Yuan DS, Wang X, Bader JS, et al. (2006) A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* 124: 1069–1081.
- VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, et al. (2010) Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Mol Syst Biol* 6: 429.
- Heider F (1946) Attitudes and cognitive organization. *The Journal of Psychology* 21: 107–112.
- Cartwright D, Harary F (1956) Structural balance: a generalization of Heider's theory. *Psychological Review* 63: 277–293.
- Davis JA (1967) CLUSTERING AND STRUCTURAL BALANCE IN GRAPHS. *Human Relations* 20: 181–187.
- Hummel NP, Doreian P (2003) Some dynamics of social balance processes: bringing Heider back into balance theory. *Social Networks* 25: 17–49.
- Ataian Y, Krebs JE (2006) Five repair pathways in one context: chromatin modification during DNA repair. *Biochem Cell Biol* 84: 490–504.
- Sugawara N, Wang X, Haber JE (2003) In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol Cell* 12: 209–219.
- Mankouri HW, Hickson ID (2007) The RecQ helicase-topoisomerase III-Rrm1 complex: a DNA structure-specific 'dissolvosome'? *Trends Biochem Sci* 32: 538–546.
- Jachning JA (2010) The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* 1799: 379–388.
- Faucher D, Wellinger RJ (2010) Methylated H3K4, a transcription-associated histone modification, is involved in the DNA damage response pathway. *PLoS Genet* 6.
- Dehe PM, Geli V (2006) The multiple faces of Set1. *Biochem Cell Biol* 84: 536–548.
- Jackson SP (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* 23: 687–696.
- Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. *PLoS biology* 6: e83.
- Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. (2011) The BioGRID Interaction Database: 2011 update. *Nucleic acids research* 39: D698–704.
- Hishigaki H, Nakai K, Ono T, Tanigami A, Takagi T (2001) Assessment of prediction accuracy of protein function from protein-protein interaction data. *Yeast* 18: 523–531.

46. Schwikowski B, Uetz P, Fields S (2000) A network of protein-protein interactions in yeast. *Nature biotechnology* 18: 1257–1261.
47. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, et al. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic acids research* 32: D449–451.
48. Pu S, Wong J, Turner B, Cho E, Wodak SJ (2009) Up-to-date catalogues of yeast protein complexes. *Nucleic acids research* 37: 825–831.
49. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research* 38: D355–360.
50. Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of molecular biology* 313: 903–919.
51. Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428: 617–624.