Biochemical and Biophysical Research Communications 401 (2010) 112–116

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Understanding gene essentiality by finely characterizing hubs in the yeast protein interaction network

Kaifang Pang [a], Huanye Sheng [a,*], Xiaotu Ma [b,**]

[a] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Department of Molecular and Cell Biology, Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080, USA

ARTICLE INFO

Article history:
Received 31 August 2010
Available online 15 September 2010

Keywords:
The centrality–lethality rule
Hub
Protein interaction
Essentiality
Gene expression
Functional module

ABSTRACT

The centrality–lethality rule, i.e., high-degree proteins or hubs tend to be more essential than low-degree proteins in the yeast protein interaction network, reveals that a protein's central position indicates its important function, but whether and why hubs tend to be more essential have been heavily debated. Here, we integrated gene expression and functional module data to classify hubs into four types: non-co-expressed non-co-cluster hubs, non-co-expressed co-cluster hubs, co-expressed non-co-cluster hubs and co-expressed co-cluster hubs. We found that all the four hub types are more essential than non-hubs, but they also show different enrichments in essential proteins. Non-co-expressed non-co-cluster hubs play key role in organizing different modules formed by the other three hub types, but they are less important to the survival of the yeast cell. Among the four hub types, co-expressed co-cluster hubs, which likely correspond to the core components of stable protein complexes, are the most essential. These results demonstrated that our classification of hubs into four types could better improve the understanding of gene essentiality.

1. Introduction

One of the most important problems in systems biology is to find relations between a protein's topological properties in the protein interaction network and its functional features such as essentiality. The centrality–lethality rule was first observed by Jeong et al. [1], who suggested that over-representation of essential proteins among high-degree nodes can be attributed to the central role hubs play in mediating interactions among numerous, less connected proteins. Since then the relationship between protein degree and essentiality has been actively debated [2–11]. Recently, Zotenko et al. [10] confirmed the higher essentiality of hubs, and further explained that the majority of hubs are essential due to their involvement in Essential Complex Biological Modules, a group of densely connected proteins with shared biological function that are enriched in essential proteins. However, Yu et al. [11] presented clear evidence that protein degree does not correlate with essentiality, and argued that the centrality–lethality rule originates from biases towards essential and well-studied proteins in the original dataset used. Thus, the relationship between protein degree and essentiality need to be further investigated.

Therefore, we re-investigated the centrality–lethality rule by using an updated high confidence network [12,13]. This network has a relatively large size and each interaction is validated at least twice. By classifying hubs in the network into four types, we confirmed the higher essentiality of hubs and proposed an explanation for the centrality–lethality rule.

2. Materials and methods

2.1. Protein interaction and gene essentiality datasets

The updated high confidence ("Updated-HC") network [12,13] was built from several datasets. An interaction observed at least twice in different datasets was kept, and the resultant network contains 4008 proteins with 9857 interactions. A list of essential genes was obtained from the *Saccharomyces* Genome Deletion Project [14].

2.2. Gene expression datasets

Ten gene expression datasets [15–24], each with more than 50 samples (conditions), were obtained from the Yeast Functional Genomics Database [25] and the *Saccharomyces* Genome Database [26]. Genes with missing value in >30% of the samples in a dataset were removed. Remaining missing values were imputed by the KNN impute algorithm with $K = 10$ using Euclidean distance [27],

* Corresponding author.
** Corresponding author.
E-mail addresses: hysheng@sjtu.edu.cn (H. Sheng), xiaotuma@utdallas.edu (X. Ma).

0006-291X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2010.09.021

and technical replicates (i.e., spot repeats and dye swaps) were averaged.

2.3. Construction of gene co-expression networks

Pearson correlation coefficient (PCC) r was used as a similarity measure between the expression profiles of two genes. The PCC r was then converted into z-score using Fisher transformation:

$$z(r) = \frac{\sqrt{n-3}}{2} \log \frac{1+r}{1-r}$$

which approximately follows a standard normal distribution under the hypothesis of independence, where n is the sample size. We only considered positive correlations since they are reported to be more reflective of functional similarity than negative correlations [28]. Next, P-values were obtained for the null hypothesis of no positive correlations and were corrected for multiple hypothesis testing by using false discovery rate (FDR) control procedure [29], and the adjusted P-values were set at the threshold of 0.001 per dataset. In addition, we only considered those pairs that are among the top 10% of all possible correlations to avoid introducing too many high correlations. Our two-stage threshold selection procedure is similar to the procedure that controls both statistical significance and biological significance in [28,30]. For each gene expression dataset, two genes were declared to be co-expressed if their correlation coefficient is above the thresholds of both FDR and PER.

2.4. Detection of functional modules

The overlapping k-clique modules in the "Updated-HC" network were detected by the software CFinder, which is based on clique percolation method [31,32]. Here, we used clique size $k = 5$ and our results were not sensitive to different values of k (data not shown). As a result, 65 functional modules were obtained.

2.5. Classification of hubs

About 21% (841/4008) of proteins in the "Updated-HC" network were defined as hubs with degree $\geqslant 7$. The remaining proteins were defined as non-hubs. The hub classification procedure was carried out in the following four steps:

First, in the "Updated-HC" network, let PPI_g denote the set of interaction partners of a given hub protein g and $PPID_g$ denote the number of proteins in PPI_g. Let $ePPI_g(i)$ denote the set of interaction partners that are found to be significantly co-expressed with the hub g in gene expression dataset i (i = 1, 2, ..., 10) and $ePPID_g(i)$ denote the number of proteins in $ePPI_g(i)$. The co-expressed protein-protein interaction degree (ePPID) of the hub g is then defined as $ePPID_g = \max(ePPID_g(i);\ i = 1, 2, \ldots, 10)$. A hub was defined as a co-expressed hub if the ratio of ePPID to PPID (ePPID/PPID) is $\geqslant 0.5$; otherwise, it was defined as a non-co-expressed hub.

Second, let $cPPI_g(j)$ denote the set of interaction partners that are in the same functional module j (j = 1, 2, ..., 65) with the hub g and $cPPID_g(j)$ denote the number of proteins in $cPPI_g(j)$. The co-cluster protein–protein interaction degree (cPPID) of the hub g is then defined as $cPPID_g = \max(cPPID_g(j);\ j = 1, 2, \ldots, 65)$. A hub was defined as a co-cluster hub if the ratio of cPPID to PPID (cPPID/PPID) is $\geqslant 0.5$; otherwise, it was defined as a non-co-cluster hub.

Third, for each set $ePPI_g(i)$ of the hub g whose $ePPID_g(i)/PPID_g$ is $\geqslant 0.5$, take intersection of it with each set $cPPI_g(j)$ of the hub g whose $cPPID_g(j)/PPID_g$ is $\geqslant 0.5$. The resultant set is denoted as $ecPPI_g(i,j)$, which represents the set of co-expressed co-cluster interaction partners of the hub g in gene expression dataset i and functional module j. Let $ecPPID_g(i,j)$ denote the number of proteins in $ecPPI_g(i,j)$. The co-expressed co-cluster protein–protein interaction degree (ecPPID) of the hub g is then defined as $ecPPID_g = \max(ecPPID_g(i,j);\ i = 1, 2, \ldots, 10;\ j = 1, 2, \ldots, 65)$. A hub was defined as a co-expressed co-cluster hub if the ratio of ecPPID to PPID (ecPPID/PPID) is $\geqslant 0.5$.

Fourth, define the difference between co-expressed hubs and co-expressed co-cluster hubs as co-expressed non-co-cluster hubs, and the difference between co-cluster hubs and co-expressed co-cluster hubs as non-co-expressed co-cluster hubs. Then, define the intersection of non-co-expressed hubs and non-co-cluster hubs as non-co-expressed non-co-cluster hubs. We noted that co-expressed non-co-cluster hubs and non-co-expressed co-cluster hubs may have common hubs, which are in both co-expressed non-co-cluster hubs and non-co-expressed co-cluster hubs but not in co-expressed co-cluster hubs. As a result, we obtained 290 non-co-expressed non-co-cluster (NE–NC) hubs, 233 non-co-expressed co-cluster (NE–C) hubs, 173 co-expressed non-co-cluster (E–NC) hubs and 185 co-expressed co-cluster (E–C) hubs. NE–C and E–NC hubs have 40 common hubs.

2.6. Betweenness and clustering coefficient

Betweenness is measured by the total number of shortest paths going through a protein in a protein interaction network [33,34]. Clustering coefficient is defined as $C_i = 2n_i/k_i(k_i - 1)$, where n_i denotes the number of direct links connecting the k_i nearest neighbors of node i [35].

2.7. Definition of functional consistence score (FC-score)

For each of GO biological process, cellular component and molecular function categories [36], we obtained the P-value of the most significant GO annotation term for each hub with its interaction partners by using the GO Term Finder software [37]. For each category, the functional consistence score (FC-score) of each hub is defined as

$$\text{FC-score} = -\log_{10}(P\text{-value}).$$

3. Results and discussion

3.1. Topological properties of the four hub types

By integrating gene expression and functional module data, we classified hubs in the "Updated-HC" network into four types: non-co-expressed non-co-cluster (NE–NC) hubs, non-co-expressed co-cluster (NE–C) hubs, co-expressed non-co-cluster (E–NC) hubs and co-expressed co-cluster (E–C) hubs (see Section 2). The four hub types played different roles in organizing the network, which were revealed by their differences in some known centrality measures, such as betweenness [33,34] and clustering coefficient [35] (see Section 2 for details). As shown in Table S1, among the four hub types, NE–NC hubs have the highest betweenness, E–NC hubs have the medium betweenness, and NE–C and E–C hubs have the lowest betweenness. In contrast to the betweenness differences among the four hub types, NE–NC hubs have the lowest clustering coefficient, E–NC hubs have the medium clustering coefficient, and NE–C and E–C hubs have the highest clustering coefficient. However, the four hub types have no much degree differences. These results implied that NE–NC hubs tend to be global connectors, NE–C and E–C hubs tend to form local and dense modules, and E–NC hubs tend to form local and slightly less dense modules. Thus, in this paper, we also called NE–NC hubs intermodular hubs, and NE–C, E–NC and E–C hubs intramodular hubs. We noted that

the software CFinder [31,32] tends to detect densely connected functional modules. E–NC hubs are not in such functional modules with the majority of their interaction partners, but they tend to show significant co-expression and similar functions (see next section) with their interaction partners. Thus, they might form slightly less dense modules.

The different roles of the four hub types in organizing the network were better reflected by their different effects upon removal on the overall topology of the network [38,39]. Here, we used two metrics to measure the effects upon hub removal on the network: characteristic path length (CPL) and main component size (MCS). The CPL, defined as the average distance (shortest path length) between node pairs, reflects the overall network connectivity [38]. Large changes of the CPL and MCS upon removal of a hub indicate that the hub plays key role in the overall network connectivity. When removed from the network serially in a descending order of degree, intermodular and intramodular hubs had distinct effects on the overall topology of the network. Removing intermodular hubs dramatically disrupted the CPL of the network, whereas removing intramodular hubs had negligible effects (Fig. 1A). After removal of intramodular hubs, the MCS had almost no changes, revealing that the network is tolerant to intramodular hub removal. After removal of intermodular hubs, the MCS was still large, but intermodular hub removal had more effect than intramodular hub removal (Fig. 1B). Thus, the network was more sensitive to intermodular hub removal than intramodular hub removal, further revealing the global role of intermodular hubs and the local role of intramodular hubs.

3.2. Functional consistence of the four hub types

The different topological roles of the four hub types were then validated by their different functional consistence with their interaction partners. For each of biological process, cellular component and molecular function categories, we computed the functional consistence scores (FC-scores, see Section 2) of the four hub types.

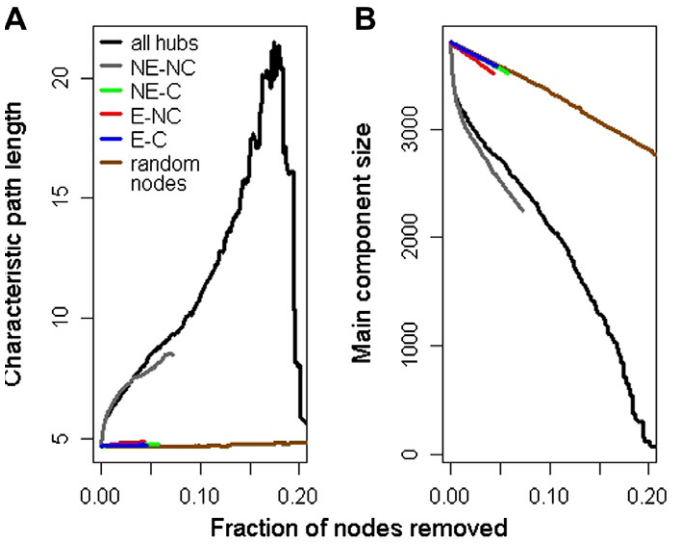

Fig. 1. Intermodular (NE–NC) hubs tend to be global connectors, while intramodular (NE–C, E–NC and E–C) hubs tend to form local modules. The effects on the characteristic path length (A) and main component size (B) of the "Updated-HC" network upon hub (node) removal. Hubs are removed from the network serially in a descending order of degree. Attacks against all hubs (black curve), NE–NC hubs (gray curve), NE–C hubs (green curve), E–NC hubs (red curve), E–C hubs (blue curve) and random nodes (brown curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We then found that among the four hub types, NE–NC hubs have the lowest FC-scores, E–NC hubs have the medium FC-scores, and NE–C and E–C hubs have the highest FC-scores (Fig. 2 and Table S2). In other words, this result explained that NE–NC hubs have more diverse functions with their interaction partners, E–NC hubs have similar functions with their interaction partners, and NE–C and E–C hubs have more similar functions with their interaction partners.

3.3. Essentiality of the four hub types

The topological differences among the four hub types also indicated their different enrichments in essential proteins. We computed the percentage of essential proteins [14] among non-hubs and the four hub types, and found that all the four hub types are significantly more essential than non-hubs (Fig. 3A and Table S3). Subsequently, among the four hub types, NE–NC hubs have the lowest essentiality, E–NC and NE–C hubs have the medium essentiality, and E–C hubs have the highest essentiality (Fig. 3A and Table S3). Since NE–C, E–NC and E–C hubs (especially NE–C and E–C hubs) tend to show more similar functions with their interaction partners, they likely correspond to the core components of functional modules or complexes. It is known that stable complexes tend to show significant co-expression [40], so E–C hubs likely correspond to the core components of stable complexes. These might explain why NE–C, E–NC and E–C hubs are more essential and why E–C hubs are the most essential. The above results were also consistent with the observed higher essentiality of multi-interface hubs [7], which most likely correspond to the core components of large and stable multi-protein complexes. In addition, we noticed that on average, NE–NC hubs have the highest degree (though insignificant) and betweenness (Table S1) but the lowest essentiality, which reveals the possible biases of using degree [1] and betweenness [3,8,41] to predict essential genes.

The non-hub nearest neighbors of the four hub types also showed different essentiality, with the lowest for those of NE–NC hubs, the medium for those of E–NC and NE–C hubs, and the highest for those of E–C hubs (Fig. 3B and Table S3). This result indicated that the essential non-hubs are more likely to be connected by intramodular hubs than by intermodular hubs. The non-hub nearest neighbors of NE–C, E–NC and E–C hubs (especially NE–C and E–C hubs) likely correspond to the peripheral components of functional modules or complexes and those of E–C hubs likely correspond to the peripheral components of stable complexes, which might explain why the non-hub nearest neighbors of NE–C, E–NC and E–C hubs are more essential and why those of E–C hubs are the most essential. Furthermore, we found that the non-hub nearest neighbors of the essential intramodular hubs are more essential than those of the non-essential intramodular hubs (Fisher's exact test, $P = 3.13e{-}10$). The above results implied that the essential intramodular hubs tend to form essential functional modules or complexes, consistent with the modular nature of gene essentiality [9]. These results also supported the explanation that the majority of hubs are essential due to their involvement in Essential Complex Biological Modules [10].

Putting our findings together with earlier studies [7,9,10] could better explain the underlying mechanism of gene essentiality. As we have shown, intramodular hubs tend to form functional modules or complexes. Thus, if a functional module or complex, in which an intramodular hub participates, is vital for the survival of the yeast cell, and the deletion of the intramodular hub gene leads it impossible for the formation of the functional module or complex, the deleted intramodular hub gene will be an essential one. The intramodular hubs and their non-hub nearest neighbors tend to correspond to the core and peripheral components of functional modules or complexes, respectively. The formation of

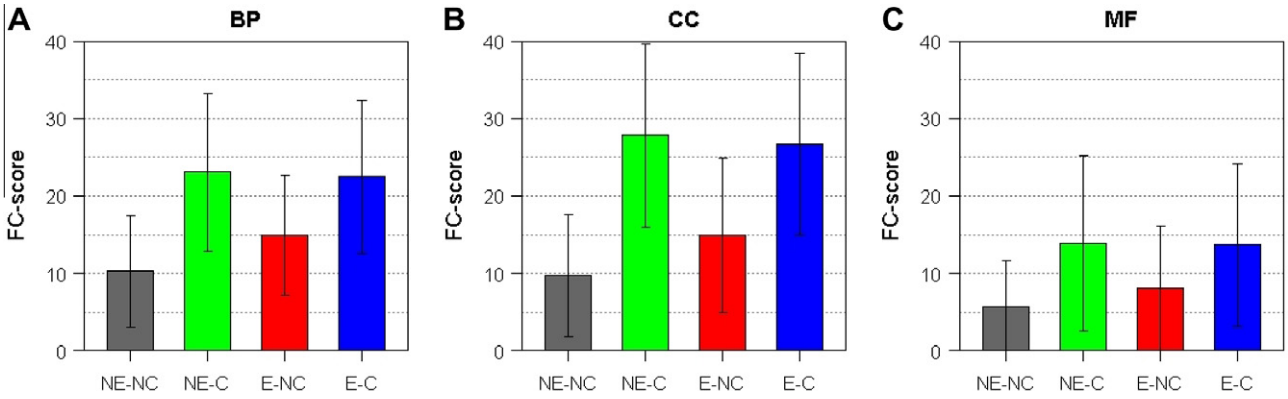

Fig. 2. Intermodular (NE–NC) hubs show more diverse functions with their interaction partners, while intramodular (NE–C, E–NC and E–C) hubs show more similar functions with their interaction partners. The functional consistence scores (FC-scores) of the four hub types in the "Updated-HC" network for (A) biological process (BP), (B) cellular component (CC) and (C) molecular function (MF) categories. Bars show the mean and standard deviation of FC-scores for the four hub types: NE–NC hubs (gray), NE–C hubs (green), E–NC hubs (red) and E–C hubs (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

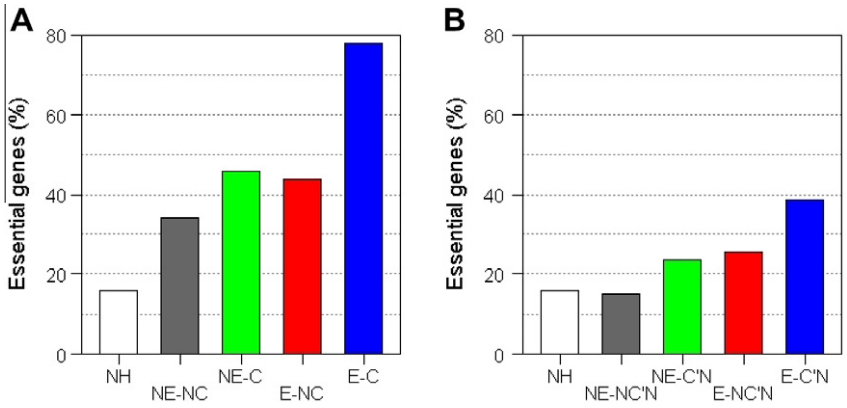

Fig. 3. Essentiality of the four hub types and their non-hub nearest neighbors. (A) Essentiality of the four hub types. Bars show the percentage of essential proteins among different groups of proteins: NH (non-hubs, white), NE–NC hubs (gray), NE–C hubs (green), E–NC hubs (red) and E–C hubs (blue). (B) Essentiality of the non-hub nearest neighbors of the four hub types. NE–NC′N, NE–C′N, E–NC′N and E–C′N represent the non-hub nearest neighbors of the four hub types. Bars show the percentage of essential proteins among different groups of proteins: NH (non-hubs, white), NE–NC′N (gray), NE–C′N (green), E–NC′N (red) and E–C′N (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

functional modules or complexes often needs their core components rather than their peripheral components. Thus, if a functional module or complex is essential, most of its core components will tend to be essential and will be more essential than its peripheral components. Conversely, if a functional module or complex is non-essential, most of its core and peripheral components will tend to be non-essential. In contrast, intermodular hubs tend to play major role in organizing different functional modules or complexes. When deleting an intermodular hub gene, the communication between different functional modules or complexes may be rewired through other paths by other proteins, and thus the yeast cell can still survive and the deleted intermodular hub gene will be non-essential. To sum up, the yeast protein network tends to be tolerant to intramodular hub attack, but the yeast cell tends to die upon the deletion of intramodular hub genes. This could help better understand gene essentiality and help biologists select candidates for gene deletion experiments.

4. Conclusion

In this paper, we confirmed the centrality–lethality rule by using a high confidence network. Furthermore, we found that the majority of hubs are essential due to their local role as core components of functional modules or complexes, but not due to their global role in organizing the whole network. Consequently, the non-hub nearest neighbors of the local intramodular hubs tend to be more essential than those of the global intermodular hubs, with the non-hub nearest neighbors of co-expressed co-cluster hubs even more essential than the global intermodular hubs (though insignificant, Fisher's exact test, $P = 0.326$). However, protein networks only constructed from yeast two-hybrid data are depleted for co-complex associations, which might explain why there are weak or insignificant correlations between protein degree and essentiality in such networks [6,10,11]. Therefore, our study demonstrated that integrating gene expression and functional module data to finely characterize hubs in the yeast protein interaction network could better improve our understanding of gene essentiality.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2010.09.021.

References

[1] H. Jeong, S.P. Mason, A.L. Barabasi, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.
[2] H. Yu, D. Greenbaum, H. Xin Lu, X. Zhu, M. Gerstein, Genomic analysis of essentiality within protein networks, Trends Genet. 20 (2004) 227–231.

[3] M.W. Hahn, A.D. Kern, Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks, Mol. Biol. Evol. 22 (2005) 803–806.
[4] S. Coulomb, M. Bauer, D. Bernard, M.C. Marsolier-Kergoat, Gene essentiality and the topology of protein interaction networks, Proc. Biol. Sci. 272 (2005) 1721–1725.
[5] X. He, J. Zhang, Why do hubs tend to be essential in protein networks? PLoS Genet. 2 (2006) e88.
[6] N.N. Batada, L.D. Hurst, M. Tyers, Evolutionary and physiological importance of hub proteins, PLoS Comput. Biol. 2 (2006) e88.
[7] P.M. Kim, L.J. Lu, Y. Xia, M.B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights, Science 314 (2006) 1938–1941.
[8] H. Yu, P.M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics, PLoS Comput. Biol. 3 (2007) e59.
[9] G.T. Hart, I. Lee, E.R. Marcotte, A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality, BMC Bioinf. 8 (2007) 236.
[10] E. Zotenko, J. Mestre, D.P. O'Leary, T.M. Przytycka, Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality, PLoS Comput. Biol. 4 (2008) e1000140.
[11] H. Yu, P. Braun, M.A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.F. Rual, A. Dricot, A. Vazquez, R.R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.S. De Smet, A. Motyl, M.E. Hudson, J. Park, X. Xin, M.E. Cusick, T. Moore, C. Boone, M. Snyder, F.P. Roth, A.L. Barabasi, J. Tavernier, D.E. Hill, M. Vidal, High-quality binary protein interaction map of the yeast interactome network, Science 322 (2008) 104–110.
[12] N.N. Batada, T. Reguly, A. Breitkreutz, L. Boucher, B.J. Breitkreutz, L.D. Hurst, M. Tyers, Stratus not altocumulus: a new view of the yeast protein interaction network, PLoS Biol. 4 (2006) e317.
[13] N.N. Batada, T. Reguly, A. Breitkreutz, L. Boucher, B.J. Breitkreutz, L.D. Hurst, M. Tyers, Still stratus not altocumulus: further evidence against the date/party hub distinction, PLoS Biol. 5 (2007) e154.
[14] G. Giaever, A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A.P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.D. Entian, P. Flaherty, F. Foury, D.J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J.H. Hegemann, S. Hempel, Z. Herman, D.F. Jaramillo, D.E. Kelly, S.L. Kelly, P. Kotter, D. LaBonte, D.C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S.L. Ooi, J.L. Revuelta, C.J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D.D. Shoemaker, S. Sookhai-Mahadeo, R.K. Storms, J.N. Strathern, G. Valle, M. Voet, G. Volckaert, C.Y. Wang, T.R. Ward, J. Wilhelmy, E.A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J.D. Boeke, M. Snyder, P. Philippsen, R.W. Davis, M. Johnston, Functional profiling of the *Saccharomyces cerevisiae* genome, Nature 418 (2002) 387–391.
[15] R.B. Brem, L. Kruglyak, The landscape of genetic complexity across 5, 700 gene expression traits in yeast, Proc. Natl Acad. Sci. USA 102 (2005) 1572–1577.
[16] A.P. Gasch, P.T. Spellman, C.M. Kao, O. Carmel-Harel, M.B. Eisen, G. Storz, D. Botstein, P.O. Brown, Genomic expression programs in the response of yeast cells to environmental changes, Mol. Biol. Cell 11 (2000) 4241–4257.
[17] A.P. Gasch, M. Huang, S. Metzner, D. Botstein, S.J. Elledge, P.O. Brown, Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p, Mol. Biol. Cell 12 (2001) 2987–3003.
[18] T.R. Hughes, M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, M.J. Kidd, A.M. King, M.R. Meyer, D. Slade, P.Y. Lum, S.B. Stepaniants, D.D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, S.H. Friend, Functional discovery via a compendium of expression profiles, Cell 102 (2000) 109–126.
[19] K.L. Huisinga, B.F. Pugh, A TATA binding protein regulatory network that governs transcription complex assembly, Genome Biol. 8 (2007) R46.
[20] S. Mnaimneh, A.P. Davierwala, J. Haynes, J. Moffat, W.T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N.J. Krogan, S.L. Hiley, Z. Li, Q. Morris, J. Grigull, N. Mitsakakis, C.J. Roberts, J.F. Greenblatt, C. Boone, C.A. Kaiser, B.J. Andrews, T.R. Hughes, Exploration of essential gene functions via titratable promoter alleles, Cell 118 (2004) 31–44.
[21] S.M. O'Rourke, I. Herskowitz, Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis, Mol. Biol. Cell 15 (2004) 532–542.
[22] C.J. Roberts, B. Nelson, M.J. Marton, R. Stoughton, M.R. Meyer, H.A. Bennett, Y.D. He, H. Dai, W.L. Walker, T.R. Hughes, M. Tyers, C. Boone, S.H. Friend, Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles, Science 287 (2000) 873–880.
[23] A.J. Saldanha, M.J. Brauer, D. Botstein, Nutritional homeostasis in batch and steady-state culture of yeast, Mol. Biol. Cell 15 (2004) 4089–4104.
[24] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, B. Futcher, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, Mol. Biol. Cell 9 (1998) 3273–3297.
[25] The Yeast Functional Genomics Database (<http://yfgdb.princeton.edu/>).
[26] J.M. Cherry, C. Adler, C. Ball, S.A. Chervitz, S.S. Dwight, E.T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, D. Botstein, SGD: Saccharomyces Genome Database, Nucleic Acids Res. 26 (1998) 73–79.
[27] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R.B. Altman, Missing value estimation methods for DNA microarrays, Bioinformatics 17 (2001) 520–525.
[28] H.K. Lee, A.K. Hsu, J. Sajdak, J. Qin, P. Pavlidis, Coexpression analysis of human genes across many microarray data sets, Genome Res. 14 (2004) 1085–1094.
[29] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J R Statist Soc Ser B 57 (1995) 289–300.
[30] D. Zhu, A.O. Hero, Z.S. Qin, A. Swaroop, High throughput screening of co-expressed gene pairs with controlled false discovery rate (FDR) and minimum acceptable strength (MAS), J. Comput. Biol. 12 (2005) 1029–1045.
[31] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (2005) 814–818.
[32] B. Adamcsek, G. Palla, I.J. Farkas, I. Derenyi, T. Vicsek, CFinder: locating cliques and overlapping modules in biological networks, Bioinformatics 22 (2006) 1021–1023.
[33] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35–41.
[34] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826.
[35] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[36] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet. 25 (2000) 25–29.
[37] E.I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, G. Sherlock, GO::TermFinder – open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes, Bioinformatics 20 (2004) 3710–3715.
[38] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature 406 (2000) 378–382.
[39] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[40] R. Jansen, D. Greenbaum, M. Gerstein, Relating whole-genome expression data with protein–protein interactions, Genome Res. 12 (2002) 37–46.
[41] M.P. Joy, A. Brock, D.E. Ingber, S. Huang, High-betweenness proteins in the yeast protein interaction network, J. Biomed. Biotechnol. 2005 (2005) 96–103.