PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2008, 35(1): 35~42
www.pibb.ac.cn

Analysis of Transcriptional Regulation Collaboration Networks in Saccharomyces cerevisiae*

CHEN Lan[1,3)**], CAI Lun[1,3)**], Geir Skogerbø[2)], CHEN Run-Sheng[1,2) ***]
([1)]Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China;
[2)]Bioinformatics Laboratory, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China;
[3)]Graduate School of The Chinese Academy of Sciences, Beijing 100039, China)

Abstract Collaboration networks have proven informative when used to describe various kinds of human relationships. Similar strategy could be used in the transcriptional regulatory network. A collaboration network of target genes (TGs) was constructed based on common transcription factors (TFs), and similarly, a smaller network of transcription factors was constructed based on their common target genes. After clustering the target gene collaboration networks, genes in the same cluster were often enriched for one or more GO terms. The results also showed that genes with specific GO terms tend to share similar regulatory mechanisms. It indicates that in a collaboration network approach the relatively simple "regulatory mechanism" measure used here was able to extract considerable biologically relevant information. Moreover, a definition of anomaly used before in a bipartite graph analysis method was applied into the collaboration networks analysis. And the correlation between the anomalies and the essential genes was discovered. In a conclusion, a collaboration network approach may be a valuable supplement to other analyses of transcriptional networks.

Key words transcriptional regulatory network, collaboration network, anomaly

To reveal the gene transcriptional regulations in a cell is important for understanding the gene functions. In the past years, many types of data sources which can aid in detecting the transcriptional regulations were rapidly accumulated, such as the chromatin immunoprecipitation (ChIP) data which provides the protein-DNA interactions [1, 2]. Many of the previous studies on gene transcriptional regulation networks focus on the network topology analysis, for example, the degree distribution, clustering coefficient, or analyzing the motif [3~5]. In the topology structure analysis, the regulation network is conceptualized as a directed network graph. Its analysis process is different from the undirected graph such as protein-protein interaction network. There were many other works which aimed at discovering the correlations between regulatory mechanisms and the gene functions. For example, Yu et al's study [6] indicated that genes targeted by the same transcriptional factors are more likely to share related functions than expected randomly, and the more common TFs they shared, the more pronounced the functional similarity they have.

In the transcriptional regulatory network, when we considered the similarities of regulatory mechanisms, two aspects of information should be involved. One is the regulatory similarity between transcription factors (TFs), which measures if they tend to regulate the same target genes; the other is the regulatory similarity between target genes (TGs), which measures if they tend to be regulated by the same TFs. However, the studies of the regulatory mechanisms before usually took one aspect of information into consideration at one time [6, 7], and seldom paid attention to the correlation of them both. If we transformed the regulatory network into two networks which capture the mechanism similarities of TFs and TGs separately, not only it could be clear and convenient to apply the network analysis methods to

*This work was supported by a grant from The National Natural Science Foundation of China (60496320, 30500104, 30570393).
** These authors contributed equally to this work.
***Corresponding author.
Tel: 86-10-64888543, E-mail: crs@sun5.ibp.ac.cn
Received: May 8, 2007 Accepted: August 3, 2007

investigate more and detailed correlations between regulatory mechanism similarities and gene functions in each network, but also it is possible to investigate the gene properties which are correlated with both of the two networks.

Collaboration networks have proven informative when used to describe various kinds of human relationships [8, 9]. A collaboration network can be established by connecting two elements with some common features, e.g. two actors with a common membership in a given organization or co-authors of a published paper. Within a cell or an organism genes can be linked in similar networks according to their "membership" in a metabolic pathway or functional structure, or by the transcription factors that control their activity. We therefore used the protein-DNA interaction (PDI) data underlying the transcriptional regulatory network of Luscombe et al [4] to construct a collaboration network among TGs based on the TFs they have in common, and similarly, a smaller network of TFs linked by their common TGs. The weights of links between nodes in the collaboration networks represent the regulatory similarities between TFs or TGs. Accordingly by this step we transformed the transcriptional regulatory network into two undirected networks, which capture the information of mechanism similarities of TFs and TGs separately.

We adopt a statistical method based on hypergeometric distribution to measure the similarities of regulatory mechanism, which was applied before on protein-protein interaction network to estimate the probability of two proteins' closeness in the network [10], and used in the transcriptional regulatory network to assess the similarity of target sets between two TFs [7]. This measurement has a prominent quality that it shows considerable resistance to the noise in the network. The relative order of most of the weights was maintained even after disturbing the network by adding up to 50% random connections [10]. As we know, the current transcriptional regulatory network data is noisy and incomplete; a weight measure strategy that resists noise should induce more reliable analysis results.

Here, we introduced the collaboration networks to analyze yeast transcriptional regulatory network (PDI: protein-DNA interaction) network. We clustered the TG and TG collaboration networks separately and recaptured the correlations between regulatory mechanisms and gene functions. Moreover, we found an interesting phenomenon that the anomalies in collaboration networks have relations with the essential genes.

1 Materials and methods

1.1 Dataset

To build collaboration networks, we used the data from transcriptional regulatory network of S. cerevisiae [4]. This network integrated data from genetic, biochemical and ChIP-chip experiments, which comprised 3 459 genes, including 142 TFs. Most of TFs directly acted on more than 30 target genes, and totally there were 2 444 directed links from TFs to TGs.

1.2 Construction of transcriptional collaboration networks

Genes in the transcriptional regulatory network were divided into two groups, TFs and TGs. The TF group consisted of 142 genes, and the TG group consisted of 3 420 genes. There were 103 overlapping genes, which were found in both TF and TG groups. Applying the definition of Newman [8, 9] collaboration networks were constructed for each group separately. Any two TFs targeting the same gene were connected by an undirected edge; similarly, any two TGs targeted by a common TF were also linked, thereby transforming the directed PDI network into two undirected collaboration networks.

1.3 Definition of the weights in collaboration networks

We measure the weight of links (that is, the similarity between the regulatory mechanisms involving any two TFs or TGs) in collaboration networks from a statistical method which already used in Protein-Protein network [10] and transcriptional regulatory network [7].

$$P(N,n_1,n_2,m)=\frac{\binom{N}{m}\binom{N-m}{n_1-m}\binom{N-n_1}{n_2-m}}{\binom{N}{n_1}\binom{N}{n_2}}$$

$$=\frac{(N-n_1)!(N-n_2)!n_1!n_2!}{N!m!(n_1-m)(n_2-m)!(N-n_1-n_2+m)!}$$

For the measuring of link weight between TF1 and TF2, as showed in the formula, N denoted the total number of genes in TG network, n_1 and n_2 denoted the numbers of target genes in TG network of TF1 and TF2 respectively. m denoted the number of common target genes shared by TF1 and TF2. This formula calculated the probability of sharing m common target

genes for the two TFs. Weights of two genes in TG collaboration network were calculated by similar strategies. For the clustering calculation conveniently, the probabilities were transformed as ln(probabilities) × (- 1) to be the final weight value.

1.4 Clustering of collaboration network

We applied the agglomerative hierarchical clustering algorithm in R package with average distance setting to cluster the vertices in the TG network according to weight values[11, 12]. As TG pairs with at least two TFs in common intuitively appear as the most interesting part of the dataset, we used a threshold of Weight = 5 to cut the hierarchical clustering tree for most analyses; however, also other thresholds were tested for comparison. Clusters with more than two genes with GO annotations were selected (called "valid clusters") for further analyses. All clusters were examined for significant enrichment of genes with similar GO annotations (SGD Gene Ontology Slim Mapper[13]) with respect to "process", "function", and "component" terms (the terms "unannotated" and "process /function /component unknown" were not included). Enrichment of a cluster in some GO terms was evaluated statistically by a P-value based on the "Hypergeometric distribution", and $P < 0.05$ was set as the threshold to evaluate the significance. Benjamini and Hochberg's False Discovery Rate approach [14] was applied to adjust P-values for multiple testing problems.

1.5 GO terms distribution in enrichment clusters

To check genes with which terms were also with similar regulatory mechanisms, and with which terms were with not so similar regulatory mechanisms, we collected the genes which in enrichment clusters (thus with similar regulatory mechanism) and with same GO terms (thus with functional relations), then compared the gene numbers distribution of these GO terms to the background (the gene numbers of the GO terms in all the clusters). Apply the "Hypergeometric distribution" a P-value which depicted the probability for observing the certain number of genes with certain term in enriched clusters from the background was calculated for each term. If the P-value is below 0.01, we said genes with this term were significant enriched in enrichment clusters, greater than 0.99, means genes with this term were significant absent in enrichment clusters. Because the GO terms have three types (process, function, component terms), we must test them separately. The threshold 5 is taken (The threshold determines how to cut off the hierarchy tree to get the clusters).

1.6 Anomaly analysis

To get more regulatory data for improving the reliability of this analysis, we integrated two sets of ChIP-on-chip data [1, 15] and calculated the two collaboration networks separately again. First we calculated the anomaly score of each TG according to the average link weigh of its TFs in the TF collaboration network. Then we sorted all TGs with their anomaly score from small to big. The yeast essential gene data were downloaded from DEG database[16]. We counted the number of essential genes per 100 TGs which had already sorting with anomaly score.

2 Results and discussion

2.1 Constructions of collaboration networks

We used the protein-DNA interaction (PDI) data from Luscombe et al [4] to construct a collaboration network among target genes, and also a smaller network among the transcription factors (See Materials and methods). The TG collaboration network (TG network) was constructed by linking any two TGs, which were controlled by common TFs and comprised 3 420 genes and 466 175 links. Similarly, the TFs were linked into a collaboration network (TF network) by their common target genes, resulting a network of 142 genes and 2 444 links. To capture the generic features of the two networks, some commonly used network measurement parameters were calculated[17]. The average node degree was 32 for the TF network and 263 for the TG network, indicating that both networks are very dense, especially the TG network. Since each node has so many links to other nodes, it takes only few links to travel from one node to any other node, thus the average shortest path between two nodes for the two networks is quite small, 2 for the TF network and 1.8 for the TG network. The average clustering coefficient, characterizing the overall tendency of the nodes to cluster, was 0.8 and 0.6 for the TF and TG networks, respectively. Such high values indicate that the two collaboration networks (especially the TG network) are dense not only on the whole but also on the local scale. Overall, these network topology measurement parameters suggests that the collaboration networks are much denser than the most other biological networks, e.g. protein-protein interaction networks[18].

For the TG network, the weights range from 0.9 to 43.5 (Figure 1a); with the weight value is higher, the regulatory pattern of the two linked TGs is more similar. A few regulatory patterns were more common than others, for instance, two TGs regulated by the same single TF (W= 4.95) or two TGs regulated by one shared TF, but one had been regulated by another TF (W= 4.26). Links of these two patterns made up more than half of all TG links, and about 95% of the gene pairs shared only one TF (W < 4.95). Pairs sharing at least two TFs (W > 4.95) comprised only 5% of the whole TG network (Figure 1a). So the weight of 5 was set as a default threshold to produce valid data for further analysis.

In the same fashion as for the TG network, we used common target genes to construct a collaboration network between the 142 TFs in the PDI data. The TF network contained 2 444 links with weights ranging from 0.94 to 241.80. However, different from the unbalanced discrete distribution of the weights in the TG network, the number of links decreased gradually (in a non-linear fashion) with increasing weights (Figure 1b). The difference of weight distributions might be caused by the unbalanced scale of the two collaboration networks. For one TF, its average TG number was around 24 while for one TG its average

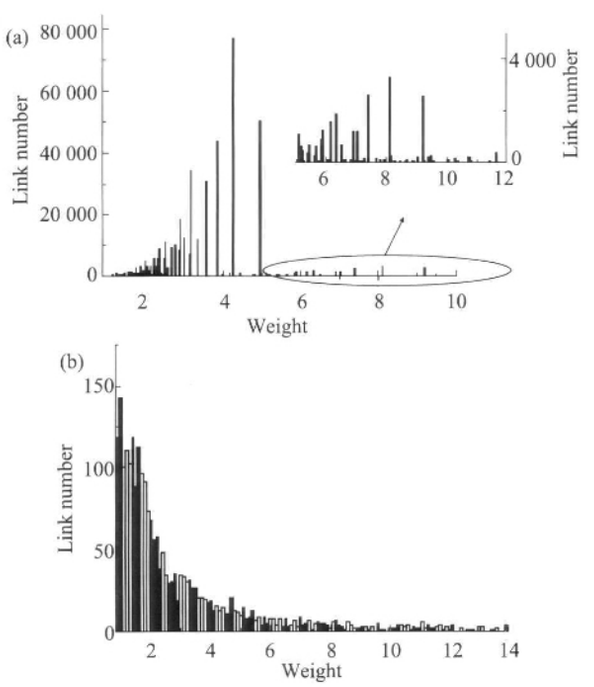

Fig. 1 Weight distribution of collaboration networks

(a)Weight distribution of the TG collaboration network. Approximately 95% of the links has weight smaller than 5. The distribution of links with weights greater than 5 is shown in the insert. (b) The weight distribution of the TF collaboration network.

TF was only about 0.04. Therefore for TG network the TF sharing patterns were relatively fewer and simple, for example most TG pairs only share one TF, whereas for TF network the TG sharing patterns were more complicated, that they shared more and variant number of TGs. It caused relative more smooth weight distribution of TF collaboration network than that of TG collaboration network.

Testing the robustness of weight measurement method against network noise showed that even after insertion of up to 50 % false links, of the top 5% links of the corresponding collaboration networks, about 85% and 70% were still retained in the TF and TG networks, respectively, indicating the robustness of the weight calculation [10] also applies to this type of network. For currently PDI data were noisy and uncompleted, a robustness method was necessary for generating credibility resulting.

2.2 TG collaboration network analysis

To test for possible correlations between regulatory mechanism and biological functions of the target genes, clustering vertices with at least two common TFs (i.e. threshold $W \geq 5$) according to similarities in regulatory mechanism produced 215 clusters (See Materials and methods). Of these 52% were significantly enriched for one or more GO terms, comprising 69% of all TGs genes, showing that in a collaboration network approach the relatively simple "regulatory mechanism" measure used here is able to extract considerable biologically relevant information. We called clusters that were significantly enriched for one or more GO terms "enrichment clusters", the others that were not enriched for any GO term are "non-enrichment clusters". With more stringent thresholds the percentage decreased (Table 1),

Table 1 Proportion of clusters enriched in GO terms[1)]

Network	Threshold	Process	Function	Component	Any
TG	3	0.38	0.29	0.35	0.63
	5	0.35	0.26	0.22	0.52
	9	0.24	0.20	0.19	0.36
	12	0.23	0.26	0.24	0.42
	15	0.23	0.20	0.23	0.38
TF	5	0.56	0.04	0.04	0.59
	9	0.24	0.20	0.19	0.36
	15	0.57	0.04	0	0.61
	54	0.47	0	0	0.47

[1)]Clusters were produced with different thresholds (the second column). The proportion of clusters enriched in the GO terms "process", "function" and "component" (third to fifth column, respectively), or in any of the three terms (the sixth column) are listed.

however, more than one-third of the clusters were still significantly enriched for one or more GO terms even at W≥15. When the threshold was relaxed from 5 to 3 (meaning that some gene pairs sharing only one TF are included), the number of valid clusters also decreased as each cluster contained more members, but the percentage of enrichment clusters increased to 62%, which reflects the observation that genes sharing only one TF may also be functional related[6].

If genes sharing the same GO terms also with similar regulatory mechanisms, they would be more possible to cluster together so as to be enriched in clusters, otherwise they would dispersed into different clusters, with the result that their GO terms would be hardly over represent in clusters. Based on the assumption, after comparing the GO terms distribution of the enrichment and non-enrichment clusters (The GO terms distribution was defined as how many genes were involved in each GO term), we found a minority of all possible GO terms were actually enriched in enrichment clusters, (find the methods in Materials and methods). Of 31 GO process terms, six terms were significantly over-represented in one or more clusters (Figure 2a). These terms were protein

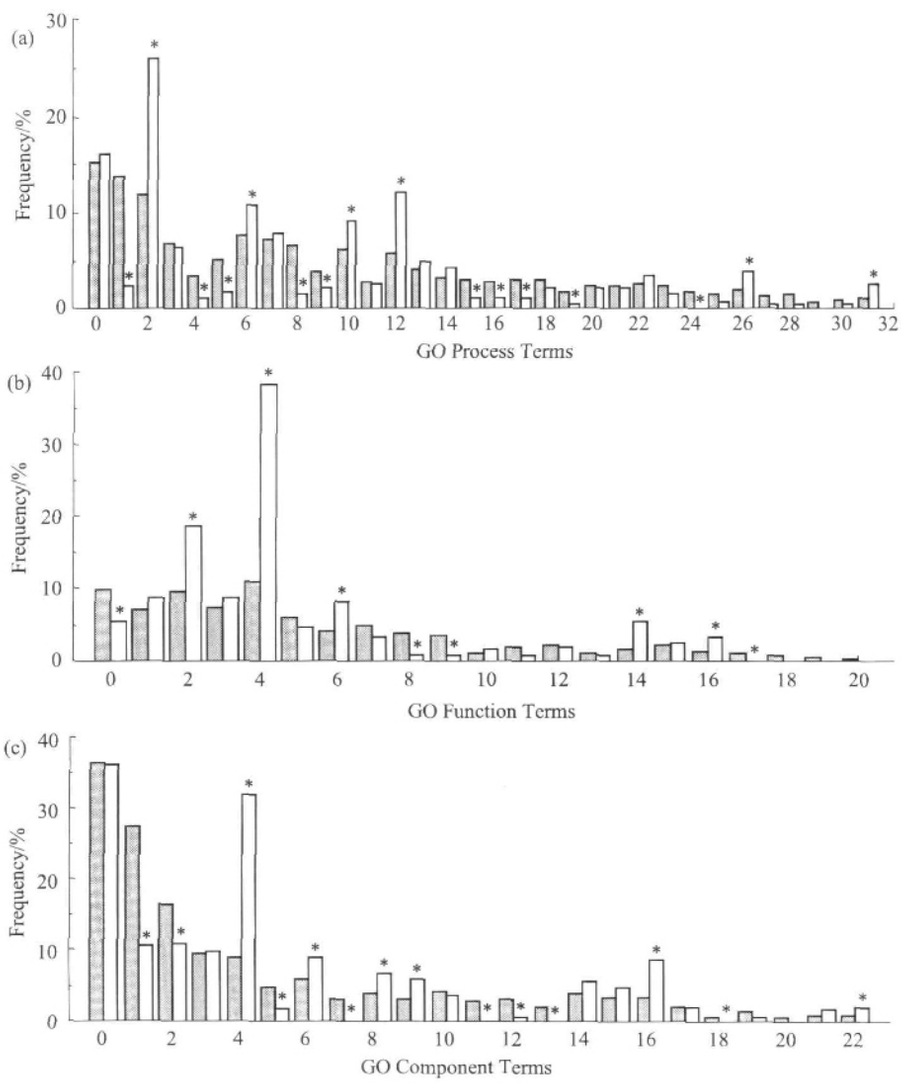

Fig. 2 GO term distribution of enrichment clusters

Gene number frequencies of each GO term in enrichment clusters (Enrich) compared to the background, which were genes proportions with each GO term in all clusters(Total). Threshold 5 was taken to cut 215 clusters. With each term a P-value was calculated to depict the probability for observing this number of genes with this term in enrichment clusters from the background. If the P-value was below 0.01 (marked by star), it indicated that genes with this term were significant enriched in enrichment clusters, greater than 0.99 (marked by hat), means genes with this term were significant absent in enrichment clusters (See Materials and methods). Three types of GO terms had corresponding three figures, and they were the GO process terms (a), the GO function terms (b), and the GO component terms (c). The x-axis denoted the serial number of each GO term, please see supplementary for descriptions of each GO term. ▭: Total; □: Enrich.

biosynthesis (term 2), cell cycle (term 6), generation of precursor metabolites and energy (term 10), carbohydrate metabolism (term 12), cellular respiration (term 26), and electron transport (term 31). Similarly, of the 20 GO function terms, only three terms were significantly over represented in any cluster (Figure 2b); these were the terms "structural molecule activity (term 4)", "oxidoreductase activity (term 6)", and "helicase activity (term 14)". Of the 22 GO component terms, "ribosome (term 4)" was the most prominent (P-value 4.1×10^{-47}) of six significantly enriched terms (Figure 2c).

Taken together, the clustering data had recaptured some central components of the cellular. The most enrichment term of the process, function and component terms "protein synthesis", "structural molecule activity" and "ribosome", respectively, all demonstrated the coordinated transcriptional regulation of genes involved in the ribosome itself or in ribosomal activity. Actually, of total 134 genes with one of the three terms in the enrichment clusters, 100 genes were with the three terms both. Among them, majority are ribosome components, the left genes are related to telomere maintenance. Another particular case was process term 31 ("electron transport"). Only 15 yeast genes had been annotated with this term, out of which 10 were significantly enriched in clusters, showing that these genes had very similar regulatory mechanisms. All of those genes take part in oxidation-reduction reaction of the cytochrome C.

Moreover, some terms were significantly absent from enrichment clusters (Find the methods in Materials and methods, Figure 2), suggested that the regulatory mechanisms of the genes with these terms are relatively not so similar. For the process terms, they were transport (term 0), organelle organization and biogenesis (term 1), transcription (term 4), protein modification (term 8), vesicle-mediated transport (term 9). For the function terms, hydrolase activity (term 0) inclined to be absent from significant clusters. Similar conclusions were also obtained with other thresholds (data not shown). It might be for that these kinds of terms contained genes involved in variant roles, or long range pathways.

2.3 TF network analysis

The TF network was clustered in the same way as the TG network. Although the GO term "transcription process" was common to all TFs, some clusters also shared other GO process terms. At threshold 5, nearly 60% TF clusters were significantly enriched in some GO term. For example, HSF1, YAP1, CAD1, MSN4 and MSN2 shared the GO terms "response to abiotic stimulus" and "response to stimulus", and PUT3, ARG80, GCN4, and ARG81 shared GO term of "amino acid metabolism". With more stringent thresholds, the proportions were slightly reduced (Table 1), More TF clusters were significantly enriched in GO process terms than those in function or component terms, probably because the poor annotation of TFs for GO function and component terms (including only terms like "transcription regulator activity", "DNA binding", "nucleus" and so on).

2.4 Anomalies in the collaboration networks

Since we have transferred the PDI from a directed network into two collaboration networks, more sophisticate network properties which are correlated with both of the two networks can be analyzed.

Discovering the anomalies from data sources is a challenge in the data mining fields. Anomaly usually means the "non-normal" observation in the data. Its definition is complex because the definition of "normality" is various depending on the data sources [19]. Here we applied a definition of anomaly in a bipartite graph analysis method[20] into the collaboration networks. And tried to detect the biology meanings of these anomalies.

For each gene in the TG collaboration network, we extracted its corresponding TFs from the TF collaboration network. Then we calculated the average link weight among these TFs, and took this average weight as a score to evaluate the anomaly degree of this TG. We denoted this score as "anomaly score" of this TG. Because the link weight between TFs in the TF collaboration network measures the similarity of their target gene sets, to some degree it reflects the similarity of the "regulatory behavior" between TFs. Thus for each TG, its anomaly score reflects the "behavior similarity" among its corresponding TFs. Less the anomaly score of a TG is, more dissimilar are the behaviors among its TFs. Now we are interested in what is the biology property of a TG if it tends to be an anomaly.

We test the correlation between the anomaly score of TGs and the essentiality of them. We found that, with the anomaly score increasing, the proportion of essential genes within the TGs has a tendency to decrease (See Materials and methods). In other

words, when the TG is more likely an anomaly, it seems that it is more probably the essential gene (Figure 3). It can be explained as that: If the weight between two TFs in the TF collaboration network is strong, it means the two TFs may regulate the similar target genes, and it implies that they probably tend to participate in the similar biology processes. A gene tends to be an anomaly means than this gene is regulated by several TFs which are probably involved in different biology processes, to some extend it means this gene might also be necessary in different biology processes. Therefore it might be an important gene because deletion of it will affect the running of multiple pathways. That's probably the reason why the anomaly score of TG has the correlation with the "essential genes".

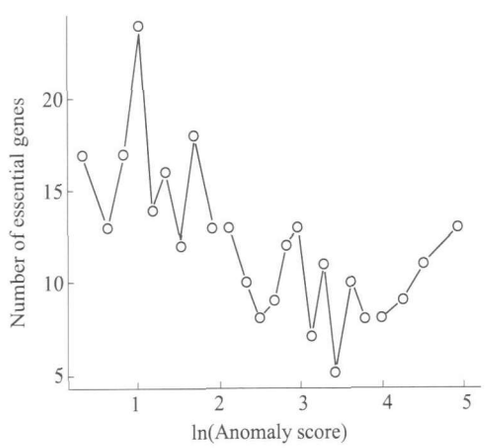

Fig. 3 The correlation between the essential genes and the anomaly scores

The x axis denotes the ln value of the average anomaly score for each 100 TGs; y axis denotes the number of essential genes per 100 TGs. This picture shows that with the anomaly scores of TGs decreasing, the gene tend to be more likely an essential gene.

The anomaly scores were also calculated for each TF by the average link weight of their target genes in the TG network. However, with too few essential genes within these TFs, we did not find any significant correlation between the anomaly score and the essentiality of genes.

This result indicates that, the topology of regulatory network implies the information of the importance of the genes. After we transfer the regulatory network into two collaboration networks, this information could be more easily to be dug out.

3 Conclusions

In this study, we introduced the concepts of collaboration networks to study gene regulatory network. Applying this strategy, yeast transcription regulatory network was transferred into two collaboration networks. Gene pairs in each collaboration network were assigned weights of regulation similarity by a statistical model. Through the collaboration network approach, the known correlation of regulatory similarities and gene functions are discovered again. And more importantly, new properties related to both of the networks could be investigated. We applied a definition of anomaly in an bipartite graph analysis method [20] into the collaboration networks analysis. And found the correlation between the anomalies and the essential genes. In conclusion, a collaboration network approach may be a valuable supplement to other analyses of transcriptional networks.

References

1 Lee T I, Rinaldi N J, Robert F, et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 2002, 298 (5594): 799～804

2 Iyer V R, Horak C E, Scafe C S, et al. Genomic binding sites of the yeast cell~cycle transcription factors SBF and MBF. Nature, 2001, 409(6819): 533～538

3 Guelzim N, Bottani S, Bourgine P, et al. Topological and causal structure of the yeast transcriptional regulatory network. Nat Genet, 2002, 31(1): 60～63

4 Luscombe N M, Madan Babu M, Yu H, et al. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature, 2004, 431(7006): 308～312

5 Shen-Orr S S, Milo R, Mangan S, et al. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet, 2002, 31(1): 64～68

6 Yu H, Luscombe N M, Qian J, et al. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends in Genetics, 2003, 19(8): 422～427

7 Schlitt T, Palin K, Rung J, et al. From gene networks to gene function. Genome Res, 2003, 13(12): 2568～2576

8 Newman M E J. Scientific collaboration networks. Ⅰ. Network construction and fundamental results. Physical Review E, 2001, 64 (1):016131

9 Newman M E J. Scientific collaboration networks. Ⅱ. Shortest paths, weighted networks, and centrality. Physical Review E, 2001, 64(1): 016132

10 Samanta M P, Liang S. Predicting protein functions from redundancies in large-scale protein interaction networks. Proc Natl Acad Sci USA, 2003, 100(22): 12579～12583

11 R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org, 2007

12 Maechler M, Rousseeuw P, Struyf A, Hubert M. Cluster Analysis Basics and Extensions; http://www.R-project.org, 2005
13 Cherry J, Adler C, Ball C, et al. SGD: Saccharomyces genome database. Nucl Acids Res, 1998, 26(1): 73～79
14 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological), 1995, 57(1): 289～300
15 Harbison C T, Gordon D B, Lee T I, et al. Transcriptional regulatory code of a eukaryotic genome. Nature, 2004, 431(7004): 99～104
16 Zhang R, Ou H-Y, Zhang C-T. DEG. A database of essential genes. Nucl Acids Res, 2004, 32(suppl-1): D271～272
17 Barabasi A-L, Oltvai Z N. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 2004, 5:101～115
18 Yook S-H, Oltvai Z N, Barabasi A-L. Functional and topological characterization of protein interaction networks. Proteomics, 2004, 4(44): 928～942
19 Margineantu D, Bay S, Chan P, et al. Data mining methods for anomaly detection KDD-2005 workshop report. ACM SIGKDD Explorations Newsletter, 2005, 7(2): 132～136
20 Sun J, Qu H, Chakrabarti D, et al. Relevance search and anomaly detection in bipartite graphs. ACM SIGKDD Explorations Newsletter, 2005, 7(2): 48～55

酵母转录调控协作网络的分析 *

陈 兰[1,3)**] 蔡 伦[1,3)**] GEIR SKOGERBø[2)] 陈润生[1,2)***]

([1)]中国科学院计算技术研究所，北京 100080; [2)] 中国科学院生物物理研究所，北京 100101;
[3)] 中国科学院研究生院，北京 100039)

摘要 协作网通常被用于描述各种社会关系，相似的概念也可以应用到转录调控网络的研究中．针对被调控基因共享转录因子的相似性，可以建立一个被调控基因协作网，同样，根据转录因子调控基因的相似性可以建立一个相对较小的转录因子协作网．对被调控基因协作网的聚类研究发现，大部分的类都显著地富集一个或者多个 GO 功能注释．进一步的结果分析发现某些 GO 注释的基因更倾向于共享相似的调控机制．这表明，在协作网中，相对简单的调控机制相似性能捕捉生物功能相关的信息．并且，将在二部图分析中使用的概念——“异常点”引入到协作网的分析中，发现协作网的异常点和致死基因有相关性．综上所述，协作网的方法是分析转录调控网络的一个有用的补充．

关键词 转录调控网络，协作网，异常点
学科分类号 180.14，180.1465

---

* 国家自然科学基金资助项目(6049320, 30500104, 30570393).
** 并列第一作者
*** 通讯联系人. Tel: 010-64888543, E-mail: crs@sun5,ibp,ac.cn
收稿日期：2007-05-08，接受日期：2007-08-03