A comparison of three weighted human gene functional association networks

Jing Zhao* , Chun-Lin Wang, Ting-Hong Yang, Bo Li, Xing Chen, Xiaona Shen, Ling Fang

Department of Mathematics
Logistical Engineering University
Chongqing, China
**Email:* zhaojanne@gmail.com

Abstract—**Gene-gene association or protein-protein interaction databases have been important resource for the study of cellular functions and human diseases. A number of gene association databases have been available in the public domain. Each of these databases has its own unique virtues, but no single database could provide enough confidence and coverage. These years some meta-databases have been built by integrating various resources of gene functional associations and weighing the evidence of each association by some score systems. In this work, we compared three weighted genome-scale human gene association networks constructed from three such meta-databases, STRING, FunCoup and FLN, respectively. We found that the three networks share a large fraction of common genes but only quite limited overlapped interactions. However, most genes involved in important cellular processes and human diseases, as well as their pairwise interactions, is included in all of the three networks. This explains why all the three networks have been successfully applied in the study of cellular functions and diseases mechanisms. We believe that further integration of these meta-databases would provide higher confidence and coverage of gene associations in human proteome and facilitate the study of human gene association networks.**

Keywords- gene-association network, confidence score, protein complex, pathway, disease-related gene

I. INTRODUCTION

Genes and proteins seldom carry out their functions alone. Instead, a group of genes or proteins usually interact and communicate with each other to perform a particular cellular task. Functional associations between genes or proteins, including direct physical binding and indirect interaction such as participation in the same cellular process, thus play important roles in cellular phenotypes. Gene associations in an organism could be represented as a network called gene-gene association network or protein-protein interaction network, in which nodes and links correspond to genes (proteins) and their interactions, respectively[1, 2]. Studies of these networks have shed light on function and evolution of cellular systems, as well as revealed the interplay between gene functional associations and diseases[3-8]. Many databases have been set up to collect gene-gene (or protein-protein) associations from different resources and methods, such as experiments, literature mining and computational prediction. HPRD[9], MINT[10], BIND[11], Reactome[12], KEGG[13], MetaCyc[14] and ArrayProspector[15] are examples of such databases. Each of these databases has its own unique characters, while no single database could provide enough confidence and coverage. Hence there have been some efforts to construct meta-databases by integrating various resources of gene functional associations and weighing the evidence of each association by some score systems[16-18]. Weighted gene-association or protein-protein interaction networks reconstructed from these meta-databases have shown great power in the investigation of complex cellular systems, human diseases and drug intervening[17-20]. Here we conduct a network comparison on three of such meta-databases, STRING[16], FunCoup[17] and FLN[18].

II. MATERIALS AND METHODS

A. Human functional association meta- databases

The three human functional association datasets under study are STRING [16] (Search Tool for the Retrieval of Interacting Genes/Proteins) , FunCoup [17](Functional Coupling) and FLN [18](Functional Linkage Network). All of them are meta-databases constructed by aggregating both physical and functional interactions between human proteins available from numerous sources and integrating the data by their own scoring systems to weigh the evidence of each association.

B. Gene database of human genome

We downloaded the gene database of human genome [21] from the NCBI ftp on June 21, 2011. This database includes 45526 genes (Entrez gene ID), in which 21417 are protein-coded genes.

C. Data of functional modules

Biological functional modules we studied included known human protein complexes and signalling pathways. For the list of human protein complexes, we use the Comprehensive Resource of Mammalian protein complexes (CORUM) database [22], where 1343 complexes and 2315 component proteins are listed in total as a core data. Signaling pathways were downloaded from the C2: CP collection of MSigDB database [23], which were curated from several online pathway databases. A total of 6804 genes involved in 880 distinct pathways were included in this collection.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

D. Data of disease-related genes

We collected genes related with diseases from several databases including Immunome[24], InnateDB[25], OMIM (the Online Mendelian Inheritance in Man)[26], and DrugBank[27]. Immunome database collects 893 genes of human immune system. InnateDB database includes 2339 innate immunity-relevant human genes manually collected by literature review (April, 2011). A total of 2651 disease genes were collected from the morbid map of the OMIM database, while 1372 drug targets for FDA approved human drugs were collected from the DrugBank database.

E. Edge density

Edge density of a network is the ratio of the number of edges to that of a completely connected network with the same number of nodes:

$$\text{Density} = \frac{A}{N(N-1)},$$

Where A, N are the numbers of edges and nodes in this network, respectively.

F. Node degree and strength

The degree of a node is the number of edges linked with the node. In a weighted network, the strength of a node is the total weights of the edges linked with the node.

III. Results and Discussion

A. Description of the human functional association networks

The three human functional association datasets under study are STRING [16] , FunCoup [17] and FLN [18]. All of them are meta-databases constructed by aggregating both physical and functional interactions between human proteins available from numerous sources and integrating the data by their own scoring systems to weigh the evidence of each association. See Table 1 for basic information of the three datasets.

TABLE I. Basic Information of the Human Functional Association Datasets Under Study and the Networks Constructed From Them

	Information	STRING	FunCoup	FLN
Datasets	*URL*	http://string-db.org	http://funcoup.sbc.su.se	http://genomebiology.com/2009/10/9/R91/additional
	Gene code	Ensembl protein code	Ensembl gene code	Entrez gene code
	Genes	17,369	17,150	21657
	interactions	1,288,886	2,290,853	22,388,609
	Range of confident score	150~999	0.018~1	0.0427~19.0321
Networks	*Genes*	14,520	14,689	18,468
	Links	1,031,691	1,803,866	19,988,965
	Proteome coverage of genes	67.8%	68.6%	86.2%
	Range of normalized link weights	0.15~1	0.018~1	0.0022~1

The three datasets use different code systems for genes. According to the Ensembl database [28], we converted Ensembl protein codes in the STRING and Ensembl gene codes in the FunCoup into entrez gene codes, respectively. Then we mapped the entrez gene codes of the three datasets onto the NCBI gene database of human genome[21] and only kept mapped protein-coded genes. Thus the STRING, FunCoup and FLN datasets include 14520, 14689 and 18468 protein-coded genes, taking 67.8%, 68.6% and 86.2% of human genome protein-coded genes, respectively. Utilizing these genes and their linkages in the three datasets, we constructed three weighted functional association networks of human genome, named as network STRING, FunCoup and FLN, respectively. As shown in Table 1, the link weights of the three datasets vary in different areas. Dividing each weight by the maximum in the corresponding dataset, we normalized the weights into the area (0, 1].

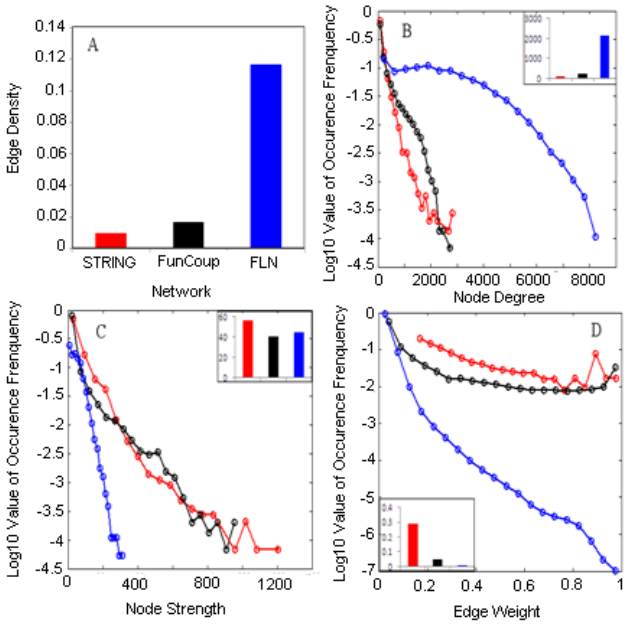

Figure 1. Comparisons of the basic features of the three human functional association networks. (A) Edge densities; (B) Distributions of node degree; (C) Distributions of node strength; (D) Distributions of edge weight. In (B), (C) and (D),logarithmic (base 10) scale is used for the Y-axis; the colors have the same meanings as in (A), while inserted histograms are comparisons of average values of the corresponding measures.

Table 1 and Figure 1 show basic information and features of the three networks. It can be seen that network FLN has much more nodes and links than the other networks, with tenfold scales of network density and average node degree. For all the networks, the semi-log scale plots for the distribution functions of node degree, node strength and edge weight are decreasing curves, suggesting that large fractions of nodes and edges in these networks own small degrees, strength, and edge weights, respectively. That is, only small

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

parts of genes interact with great number of partners, while small fractions of interactions have high confidence scores. However, the distribution curves for measures of network FLN exhibit significantly different patterns with those of the two others. This network has broader profile of node degrees, and smaller range of node strength. Its curve of node degree distribution has a much gentler slope, whereas the curves for node strength and edge weight distributions decrease much sharply. This phenomenon suggests that although network FLN has much more links than the other networks, a great fraction of them has low confidence score. In fact, the edge weights of about 90% links are smaller than 0.05. Although network STRING and FunCoup have similar distribution curves for the three measures, network STRING has much larger average node strength and link weight, but smaller average node degree, suggesting this network includes more links with high confidence. Specifically, the smallest edge weight of network STRING is 0.15, whereas respectively about 90% and 60% links in networks FLN and FunCoup have edge weights smaller than 0.05. In summary, although the three networks include different numbers of interactions, network with fewer links include larger part of interactions with higher confidence.

B. Overlap of nodes and links between networks

When checking the nodes and links, we found that the three networks had 12520 and 95655 common genes and interactions, respectively. Although the networks have a large fraction of common genes (taking 86.2%, 85.2% and 67.8% of the total in networks STRING, FunCoup and FLN, respectively), there is quite limited fraction of common interactions (only taking 9.3%, 5.3% and 0.5% of the total in networks STRING, FunCoup and FLN, respectively). The situation is the same for overlaps between every two networks (Table 2).

TABLE II. OVERLAPS OF NODES AND EDGES BETWEEN EVERY TWO NETWORKS

Edges / Nodes	STRING	FunCoup	FLN
STRING	-	12,556	14,402
FunCoup	110,094	-	14,578
FLN	626,806	776,469	-

As can be seen in Figure 1(C), the node strength profiles of the three networks are significantly different. To compare the distributions of node strengths of the common genes in the networks according to their important levels in the networks, in each network, we converted the node strength to its rank among all node strengths of the network. Then we grouped the nodes into 10 bins from the top of ranks to the bottom, in which each bin includes 1/10 nodes. At last we counted the occurrence frequency of the common genes in each bin. As shown in Figure 2 (A), the distributions of node strength ranks of the three networks are clearly decreasing, in which that of FLN decreases most sharply and that of FunCoup most gently. This distribution suggests that a large fraction of important genes, *i. e*, genes own more high-confident interactions in each network, appear in the common gene set.

In Figure 2 (B) we show the distributions of edge weights of common interactions. It can be seen that peaks of high-confident interactions are present in networks STRING and FunCoup, respectively. On the contrary, the distribution of edge weights of common genes in network FLN is monotonously decreasing, suggesting much more low-confident interactions of this network overlap with interactions of the other two networks. In summary, the common interaction set includes significantly large fractions of high-confident pairs in networks STRING and FunCoup and low-confident pairs in network FLN.

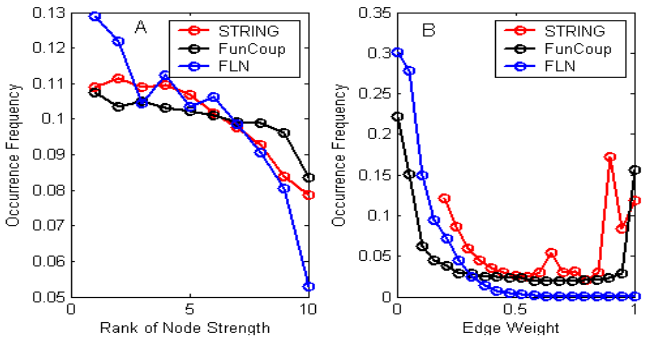

Figure 2. Distributions of the common genes and interactions in the three human functional association networks. (A) Distributions of node strength ranks of common genes. (B) Distributions of edge weights of common interactions.

C. Functional modules exhibited in the networks

In has been known that, in cells, a group of genes or proteins usually collaborate with each other, forming a specific functional module so as to carry out a particular cellular task. Specifically, certain proteins physically interact with each other to form stable structural and functional units called protein complex[29]; while certain transcription factors regulate a group of target genes to coordinate cellular activities related to a particular biological process [30]. On the other hand, studies of biological networks have revealed that functional modules usually correspond to densely linked topological sub-networks of the global network[31, 32]. In this section, we investigate whether known human protein complexes and pathways are presented in the three networks and how proteins in these functional modules are linked.

TABLE III. DISTRIBUTIONS OF GENES OF FUNCTIONAL MODULES IN THE THREE NETWORKS AND THEIR LINKS

	Average percentage of module genes in networks			Average link density of modules in networks		
	STRING	FunCoup	FLN	STRING	FunCoup	FLN
Protein complexes	90.7%	97%	100%	0.86	0.77	0.98
Pathways	88%	94%	99%	0.59	0.30	0.74

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

A total of 1343 protein complexes and 880 pathways were extracted from the Comprehensive Resource of Mammalian protein complexes (CORUM) database [22] and the C2: CP collection of MSigDB database [23], respectively. As for the complexes, we only studied the 938 ones which consist of at least three component proteins. For each complex and pathway, we counted how many of its genes appear in each of the three networks and how many links among the genes were presented in these networks. As listed in Table III, most genes of the functional modules under study are presented in the three networks. Specifically, network FLN includes all complex genes and 99% pathway genes, while FunCoup includes more genes of functional modules than STRING. Furthermore, the average link densities of functional module genes in the three networks are significantly much higher than link densities of the three global networks (see Figure 1A for link densities of the three networks.), in which functional module genes are linked most densely in FLN and most sparsely in FunCoup.

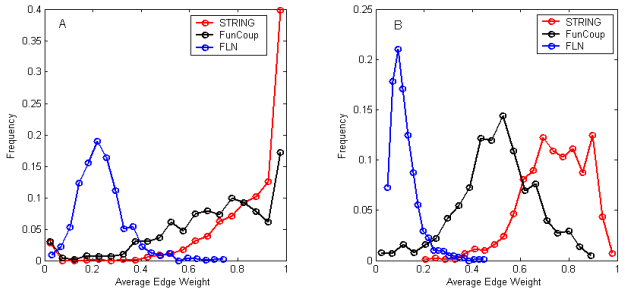

Figure 3. Distributions of average edge weights of protein complexes (A) and pathways (B) in the three networks.

Taking the confident levels of links into consideration, we calculated the average edge weight of each functional module in each of the three networks. In Figure 3 we show the distributions of average edge weights of complexes and pathways in the three networks. It can be seen that STRING includes most high-confident links between functional module genes. Actually, average edge weights of 92.3% complexes and 86.8% pathways in this network are at least 0.6. However, although FLN includes most functional module genes and their links, the links exhibit lower confident levels compared with those in the other two networks.

D. Disease-related genes in the networks

Interactions between genes and proteins have been known deeply involved in the pathogenesis of diseases. Recent network-based approaches have demonstrated great success in the application of functional relationships among genes for understanding human diseases[7]. Here we investigate how many genes involved in the occurrence, development and intervening of diseases are presented in the three networks and the important levels of these genes in the networks.

TABLE IV. PERCENTAGES OF GENES IN THE FOUR DISEASE-RELATED GENE DATASETS THAT OVERLAP WITH GENES IN THE THREE NETWORKS AND THEIR COMMON GENES

Networks / Datasets	STRING	FunCoup	FLN	Common genes in three networks
Immunome	85.3%	86.9%	99.2%	78.7%
InnateDB	88.8%	95.6%	99.7%	86.2%
OMIM	87.6%	90.3%	98.6%	82%
DrugBank	88.7%	91%	99.1%	83%

Genes related with diseases were collected from Immunome[24], InnateDB[25], OMIM (the Online Mendelian Inheritance in Man)[26], and DrugBank[27] database, which include human immune system genes, innate immunity-relevant human genes, disease causative genes, and drug targets for FDA approved human drugs, respectively. Mapping these genes into each of the three networks and their common gene set, we found that most of the disease-related genes under study appear in the three networks and their common gene set, in which FLN includes almost all the disease-related genes (see Table IV). In Figure 4 we show how node strength ranks of Immunome, InnateDB, OMIM and DrugBank genes distribute in the three networks. In most cases, the node strength ranks of genes in the four disease-related gene sets exhibit decreasing pattern, suggesting that large fractions of disease-related genes have many high-confident interactions in each network. It is noted that the distributions of Immunome, OMIM and DrugBank genes in FunCoup network are slightly different with those in the other two networks. In this network, the frequencies of top ranked genes are less than those of medium ranked genes.

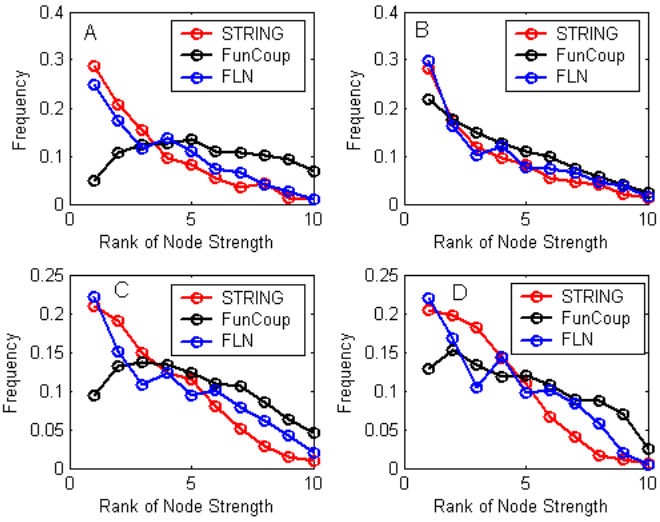

Figure 4. Distribution of node strength ranks of Immunome (A), InnateDB (B), OMIM (C) and DrugBank (D) genes in the three networks.

IV. CONCLUSIONS

In this work, we compared three weighted genome-scale human gene association networks constructed from three meta-databases STRING, FunCoup and FLN, respectively. These meta-databases were constructed by aggregating physical and functional interactions between human proteins available from numerous sources, but different methodologies and input data were utilized to produce its own confidence scores for weighing the evidence of each

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

association. It was found that genes in network STRING and FunCoup cover roughly the same fraction of human proteome, while FLN includes much more genes and interactions. The three networks own a fairly large fraction of common genes, but the overlap in terms of interactions is quite limited. When taking normalized confidence scores into account, STRING was found to include much higher fraction of high-confident interactions than the other two networks. Considering the application of gene association networks in the research of cellular functions and human diseases, we also explored the distributions of biologically important genes, i.e., genes of biological functional modules, and disease-related genes, in the networks. Most genes of known protein complexes and pathways were found presented in the three networks and they form densely connected topological modules in all the networks, in which STRING includes most high-confident links between genes of these functional modules. Furthermore, most genes known to be involved in the occurrence, development and intervening of diseases are presented in the three networks, and large fractions of these disease-related genes own many high-confident interactions in each network. Therefore, although the three networks only share a few overlapped interactions, information concerning genes involved in important cellular processes and human diseases, as well as their interactions, is included in all of the three networks. This explains why all the three networks have been successfully applied in the study of cellular functions and diseases mechanisms. It is expected that some computational approach could be developed to further aggregate these meta-databases for providing higher coverage of gene associations in human proteome.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (10971227).

REFERENCE

[1] D. Ucar, I. Neuhaus, P. Ross-MacDonald, C. Tilford, S. Parthasarathy, N. Siemers, and R.-R. Ji, "Construction of a reference gene association network from multiple profiling data: application to data analysis," *Bioinformatics*, vol. 23, pp. 2716-2724, 2007.

[2] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal, "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, pp. 1173-1178, 2005.

[3] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, "Evidence for dynamically organized modularity in the yeast protein-protein interaction network," *Nature*, vol. 430, pp. 88-93, 2004.

[4] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, "Evolutionary Rate in the Protein Interaction Network," *Science*, vol. 296, pp. 750-752, 2002.

[5] J. Xu and Y. Li, "Discovering disease-genes by topological features in human protein-protein interaction network," *Bioinformatics*, vol. 22, pp. 2800-2805, 2006.

[6] J. Zhao, C. Geng, L. Tao, D. Zhang, Y. Jiang, K. Tang, R. Zhu, H. Yu, W. Zhang, F. He, Y. Li, and Z. Cao, "Reconstruction and Analysis of Human Liver-Specific Metabolic Network Based on CNHLPP Data," *Journal of Proteome Research*, vol. 9, pp. 1648-1658, 2010.

[7] Z. P. Liu, Y. Wang, X. S. Zhang, and L. Chen, "Network-based analysis of complex diseases," *IET Systems Biology*, vol. 6, pp. 22-33.

[8] J. Zhao, P. Jiang, and W. Zhang, "Molecular networks for the study of TCM pharmacology," *Briefings in Bioinformatics* vol. 11, pp. 417-430, 2010.

[9] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, "Human Protein Reference Database--2009 update," *Nucleic Acids Res* vol. 37, pp. D767-772, 2009.

[10] A. Zanzoni, L. Montecchi-Palazzi, M. Quondam, G. Ausiello, M. Helmer-Citterich, and G. Cesareni, "MINT: a Molecular INTeraction database," *FEBS Letters*, vol. 513, pp. 135-140, 2002.

[11] G. D. Bader, D. Betel, and C. W. V. Hogue, "BIND: the Biomolecular Interaction Network Database," *Nucleic Acids Research*, vol. 31, pp. 248-250, 2003.

[12] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein, "Reactome: a knowledgebase of biological pathways," *Nucleic Acids Research*, vol. 33, pp. D428-D432, 2005.

[13] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res.*, vol. 28, pp. 27-30, 2000.

[14] C. J. Krieger, P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, and P. D. Karp, "MetaCyc: a multiorganism database of metabolic pathways and enzymes," *Nucleic Acids Research*, vol. 32, pp. D438-D442, 2004.

[15] L. J. Jensen, J. Lagarde, C. von Mering, and P. Bork, "ArrayProspector: a web resource of functional associations inferred from microarray expression data," *Nucleic Acids Research*, vol. 32, pp. W445-W448, 2004.

[16] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, and C. von Mering, "STRING 8- a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, pp. D412-D416, 2009.

[17] A. Alexeyenko and E. L. L. Sonnhammer, "Global networks of functional coupling in eukaryotes from comprehensive data integration," *Genome Research*, vol. 19, pp. 1107-1116, 2009.

[18] B. Linghu, E. Snitkin, Z. Hu, Y. Xia, and C. DeLisi, "Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network," *Genome Biology*, vol. 10, pp. R91, 2009.

[19] J. Zhao, T.-H. Yang, Y. Huang, and P. Holme, "Ranking Candidate Disease Genes from Gene Expression and Protein Interaction: A Katz-Centrality Based Approach," *PLoS ONE*, vol. 6, pp. e24306.

[20] J. Zhao, J. Chen, T.-H. Yang, and P. Holme, "Insights into the pathogenesis of axial spondyloarthropathy from network and pathway analysis," *BMC Systems Biology*, vol. 6, pp. S4.

[21] D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, W. Helmberg, Y. Kapustin, D. L. Kenton, O. Khovayko, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, K. D. Pruitt, G. D. Schuler, L. M. Schriml, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, T. O. Suzek, R. Tatusov, T. A. Tatusova, L. Wagner, and E. Yaschenko, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Res.*, vol. 34, pp. D173-180, 2006.

[22] A. Ruepp, B. Waegele, M. Lechner, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, and H. W. Mewes, "CORUM: the comprehensive resource of mammalian protein complexes 2009," *Nucleic Acids Research*, vol. 38, pp. D497-D501, 2010.

[23] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc Natl Acad Sci U S A*, vol. 102, pp. 15545–15550, 2005.

[24] C. Ortutay and M. Vihinen, "Immunome: A reference set of genes and proteins for systems biology of the human immune system," *Cellular Immunology*, vol. 244, pp. 87-89, 2006.

[25] D. J. Lynn, G. L. Winsor, C. Chan, N. Richard, M. R. Laird, A. Barsky, J. L. Gardy, F. M. Roche, T. H. W. Chan, N. Shah, R. Lo, M. Naseer, J. Que, M. Yau, M. Acab, D. Tulpan, M. D. Whiteside, A. Chikatamarla, B. Mah, T.

Munzner, K. Hokamp, R. E. W. Hancock, and F. S. L. Brinkman, "InnateDB: facilitating systems-level analyses of the mammalian innate immune response," *Mol Syst Biol*, vol. 4, 2008.
[26] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, "Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders," *Nucleic Acids Res* vol. 33, pp. D514-517, 2005.
[27] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Res*, vol. 34, pp. D668-672, 2006.
[28] T. J. P. Hubbard, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney, "Ensembl 2007," *Nucleic Acids Research*, vol. 35, pp. D610-D617, 2007.
[29] A. D. King, N. Prulj, and I. Jurisica, "Protein complex prediction via cost-based clustering," *Bioinformatics*, vol. 20, pp. 3013-3020, 2004.
[30] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson, "Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle," *Molecular Biology of the Cell*, vol. 11, pp. 369-391, 2000.
[31] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical Organization of Modularity in Metabolic Networks," *Science*, vol. 297, pp. 1551-1555, 2002.
[32] J. Zhao, H. Yu, J.-H. Luo, Z.-W. Cao, and Y.-X. Li, "Hierarchical modularity of nested bow-ties in metabolic networks," *BMC Bioinformatics*, vol. 7, pp. 386, 2006.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE