STRING PPI Score to Characterize Protein Subnetwork Biomarkers for Human Diseases and Pathways

Prayas Timalsina, Kevin Charles, and Ananda Mohan Mondal*

Department of Mathematics and Computer Science,

Claflin University, Orangeburg, SC 29115

Corresponding Author: amondal@claflin.edu

Abstract – Protein subnetwork biomarkers for 144 diseases and pathways are analyzed in terms of protein-protein interaction (PPI) score available in STRING database. Most of the subnetwork biomarker (SNB) studies are to classify disease samples from the control. But no de novo algorithm is available to identify SNB from the whole genome PPI network without the knowledge of differentially expressed genes. Recently, based on mouse model, researchers showed that there exists a dynamical network biomarker which can distinguish among the normal state, pre-disease state, and disease state of a disease progression. But, most of the gene expression data for human diseases are at the disease state. No data is available for the first two stages of a disease. Understanding the network behavior of a disease at disease state might help in the development of de novo algorithm for predicting protein SNBs not only for disease state but also for early stages of a disease or early warning signals. PPI score in STRING database represents a rough estimate of how likely a given interaction describes a functional linkage between two proteins. So, analyzing protein SNB for human diseases at disease state with respect to PPI score may shed some light in the development of de novo models for predicting SNB.

A simple brute force approach is used to isolate the SNB for a disease or pathway from the genome-wide PPI network by projecting the corresponding differentially expressed proteins. Then the SNBs are analyzed in terms of PPI score. Our investigation shows that higher is the PPI score of a network is more likely to produce a true SNB for a disease. Results also show that Physical PPIs with high score are more capable of producing a true SNB.

Keywords- biomarker; brute force method; PPI biomarker; PPI score; single protein biomarker; subnetwork biomarker.

I. INTRODUCTION

In general, a biomarker or marker is a gene or group of genes that represent a certain phenotype or disease. Biomarkers are important in the study of disease and drug design. If the biomarkers for a disease are known then drug can be designed to control the activity of biomarkers, thus controlling the disease. Usually, genes expressed differentially are considered as single gene markers (SGMs). Studies show that sets of SGMs determined by differential expression vary considerably when inferring them from different platforms thus making them useless in cross-platform studies [1]. Chuang et al. [2] showed that multigenetic markers can be used to address this issue. Multigenetic markers consist of several differentially expressed genes which also form a connected component in a protein-protein interaction (PPI) network thus giving the name subnetwork biomarkers. Subnetworks are significant because, in contrast to individual proteins, they provide concrete hypotheses as to the molecular complexes, signaling pathways, and other mechanisms that impact the disease outcome [3].

With the recent development of high-throughput experiment to determine protein-protein interaction both physical and genetic, PPI networks are increasingly serving as tools for discovering the molecular basis of disease. In a review by Ideker and Sharan [3], authors enumerated four different prospective applications of PPI networks to disease, namely: identifying new disease genes; the study of their network properties; identifying disease-related subnetworks; and network-based disease classification. Most of the studies on disease-related subnetworks are to provide better classification of disease samples from the control [2, 4, 5].

Recently, researchers enumerated three stages of disease progression [6-10]: *normal state*, which is a stable state where the system undergoes gradual or slow change; *pre-disease state*, which is a limit of the normal state just before the drastic transition to the disease; and *disease state*, which is another stable state after the critical transition, where the system is considered to be seriously damaged. Based on mouse model, they showed that there exists a dynamical network biomarker (DNB) which is capable of distinguishing among three stages of disease progression [6-10]. But, for human diseases, most of the gene expression data are at the disease state (3rd stage of disease progression). No data is available for the 1st and 2nd stages of a disease. So, a model for predicting disease network at early stages would help not only in taking protective measure but also in drug design. Understanding the network behavior of a disease in disease state might help in developing the predictive model for early disease states (normal state and pre-disease state). The main objective of this study is to understand and characterize protein subnetwork biomarkers at disease state.

Researchers [11] found that disease genes exhibit an increased tendency for their protein products to interact with one another, tend to be co-expressed in specific

978-1-4799-7502-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BIBE.2014.46

tissues, and display coherent functions with respect to all three branches of the Gene Ontology hierarchy [12]. PPI scores available in STRING database represent the confidence score between two proteins to have similar functions [13]. So, it would be interesting to see the pattern or correlation of disease genes in terms of PPI score.

II. DATASETS PREPARATION

Two sets of data, namely i) list of biomarkers or single protein biomarkers (SPBs) and ii) protein-protein interaction data are required for identifying protein subnetwork biomarkers for a disease. In the present study, we considered human genome to identify subnetwork biomarkers related to different diseases. The list of biomarkers are procured from SABiosciences of Qiagen [14]. Two genome-wide PPI networks for human are used in the present work: one non-scored PPI network from BioGRID database [15] and one Scored PPI network from STRING database [13].

A. Single Protein Biomarkers

Biomarkers or single protein biomarkers (SPBs) for 144 human diseases and pathways are procured from SABiosciences of Qiagen [14].

Table I shows a sample of 10 disease and pathway names. Each disease or pathway is associated with 84 key genes commonly involved in the dysregulation of signal transduction and other normal biological processes during disease or pathway [14]. There are total of 4584 genes associated with 144 diseases and pathways. Number of disease associated with a gene ranges from 1 through 51. 2396 genes are associated with only one disease, 909 genes are associated with two diseases. On the other extreme, gene TNF is associated with as most as 51 diseases, followed by gene IL6 with 48 diseases, followed by gene VEGFA with 44 diseases, etc.

TABLE I. A SAMPLE OF 10 DISEASES AND PATHWAYS

Sl #	Disease or Pathway
1	Adherens_Junction
2	Adipogenesis
3	Allergy_and_Asthma
4	Alzheimer_disease_GE
5	Amino_Acid_Metabolism_I
14	Breast_Cancer
35	Diabetes
37	DNA_Repair
143	WNT_Signaling_Targets
144	Wound_Healing

B. BIOGRID PPIs

Version 3.2.102 is used for this study. Total number of PPIs is 217215 of which 1665 are genetic PPIs and 215550 are physical PPIs. For this study, we used physical PPIs only. After removing duplicate and self-interacting PPIs, we are left with 132293 physical PPIs composed of 15527 proteins. So, on an average, there are 9 interactions per protein.

C. STRING PPIs

Original PPI dataset, downloaded from STRING database version 9.0, contains 3,281,414 PPIs. For the present study, direction of interaction is not important. After removing direction and some erroneous data (860 in total: some are missing scores, some do not conform to STRING names etc.), final dataset contains 1,640,129 PPIs composed of 18,595 proteins.

STRING PPIs do not come with official protein names but disease proteins procured from Qiagen [14] are in official protein names. A mapping between STRING and official protein names is required. Another file from STRING database with GO annotation contains both STRING and official protein names, which is used as the mapping file. Original mapping file contains 17919 unique records. After cleaning some erroneous data (some protein names are in numbers i.e., not in official protein names), we are left with 17839 unique records. Finally, STRING PPI names are converted to PPIs in official protein names and working network is composed of 1568065 PPIs and 16614 proteins. So, on an average, there are 94 interactions per protein.

D. BIOGRID vs. STRING PPIs

Table II shows the topology of cleaned PPI networks obtained from BIOGRID and STRING database. BIOGRID network is composed of 15527 proteins and 132293 PPIs with an average degree per node of 17, and STRING network is composed of 16614 proteins and 1568055 PPIs with an average degree of 189. So, STRING network is highly connected compared to BIOGRID network. A third network, named COMMON, is derived from the intersection of BIOGRID and STRING PPIs. COMMON network is composed of 9485 proteins and 52817 PPIs with an average degree of 11. It is noticeable that 40% of BIOGRID PPIs and only 3.4% of STRING PPIs are in common.

TABLE II. TOPOLOGY OF BIOGRID AND STRING PPI NETWORKS.

Network	# PPI	# Protein	Avg Deg
BIOGRID	132293	15527	17
STRING	1568055	16614	189
COMMON	52817	9485	11

III. METHODOLOGY

A. Brute Force Method for Identifying Subnetwork Biomarkers

The general idea of computing subnetwork biomarkers, for example biomarkers for cancer, is to search for combinations of genes which (i) are sufficiently differentially expressed in the cancer tissue samples from gene expression training data and (ii) form a connected pattern in the PPI network [16]. A simple Brute Force method [17] to identify subnetwork biomarkers related to 144 human diseases and pathways is developed using the definitions given in this section.

Single Protein Biomarker (SPB): A protein that is sufficiently differentially expressed in the tissue of a patient with the disease or phenotype is called a single protein biomarker for the disease or phenotype.

PPI Biomarker (PPIB): A PPI composed of two SPBs.

True PPIB: A PPIB is a true PPIB when both proteins of the PPIB are associated with the same disease, phenotype or biological process.

Pseudo PPIB: A PPIB is a pseudo PPIB when two proteins of the PPIB are associated with two different diseases, phenotypes or biological processes.

Subnetwork Biomarker (SNB): A Subnetwork composed of PPIBs. So, by definition, an SNB is composed of one or more PPI biomarkers.

True SNB: An SNB composed of true PPIBs.

Pseudo SNB: An SNB composed of pseudo PPIBs.

The brute force method to identify SNBs from a genome-wide PPI network is developed employing bottom-up approach starting with SPBs. First, PPIBs are found from SPBs, then, SNBs are found from PPIBs just using the definitions given above.

B. Measuring Capability of Producing SNB

For practical purpose, we need an SNB that solely represents a disease, which is a true SNB composed of true PPIBs. The capability of producing true SNB by a network depends on the type of PPIs such as physical or genetic as well as PPI scores. A simple form of accuracy is defined in order to measure the capability of a PPI network in identifying protein subnetwork biomarkers.

$$Accuracy = \frac{Number\ of\ true\ PPIB}{Total\ number\ of\ PPIB} = \frac{Number\ of\ true\ PPIB}{Number\ of\ true\ PPIB + Number\ of\ pseudo\ PPIB}$$

It is noticeable that the accuracy defined here is different from usual classification accuracy, which depends on true positive, true negative, false positive and false negative.

IV. RESULTS AND DISCUSSION

Subnetwork biomarkers are identified using the brute force method described above and then analyzed in terms of PPI scores. PPI scores ranging from 150 to 999 available in STRING database [13] represent the confidence of how likely two proteins will have the similar function. So, understanding the characteristics of subnetwork biomarkers might help in the development of *de novo* algorithm for predicting subnetwork biomarkers without the knowledge of differentially expressed genes. We explored the characteristics of SNBs with respect to PPI score for combined SNBs considering all of the 144 diseases and pathways. It should be noted that there are three types of SNB in combined analysis: true SNB composed of true PPIBs for all diseases, pseudo SNB composed of pseudo PPIBs for all diseases, and combined SNB composed of true PPIBs and pseudo PPIBs for all diseases. In case of individual disease analysis, which will be included in extended version of this paper, only true SNBs are involved since by definition, pseudo and combined SNB do not exist for an individual disease.

A. Combined SNB for 144 Diseases and Pathways

PPIBs and SNBs are found for 144 diseases and pathways using three different PPI networks (BIOGRID, STRING, and COMMON) and SPBs for corresponding diseases and pathways. Table III summarizes the SNBs in terms of number of SPBs, number of PPIBs, accuracy, and average PPI score. Combined SNB is composed of pseudo PPIBs and true PPIBs, whereas true SNB is composed of only true PPIBs. BIOGRID PPIs do not come with PPI score and as such there are no average PPI scores for both combined SNB and true SNB.

TABLE III. COMBINED SNB AND ACCURACY

	Comb. SNB		**True SNB**			**Avg. PPI Score**	
Network	***SPB***	***PPIB***	***SPB***	***PPIB***	***Accuracy***	***Comb. SNB***	***True SNB***
BIOGRID	3744	30713	2451	8541	27.81%	---	---
STRING	4504	489019	4395	118898	24.31%	381	485
COMMON	3241	18379	2273	7381	40.16%	703	809

Coverage in terms of SPB: There are altogether 4584 SPBs (84 SPBs for each disease with some SPBs common in different diseases) for 144 diseases. STRING network contains majority of SPBs - 4504 with combined SNB and 4395 with true SNB, followed by BIOGRID network - 3744 with combined SNB and 2451 with true SNB. COMMON network has the lowest coverage as expected.

Network in terms of accuracy: It is clear from Table III that COMMON network is more capable of producing true SNB (40.16%) compare to individual networks, BIOGRID with 27.81% and STRING with 24.31%. It is also noticeable that accuracy has direct relation with the average

PPI score of SNB. COMMON network has higher PPI scores - 703 for combined SNB and 809 for true SNB, compare to STRING network - 381 for combined SNB and 485 for true SNB. It can be concluded that *higher is the PPI score of a network, it is more likely to produce a true SNB (**Conclusion-1**)*. In order to prove this hypothesis, further investigation is done in section IV-C. The similar observation was made by Charles et al. [17] in case of identifying SNB for biological roles in yeast. By definition, COMMON network is composed of physical PPI only, since it is the intersection of BIOGRID and STRING data and in this study, we considered only physical PPIs from BIOGRID. It is also clear that COMMON PPIs have high PPI scores. So, COMMON network is composed of physical PPIs with high scores. It can be concluded that *physical PPIs with high score are more capable of producing a true SNB (**Conclusion-2**)*. By definition, physical PPI means both proteins physically interact together to produce another product. So, physical PPI with high score means that two proteins are more likely to produce another product and as such they are more likely responsible for the disease caused by the product. Then question might appear which physical PPIs with high score are responsible for a particular disease. Further investigation is required to answer this question.

B. STRING Network and Derived SNBs

Since STRING network has the largest coverage in terms of SPBs as mentioned in section IV-A, it is better to do detail analysis of SNBs derived from it. Table-IV shows the topology of three different types of SNBs (combined SNB, pseudo SNB, and true SNB) derived from STRING network along with the original STRING network.

TABLE IV. TOPOLOGY OF STRING NETWORK AND DERIVED SNBS.

Network	PPI	Protein	Avg Degree	Avg Score	Comp
STRING	1568056	16614	189	381.09	1
Comb. SNB	489019	4504	217	381.05	1
Pseudo SNB	370121	4492	165	347.62	1
True SNB	118898	4395	54	485.10	1

It is noticeable that the average degrees per node for networks STRING, combined SNB, pseudo SNB are very high (189, 217, and 165 respectively) compare to network for true SNB (54 only). So, first three are highly connected networks compare to true SNB. Three types of SNBs (Combined, pseudo, and true) have about the similar number of proteins (4400 ~ 4500) but true SNB has very low average degree (54) compare to combined SNB (217) and pseudo SNB (165). This is expected since both proteins of a PPI for true SNB are related to the same disease. It can be concluded that, *for SNBs derived from a genome-wide PPI network, true SNB has the lowest average degree compare to combined SNB and pseudo SNB (**Conclusion-3**)*.

It is interesting that average PPI scores for STRING and combined SNB are about the same (381.0). This indicates that PPIs in combined SNB are nothing but the random sample of STRING PPIs. So, it can be concluded that *in identifying SNB from a genome-wide PPI network, combined SNB, which is the combination of true PPIBs and pseudo PPIBs, does not carry any weight (**Conclusion-4**)*. In other words, combined SNB should not be considered for identifying an SNB for a disease. Average PPI score for pseudo SNB is less than combined SNB (347 < 381) and much less than true SNB (347 << 485). By definition, pseudo SNB is composed of pseudo PPIBs and a pseudo PPIB is composed of two proteins associated with two different diseases. So, two proteins of a pseudo PPIB are more likely related to two different functions and as such they will have low confidence score of having similar function. On the other hand, true SNB is composed of true PPIB and two proteins in a true PPIB are related to the same disease. So, two proteins of a true PPIB are more likely to have similar function and as such true PPIBs are more likely to have high PPI score. It can be concluded that *true SNBs are composed of PPIs with high PPI scores compared to pseudo SNBs (**Conclusion-5**)*. This suggests that PPIs with high scores can be used as the seed for the *de novo* algorithm for identifying an SNB for a disease from a genome-wide PPI network.

Fig. 1 shows the cumulative distribution of frequency of PPI based on PPI score for different SNBs along with the original STRING network. It is clear that the frequency distributions for STRING and combined SNB are very similar. As a result, the average PPI scores for these two networks are almost the same (Table IV). This implies that *combined SNB is nothing but a network formed by a random sample of PPIs selected from the whole-genome PPI network (**Conclusion-6**)*. But the distribution for true SNB and pseudo SNB are very distinct and they are apart from each other. It is clear that in case of pseudo SNB, there are more PPIs with low scores compare to true SNB. For example, about 84% pseudo PPIBs have scores less than 500 compare to 63% of true PPIBs. In other words, true SNB has more PPIs with high scores compare to pseudo SNB. For example, about 18% true PPIBs have scores higher than 800 compared to 6% pseudo PPIBs. This is the reason that true SNBs have high average score compared to pseudo SNBs as shown in Table IV. True PPIB means that both proteins are associated with the same disease. Since majority of true PPIBs have high PPI scores then two proteins forming a true PPIB are more likely to have the same molecular function. So, a disease is occurring due to disruption/dysregulation of some molecular function.

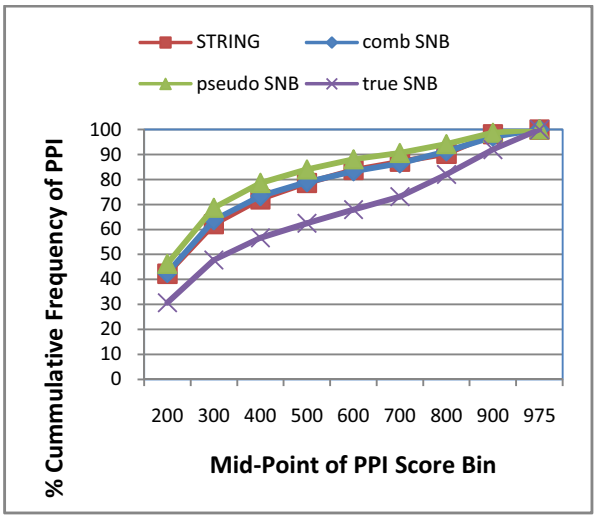

Figure 1. PPI score distribution of STRING network and SNBs derived from STRING. Bin size = 100.

C. *Effect of PPI Score in Identifying SNB*

Fig. 2 presents the accuracy produced as function of PPI score. Whole network is divided into smaller networks using a bin size of PPI score equals 50, 100, and 120. For each bin, subnetwork biomarkers were identified and then accuracy for the same was evaluated. It is evident that accuracy increases with the increase of PPI score. It is noticeable that for PPI scores ranging between 800 and 900, accuracy is less than the previous score bin for smaller bin sizes like bin size of 50 and 100. For example, for bin size 50, accuracy for this range is 30.70%, which is smaller than the accuracies for previous five bins (43.06%, 46.23%, 41.86%, 37.75%, and 31.65%). For bin size 100, accuracy for this range is 41.22%, which is smaller than the accuracy for previous bin (44.27%). As the bin size is increased accuracy goes up for this range of score, eventually, it becomes larger than the previous bin. For example, for bin size 120, accuracy for this range is 47% which is larger than the accuracy for the previous bin, 44%. So, it can be concluded that the *general trend is that accuracy or capability of producing a true SNB by a network increases with PPI scores* (***Conclusion-7***). A similar trend is observed by Charles et al. [17] (Fig. 4 of their paper) in the study of protein subnetwork biomarkers for yeast considering single protein biomarkers related to biological roles. By definition of accuracy, we can conclude that higher is the PPI score of a network, more is the capability of producing true SNB. Higher the PPI score, higher the confidence of interaction between two proteins [11]. If the two proteins are more likely to interact with each other, they are more likely to be localized at the same subcellular location [8, 9]. As a result, they are more likely to be associated with the same cellular role or phenotype or disease. This is the reason for which PPIs with high score are more capable of producing subnetwork biomarker with high quality. This also proves the hypothesis made in section IV-A that *higher is the PPI score of a network, it is more likely to produce a true SNB* (***Conclusion-1***).

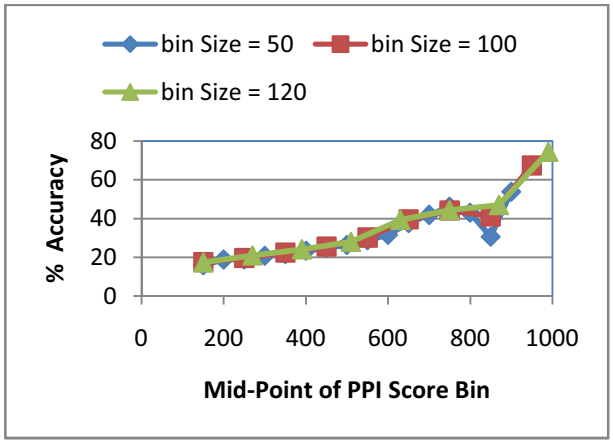

Figure 2. PPI scores on the performance of identifying true SNBs.

V. Conclusion and Future Remarks

A brute force method to identify subnetwork biomarkers (SNBs) from a genome-wide PPI network is developed employing bottom-up approach starting with single protein biomarkers (SPBs). First, PPI Biomarkers (PPIBs) are found from SPBs. Then SNBs are found from PPIBs. We explored the characteristics of SNBs with respect to PPI score for combined SNBs considering 144 diseases and pathways.

Our investigation shows that higher is the PPI score of a network is more likely to produce a true SNB for a disease ***(Conclusion-1)***. It also shows that physical PPIs with high score are more capable of producing a true SNB for a disease ***(Conclusion-2)***. So, the general trend is that accuracy or capability of producing a true SNB by a network increases with PPI scores ***(Conclusion-7)***. A similar trend is observed by another study of protein subnetwork biomarkers for yeast considering single protein biomarkers related to biological roles [17].

Combined SNB, composed of true PPIBs and pseudo PPIBs, is nothing but a subnetwork formed by a random sample of PPIs selected from the whole-genome PPI network (***Conclusion-6)***. So, in identifying SNB from a genome-wide PPI network, combined SNB does not carry any weight (***Conclusion-4)***. In other words, combined SNB should not be considered for identifying an SNB for a disease.

For SNBs derived from a genome-wide PPI network, true SNB has the lowest average degree compare to combined SNB and pseudo SNB ***(Conclusion-3)***. This property can be used in the development of *de novo* algorithm for identifying true SNB for a disease from the whole-genome PPI network.

Our results also indicate that true SNBs are composed of PPIs with high scores compared to pseudo SNBs ***(Conclusion-5)***. This suggests that PPIs with high scores can be used as the seed for the *de novo* algorithm for identifying true SNB for a disease from a genome-wide PPI network.

Present study considers combined SNB for 144 diseases and pathways. Further investigation is required for analyzing i) combined SNB for diseases only, ii) combined SNB for pathways only, iii) individual SNB for each disease and pathway, iv) which physical PPIs with high score are responsible for a particular disease, and v) functional enrichment of each SNB. We are investigating these aspects for including in the extended version of the paper.

ACKNOWLEDGMENT

This work was partially supported by NASA grant, Prime Award No: NNX12AI12A, Sub-award No: 520976-Claflin-Mondal and Center for Excellence in Teaching of Claflin University.

REFERENCES

[1] L. Ein-Dor, I. Kela, G. Getz, D. Givol, and E. Domany, "Outcome signature genes in breast cancer: is there a unique set?," Bioinformatics, vol. 21, pp. 171-178, 2005.

[2] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," Mol Syst Biol, vol. 3, p. 140, 2007.

[3] T. Ideker and R. Sharan, "Protein networks in disease," Genome Res, vol. 18, pp. 644-52, Apr 2008.

[4] J. Su, B. J. Yoon, and E. R. Dougherty, "Identification of diagnostic subnetwork markers for cancer in human protein-protein interaction network," BMC Bioinformatics, vol. 11 Suppl 6, p. S8, 2010.

[5] R. K. Nibbe, S. A. Chowdhury, M. Koyutürk, R. Ewing, and M. R. Chance, "Protein–protein interaction networks and subnetworks in the biology of disease," Wiley Interdisciplinary Reviews: Systems Biology and Medicine, vol. 3, pp. 357-367, 2011.

[6] L. Chen, R. Liu, Z. P. Liu, M. Li, and K. Aihara, "Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers," Sci Rep, vol. 2, p. 342, 2012.

[7] R. Liu, M. Li, Z.-P. Liu, J. Wu, L. Chen, and K. Aihara, "Identifying critical transitions and their leading biomolecular networks in complex diseases," Scientific reports, vol. 2, 2012.

[8] M. Li, T. Zeng, R. Liu, and L. Chen, "Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis," Brief Bioinform, Apr 25 2013.

[9] R. Liu, K. Aihara, and L. Chen, "Dynamical network biomarkers for identifying critical transitions and their driving networks of biologic processes," Quantitative Biology, vol. 1, pp. 105-114, 2013.

[10] X. Liu, R. Liu, X.-M. Zhao, and L. Chen, "Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers," BMC medical genomics, vol. 6, p. S8, 2013.

[11] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, "The human disease network," Proc Natl Acad Sci U S A, vol. 104, pp. 8685-90, May 22 2007.

[12] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, et al., "The Gene Ontology (GO) database and informatics resource," Nucleic Acids Res, vol. 32, pp. D258-61, Jan 1 2004.

[13] C. von Mering, L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, et al., "STRING: known and predicted protein-protein associations, integrated and transferred across organisms," Nucleic Acids Res, vol. 33, pp. D433-7, Jan 1 2005.

[14] (May 16, 2013). Biomarkers. Available: http://www.sabiosciences.com/Biomarker.php

[15] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets," Nucleic Acids Res, vol. 34, pp. D535-9, Jan 1 2006.

[16] P. Dao, R. Colak, R. Salari, F. Moser, E. Davicioni, A. Schonhuth, et al., "Inferring cancer subnetwork markers using density-constrained biclustering," Bioinformatics, vol. 26, pp. i625–i631, 2010.

[17] K. Charles, A. Afful, and A. M. Mondal, "Protein Subnetwork Biomarkers for Yeast Using Brute Force Method," in The 2013 International Conference on Bioinformatics and Computational Biology, BIOCOMP'13, Las Vegas, USA, 2013, pp. 218-223.