Network-based detection of Disease Modules and Potential Drug Targets in Intractable Epilepsy

Hongwei Chu[1], Xuezhong Zhou[2]*, Guangming Liu[2], Minghui Lv[1], Xiaofeng Zhou[1], Yiwei Wang[1], Lin Liu[2], Xing Li[2], Pin Sun[3], Yizhun Zhu[4], Changkai Sun[1]*

[1]Liaoning Provincial Key Laboratory of Cerebral Diseases, Institute for Brain Disorders
Dalian Medical University, Dalian, 116044, China

[2]School of Computer and Information Technology and Beijing Key Lab of Traffic Data Analysis and Mining
Beijing Jiaotong University, Beijing, 100044, China

*co-corresponding authors: Xuezhong Zhou: xzzhou@bjtu.edu.cn; Changkai Sun: cksun110@vip.sina.com

[3]Shanghai Medical College, Fudan University, Shanghai, 200032, China

[4]Department of Pharmacology, School of Pharmacy, Fudan University, Shanghai, 201210, China

***Abstract*—Epilepsy is one of the common nervous system diseases and a complex brain disease that severely damages the life and health of humans. One-third of all epilepsy patients have medically intractable epilepsy (IE), for which anti-epileptic drugs are not effective. Therefore, discovery of potential drug targets is urgent and meaningful for better clinical solutions. Using the IE terms from Medical Subject Headings (MeSH) terminology, we integrated literature-based disease-gene relationships, which were extracted from the CoreMine PubMed search engine system, protein-protein interactions (PPI) and drug-target relationships from heterogeneous data sources, and used the network medicine approach to identify disease modules and detect enriched pathways. The potential drug targets and the underlying mechanisms were confirmed by chemical-protein interaction network and published literatures. Using 23 IE MeSH terms, we manually searched the CoreMine system to obtain 1,400 disease-gene associations, which had 871 distinct genes from the PubMed database. With the help of the PPI database (i.e. String 9), we mapped the genes to the PPI network and utilized the BGL community detection method to find 33 disease-related topological PPI modules that contain 640 proteins and 2,483 links. After that, we used the enrichment analysis method to obtain the PPI modules with pathway and gene ontology enrichment. Finally, we confirmed nine significant PPI modules that are considered as epilepsy disease modules with significant functional signatures. We combined genes with drugs in the DrugBank database to confirm the four proteins, MT-CYB, UQCRB, UQCRC1 and UQCRH, which would be potential drug targets for IE. The results of this study demonstrated that integrated network data sources and network-based approach are useful to understand the molecular mechanism and extract potential drug targets for IE.**

Keywords—intractable epilepsy; network; module; Gene Ontology; KEGG pathway; drug target

I. INTRODUCTION

Epilepsy is one of the common neurological disorders; it has high prevalence and the causes are unknown in most cases [1]. Epileptic seizures are caused by episodic abnormal synchronous discharges in cerebral neuronal networks, leading to a paroxysmal brain disorder [2, 3]. Frequent and consistent epileptic seizures can cause rapid changes to various kinds of enzymes, neurotransmitters, chemical substances and amino acids, which can lead to brain cell death [3]. One-third of all epilepsy patients have medically intractable epilepsy (IE), which is difficult to control [4]. Apoptosis has been believed to be a major factor contributing to the formation of an abnormal excitatory circuit leading to refractory epileptic events [1]. Intractable epilepsies have a remarkable impact on cognitive and behavior function, and the treatment of seizures is still a challenge for clinicians [5]. Although there are some specific molecular targets in the brain, such as voltage-gated sodium and calcium channels, GABAA receptors, the GAT-1 GABA transporter and the synaptic vesicle protein SV2A [6], they are not effective for all cases due to the tremendous complexity underlying the molecular mechanisms of epilepsy. It is well recognized that interactome networks, which include the networks of genes, proteins and metabolites, build the fundamental mechanisms of diseases. Therefore, network medicine provides a novel approach to understanding the complicated mechanism of disease and treatment [7]. It is fundamental to understanding the mechanisms of complex diseases and the identification of potential drug targets in a pathway-centric perspective [8]. In this paper, by integrating literature-based disease-gene relationships, we filtered the protein-protein interaction (PPI) network relevant to epilepsy and tried to identify the disease modules and novel drug targets for IE.

There are several main calculating and manual steps to perform the disease module and drug target detection tasks. After confirming the medical subject headings (MeSH) [9] terms for IE, we used the CoreMine system (http://www.coremine.com/medical/#search) to obtain the disease-gene relationships. Then we got the PPI network of IE using the String 9.1 database [10]. The community detection method was used to detect the topological PPI modules relevant to IE. Next, we performed enrichment analysis to find the related gene functions and pathways of the modules using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) respectively. Finally, we used the drug-

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

target relationships from DrugBank [11] to confirm the novel drug targets that are not covered by current epilepsy drugs.

II. RESULTS

A. MeSH terms of IE subtypes

We searched the MeSH vocabulary using epilepsy as keyword and manually evaluated the MeSH terms associated with IE. In total, we confirmed 23 MeSH disease terms that are used to describe the subtypes of IE. The results are shown in Table I. We found that diseases like frontal lobe epilepsy, tuberous sclerosis and fragile X syndrome, which are the classical subtypes of IE, are included in our term list.

TABLE I. 23 MESH TERMS CORRESPONDING TO THE SUBTYPES OF IE

NO.	MeSH disease terms	Unique ID
1	Alexander Disease	D038261
2	Anti-N-Methyl-D-Aspartate Receptor Encephalitis	D060426
3	Classical Lissencephalies and Subcortical Band Heterotopias	D054221
4	Epilepsies, Myoclonic	D004831
5	Epilepsy, Absence	D004832
6	Epilepsy, Frontal Lobe	D017034
7	Epilepsy, Temporal Lobe	D004833
8	Fragile X Syndrome	D005600
9	Lafora Disease	D020192
10	Landau-Kleffner Syndrome	D018887
11	Leigh Disease	D007888
12	Lissencephaly	D054082
13	Malformations of Cortical Development	D054220
14	MELAS Syndrome	D017241
15	MERRF Syndrome	D017243
16	Mitochondrial Encephalomyopathies	D017237
17	Myoclonic Epilepsies, Progressive	D020191
18	Myoclonic Epilepsy, Juvenile	D020190
19	Neuronal Ceroid-Lipofuscinoses	D009472
20	Rett Syndrome	D015518
21	Spasms, Infantile	D013036
22	Tuberous Sclerosis	D014402
23	Unverricht-Lundborg Syndrome	D020194

B. Literature-based disease-gene relationships

Using the 23 MeSH terms for IE, we manually searched CoreMine in October 2013 to obtain the disease-gene relationships. After filtering the relationships with significant correlations (i.e. p-value<0.05), we got 3,219 disease-gene relationships with the corresponding number of occurrences in the PubMed database. To evaluate the reliability of these disease-gene relationships extracted from literature, we randomly selected 230 disease-gene relationship records, including 130 relationships from 1,819 relationships with only one literature occurrence and 100 relationships in the remaining 1,400 relationships whose occurrence in literature was larger than one. Then we evaluated these 230 disease-gene relationships by manually checking the PubMed database. We found that there are 85 (85/230=37%) false positive relationships that actually have no direct evidence from the corresponding literature. In particular, there are 71 (71/130=55%) relationships that have a literature occurrence frequency equal to one. Therefore, we eliminated all the 1,819 relationships. Finally, we obtained 1,400 relationships and 871 disease-associated genes. The number of genes related to each MeSH term is shown in Table II. To further evaluate the quality of the extracted disease-gene relationships, we checked the phenotype-genotype associations related to epilepsy in the OMIM database [12]. We obtained 136 disease-gene relationships with 84 associated genes, in which there are 68 (68/84=81%) genes covered by the 3, 219 relationships and 52 (52/84=62%) genes covered by the 1,400 relationships. The rest of the 16 genes are not covered by our results, mainly because the 2013 version of the PubMed database is used in the CoreMine system. This has been confirmed by additional searching of the current CoreMine system, where we found that eight additional genes exist in current disease-gene relationships in the CoreMine system. Therefore, the extracted disease-gene relationships are reliable and are far more numerous than those held in the OMIM database.

TABLEII. LIST OF NUMBER OF GENES RELATED TO EPILEPSY MESH TERMS

NO.	MeSH disease terms	Number of relationships
1	Landau-Kleffner Syndrome	1
2	Anti-N-Methyl-D-Aspartate Receptor Encephalitis	8
3	Alexander Disease	15
4	Lafora Disease	17
5	Unverricht-Lundborg Syndrome	20
6	Epilepsy, Frontal Lobe	27
7	MERRF Syndrome	27
8	Myoclonic Epilepsy, Juvenile	35
9	Classical Lissencephalies and Subcortical Band Heterotopias	37
10	Lissencephaly	38
11	MELAS Syndrome	39
12	Spasms, Infantile	47
13	Mitochondrial Encephalomyopathies	55
14	Leigh Disease	68
15	Epilepsy, Absence	70
16	Neuronal Ceroid-Lipofuscinoses	76
17	Rett Syndrome	83
18	Epilepsies, Myoclonic	94
19	Myoclonic Epilepsies, Progressive	99
20	Malformations of Cortical Development	111
21	Epilepsy, Temporal Lobe	128
22	Fragile X Syndrome	148
23	Tuberous Sclerosis	157
Total		1400

C. Topological modules of IE

Topological modules in the PPI network represent the local topological structures with dense inner interactions between each entity they hold [7]. These topological modules would give meaningful insights into disease mechanisms. After we mapped the 871 genes to the PPI network from the String 9.1 database, we obtained 640 proteins and 2,483 PPIs in total. We identified 33 topological modules using the BGL community detection algorithm [13]. To calculate the topological statistics of each module, we used Gephi [14] to visualize and obtain the analysis results of each module. The results are shown in Table III.

The network diameter is 12 and the average path length is 4.0, which showed that each protein in the whole PPI network of IE has a rather short range of interactions. To evaluate the

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

TABLE III. NETWORK TOPOLOGICAL CHARACTERISTICS OF THE 33 MODULES

Module	Nodes	Edges	Edges: Nodes	Density	Average Degree	Maximal Betweenness Centrality
M1	69	188	2.73	0.08	5.45	5347.80
M2	10	45	4.5	1	9	5508
M3	31	59	1.90	0.13	3.81	8035.22
M4	139	579	4.17	0.06	8.33	38123.38
M5	33	63	1.91	0.12	3.82	7172.78
M6	28	47	1.68	0.12	3.36	3597.74
M7	2	1	0.5	1	1	0
M8	11	12	1.09	0.22	2.18	2385.50
M9	44	284	6.45	0.3	12.91	10007.67
M10	49	185	3.78	0.16	7.55	15711.60
M11	5	4	0.8	0.4	1.6	1267.49
M12	28	52	1.86	0.14	3.71	3424.77
M13	7	7	1	0.33	2	5998.86
M14	2	1	0.5	1	1	0
M15	25	44	1.76	0.15	3.52	3976.76
M16	21	61	2.91	0.29	5.81	3241.85
M17	6	6	1	0.4	2	2544.45
M18	2	1	0.5	1	1	0
M19	11	17	1.55	0.31	3.09	2156.05
M20	12	14	1.17	0.21	2.33	2927.71
M21	3	3	1	1	2	0
M22	3	2	0.67	0.67	1.33	1
M23	2	1	0.5	1	1	0
M24	27	64	2.37	0.18	4.74	4716.25
M25	15	16	1.07	0.15	2.13	1946.74
M26	4	3	0.75	0.5	1.5	1854
M27	4	3	0.75	0.5	1.5	3
M28	11	20	1.82	0.36	3.64	6110
M29	8	7	0.88	0.25	1.75	3346.47
M30	15	39	2.6	0.37	5.2	4304.81
M31	5	4	0.8	0.4	1.6	721.26
M32	4	3	0.75	0.5	1.5	1076.03
M33	4	3	0.75	0.5	1.5	1854

significance of modules for IE, we use the maximum betweenness centrality of nodes in each module to rank the centrality of each module in the whole network. The results showed that of the 33 modules, Module 4 (AKT1 has maximum betweenness centrality) has the maximal betweenness centrality and ranks 1st in our list. Module 10 (FOS) ranks 2nd and Module 9 (MT-CYB) ranks 3rd. We also obtain the statistics of degree of the 640 proteins, and the top ten proteins respectively are AKT1, INS, CCND1, FOS, MAPK1, MTOR, MAPK14, CDK1, CREBBP and CTNNB1. All proteins belong to Module 4, apart from FOS, which comes from Module 10.The results reveal that Modules 4, 9 and 10 are more important than other modules, and would be significant disease modules for IE.

The 10 proteins of Module 2 are all cholinergic nicotinic receptors. They bind acetylcholine, which finally leads to the opening of an ion-conducting channel across the plasma membrane, and the ion-conducting channel has a connection with epilepsy. The information of the 10 proteins can refer to the String 9.1 database. The dense inner interactions with each other and functional similarity make this module not only a topological module but also a functional module.

D. Functional enrichment analysis of topological modules

To obtain the function descriptions of each topological module, we used GO to conduct enrichment analysis from biological process, cellular component and molecular function perspectives for each module. We filtered the significant GO terms of each module by the threshold of corrected p-value (CPV) <0.05 [15, 16]. The most representative enriched GO terms of biological process with CPV<0.05 in each module are presented in Table IV. The most representative enriched GO terms of cellular component and molecular function with CPV<0.05 in each module can be seen in supplementary materials.

The results show that there are 10 modules enriched in transmission of nerve impulse and behaviour in biological processes; nine modules enriched in cell projection in cellular components; and 13 modules enriched in protein binding in molecular functions. In the most representative enriched GO terms in biological processes, the 'behaviour' of Module 10 has 57 related GO terms, which are correlated to 10 modules. The 'ion transport' of Module 30 has 47 related GO terms, which are correlated to seven modules. Similarly, in cellular components, the 'axon' of Module 10 has 19 GO terms, which are correlated to seven modules, and the synapse of Module 1 has 16 GO terms, which are correlated to seven modules. In molecular functions, the 'protein binding' of Module 4 has 23 GO terms, which are correlated to 13 modules, and the symporter activity of Module 27 has 18 GO terms, which are correlated to four modules.

TABLE IV. ENRICHED GO TERMS (BIOLOGICAL PROCESSES) IN EACH MODULE SIZE MEANS THE NUMBER OF PROTEINS IN EACH MODULE. NUMBER OF GO TERMS MEANS THERE ARE HOW MANY GO TERMS WITH CPV <0.05 IN TOTAL IN EACH MODULE. IN THE ANALYSIS OF BIOLOGICAL PROCESS, MODULE 33 DOES NOT HAVE SIGNIFICANT ENRICHMENT IN GO TERMS. IN THE ANALYSIS OF CELLULAR COMPONENT, M7, M23, M29, M31, M32, AND M33 DO NOT HAVE SIGNIFICANT ENRICHMENT IN GO TERMS. IN THE ANALYSIS OF MOLECULAR FUNCTION, M22, M33 DO NOT HAVE SIGNIFICANT ENRICHMENTS.WE JUST SHOW THE MOST REPRESENTATIVE ENRICHED GO TERM WITH CPV<0.05 IN EACH MODULE.

Module	Size	Number of GO terms	GO ID	p-value	CPV	Number of Proteins in this GO term	GO Description
M1	69	506	19226	1.23E-38	1.73E-35	34	Transmission of nerve impulse
M2	10	179	35095	4.94E-24	1.93E-21	7	Behavioral response to nicotine
M3	31	153	30031	1.41E-10	1.01E-07	7	Cell projection assembly
M4	139	685	48522	5.24E-21	1.13E-17	64	Positive regulation of cellular process
M5	33	272	48699	1.54E-11	1.46E-08	14	Generation of neurons
M6	28	137	44281	3.39E-11	1.67E-08	17	Small molecule metabolic process
M7	2	1	30316	1.26E-03	2.89E-02	1	Osteoclast differentiation
M8	11	105	9117	2.32E-08	2.13E-06	6	Nucleotide metabolic process
M9	44	127	45333	6.52E-56	2.75E-53	29	Cellular respiration
M10	49	363	7610	1.87E-16	1.81E-13	19	Behavior

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

M11	5	291	32582	2.01E-06	3.93E-04	3	Negative regulation of gene-specific transcription
M12	28	83	6397	1.78E-10	1.18E-07	10	mRNA processing
M13	7	45	6879	8.87E-05	1.46E-02	2	Cellular iron ion homeostasis
M14	2	68	43247	2.93E-08	1.85E-06	2	Telomere maintenance in response to DNA damage
M15	25	30	6493	1.18E-05	5.94E-03	3	Protein amino acid O-linked glycosylation
M16	21	31	6813	1.55E-16	4.04E-14	10	Potassium ion transport
M17	6	63	9267	1.06E-09	1.32E-07	4	Cellular response to starvation
M18	2	17	6406	5.81E-06	1.01E-04	2	mRNA export from nucleus
M19	11	38	52547	4.20E-05	1.13E-02	3	Regulation of peptidase activity
M20	12	97	6073	2.18E-11	2.17E-09	5	Cellular glucan metabolic process
M21	3	32	6743	1.72E-10	2.53E-09	3	Ubiquinone metabolic process
M22	3	29	48699	7.12E-05	1.93E-03	3	Generation of neurons
M23	2	8	6813	7.57E-05	6.82E-04	2	Potassium ion transport
M24	27	176	16481	5.87E-13	3.78E-10	13	Negative regulation of transcription
M25	15	170	71699	3.08E-06	6.66E-04	2	Olfactory placode morphogenesis
M26	4	35	8053	3.52E-07	6.37E-05	2	Mitochondrial fusion
M27	4	34	6836	7.50E-07	6.98E-05	3	Neurotransmitter transport
M28	11	239	7040	6.50E-18	3.09E-15	7	Lysosome organization
M29	8	9	6393	4.89E-04	4.78E-02	1	Termination of mitochondrial transcription
M30	15	44	6811	5.55E-20	4.33E-18	15	Ion transport
M31	5	34	6376	1.02E-05	1.72E-03	2	mRNA splice site selection
M32	4	11	21758	2.80E-04	2.10E-02	1	Putamen development

Taken together, in GO biological processes, the most representative enriched GO term is enriched in cellular respiration in Module 9; in GO cellular components, the most representative enriched GO term is enriched in mitochondrial envelope in Module 9; in GO molecular functions, the most representative enriched GO term is enriched in glutamate receptor activity in Module 1, while NADH dehydrogenase (ubiquinone) activity in Module 9 ranks 2nd. Genes associated with the same disorder share common cellular and functional characteristics, which is annotated in the GO [17, 18].

Unusually high neuronal excitability in the brain is one of the core mechanisms of epileptic seizures. Glutamate is the major excitatory neurotransmitter in the brain, and is also the main excitatory amino acid in the body. NMDA, AMPA and kainate (KA) are receptors of glutamate, and the overstimulation of the receptors may cause neuronal death in epilepsy and neurodegenerative diseases [19]. At present, the research emphasis in neuron injury after seizures has been transferred to mitochondria, thinking that mitochondria damage is the key link in neuronal death [20]. Furthermore, mitochondria influence neuronal excitability, such as the regulation of Ca(2+) homeostasis and ATP production to maintain Na(+)K(+)-ATPase in the central nervous system (CNS) [17], therefore, there is a direct link between mitochondrial dysfunction and increased excitability of neurons. The top ten GO terms in Module 9, such as generation of precursor metabolites and energy, energy derivation by oxidation of organic compounds, oxidative phosphorylation, respiratory electron transport chain, mitochondrial electron transport and NADH to ubiquinone, which are all connected with cellular respiration. This indicated that Module 9 would be a typical disease module for epilepsy.

E. Pathway enrichment in the topological modules

Biological pathways have been widely used in gene function research [21]. Pathway analysis emphasizes the interaction between molecules, and aims to reveal which genes may be associated with the change in which pathway [22]. The most representative enriched KEGG Pathway with CPV<0.05 in each module is shown in Table V.

From the results, we can see that the most representative enriched KEGG Pathways of Modules 2, 10, 22, 27 and 30 all have the same four KEGG Pathways in four modules. Oxidative phosphorylation in Module 9 is the most representative enriched KEGG Pathway of all modules, which is consistent with the results of GO enrichment analysis. Module 9 contains SCO2, the mutation of which refers to the aetiology of convulsions in the neonatal period and infancy [23]. Moreover, the epileptogenic potential and the expressions of ATP5B in hippocampal neurons were measured [24].

F. Disease modules of IE

GO term and pathway enrichment analysis reveals key regulatory processes in disease and prioritize possible disease modules [25]. From the results of GO term and pathway enrichment analysis, we finally found that there are 23 functionally significant topological modules (e.g. M9, M1 and M4), which would be potential disease modules for epilepsy.

TABLE V. ENRICHED KEGG PATHWAY IN EACH MODULE INPUT NUMBER MEANS IN THIS PATHWAY, THE NUMBER OF THE PROTEINS IN THE MODULE. NUMBER OF PATHWAYS MEANS THERE ARE HOW MANY PATHWAYS WITH CPV <0.05 IN TOTAL IN EACH MODULE. M7, M11, M14, M26, M29, M31, M32 AND M33 DO NOT HAVE SIGNIFICANT ENRICHMENT IN THE KEGG PATHWAY.

Module	Size	Number of Pathways	Pathway ID	Pathway	Input number	P-Value	CPV
M1	69	16	hsa04724	Glutamatergic synapse	24	7.07E-19	1.78E-15
M2	10	3	hsa04080	Neuroactive ligand-receptor interaction	10	6.21E-14	3.63E-12

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

M3	31	4	hsa04810	Regulation of actin cytoskeleton	12	6.33E-10	8.60E-07
M4	139	17	hsa04150	mTOR signaling pathway	29	7.84E-20	3.59E-16
M5	33	2	hsa00565	Ether lipid metabolism	4	1.03E-06	0.00023
M6	28	17	hsa01200	Carbon metabolism	10	4.65E-10	4.42E-07
M8	11	1	hsa00230	Purine metabolism	6	8.49E-08	4.36E-05
M9	44	8	hsa00190	Oxidative phosphorylation	29	2.41E-31	1.67E-28
M10	49	7	hsa04080	Neuroactive ligand-receptor interaction	14	1.00E-09	4.00E-07
M12	28	1	hsa03040	Spliceosome	6	3.09E-07	0.00021
M13	7	6	hsa04623	Cytosolic DNA-sensing pathway	2	0.00071	0.0078
M15	25	6	hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	5	2.55E-06	0.00060
M16	21	1	hsa04725	Cholinergic synapse	2	0.0020	0.020
M17	6	1	hsa04140	Regulation of autophagy	4	2.45E-09	8.78E-07
M18	2	5	hsa03008	Ribosome biogenesis in eukaryotes	2	0.00066	0.0032
M19	11	1	hsa04142	Lysosome	4	2.08E-07	0.00013
M20	12	3	hsa00500	Starch and sucrose metabolism	4	2.85E-07	8.27E-06
M21	3	2	hsa00900	Terpenoid backbone biosynthesis	2	6.24E-06	4.03E-05
M22	3	2	hsa04360	Axon guidance	3	4.14E-06	0.00012
M23	2	3	hsa04911	Insulin secretion	2	9.61E-05	0.0013
M24	27	1	hsa04330	Notch signaling pathway	3	0.00044	0.037
M25	15	1	hsa04141	Protein processing in endoplasmic reticulum	3	0.00062	0.031
M27	4	1	hsa04727	GABAergic synapse	2	1.72E-05	0.00025
M28	11	2	hsa04142	Lysosome	5	1.44E-09	1.64E-07
M30	15	6	hsa05033	Nicotine addiction	7	1.49E-12	5.58E-11

In Module 1, glutamatergic synapse is the most typical pathway corresponding to the GO analysis, which includes transmission of nerve impulse, synapse and glutamate receptor activity as the three domains. It has been reported that cascading excitation within networks of synaptically connected excitatory glutamatergic neurons is a classical mechanism in epileptic synchronization [2]. In this module, CACNA1H as a T-type calcium channel gene is a susceptibility gene or disease-causing gene in childhood absence epilepsy; mutations of CACNB4 can cause juvenile myoclonic epilepsy in humans [26]; and a single mutation was identified in the GRIN2A gene for Landau-Kleffner syndrome [27]. In Module 4, the mTOR signalling pathway is the most important pathway in accordance with GO analysis, which contains positive regulation of the cellular process, cytosol and protein binding. The mTOR signalling pathway refers to main multiple cellular functions, containing protein synthesis, cell growth and proliferation and synaptic plasticity, which may affect neuronal excitability and be responsible for epileptogenesis [28]. Module 4 contains TSC1, TSC2 and PTEN, which are the most prominent causal mutations to have been found in patients with epilepsy [29, 30]. Both TSC1 and TSC2 genes act as negative regulators of mTOR signalling, and mutations cause hyperactivation of the pathway. Module 24 can influence the Notch signalling pathway, which is associated with neuronal discharges. Recent studies have suggested that Notch signalling is up-regulated in answer to seizure activity, and its activation further advances neuronal excitation of CA1 pyramidal neurons in acute seizures [31]. HDAC2 of Module 24 plays an important role in the pathogenesis of human temporal lobe epilepsy [32]. Module 27 is highly enriched in GABAergic synapse, and the three domains of GO analysis are enriched in neurotransmitter transport, integral to plasma membrane, symporter activity. An impairment of GABAergic synaptic inhibition represents a pivotal pathway of epileptogenesis [33]. SLC6A11 is GABA transporter 3. The importance of a module is evaluated by its influence on important pathways implicated in disease, and the contained genes also demonstrate the functional relevance of module to disease [34]. Following this analysis approach, we considered Modules 1, 2, 4, 9, 10, 16, 24, 27 and 30 as disease modules. Of these, Module 9 is the most influential and typical. Several disease genes in each disease module are presented in Table VI. We drew the PPI network of nine disease modules using Gephi to show the inner connection (see Fig. 1).

TABLE VI. SEVERAL DISEASE GENES IN EACH DISEASE MODULE

Module	Several disease genes in this module	Supporting literature PubMed ID
M1	CACNA1H, CACNB4, GRIN2A	24277868 , 24875574, 24828792
M2	CHRNB2, CHRNA7	23032131, 24090792
M4	TSC1,TSC1, PTEN	24672426, 24917535
M9	SCO2, ATP5B	19070318, 22616176
M10	RORB	24355400, 23279911
M16	KCNA1, KCNA4, KCNAB1	18440780
M24	HDAC2	21987499
M27	SLC6A11	24561070
M30	SCN1A, SCN8A, SCN2A, SCN3A, SCN1B, SCN2B	23032131, 24888894, 24337656

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

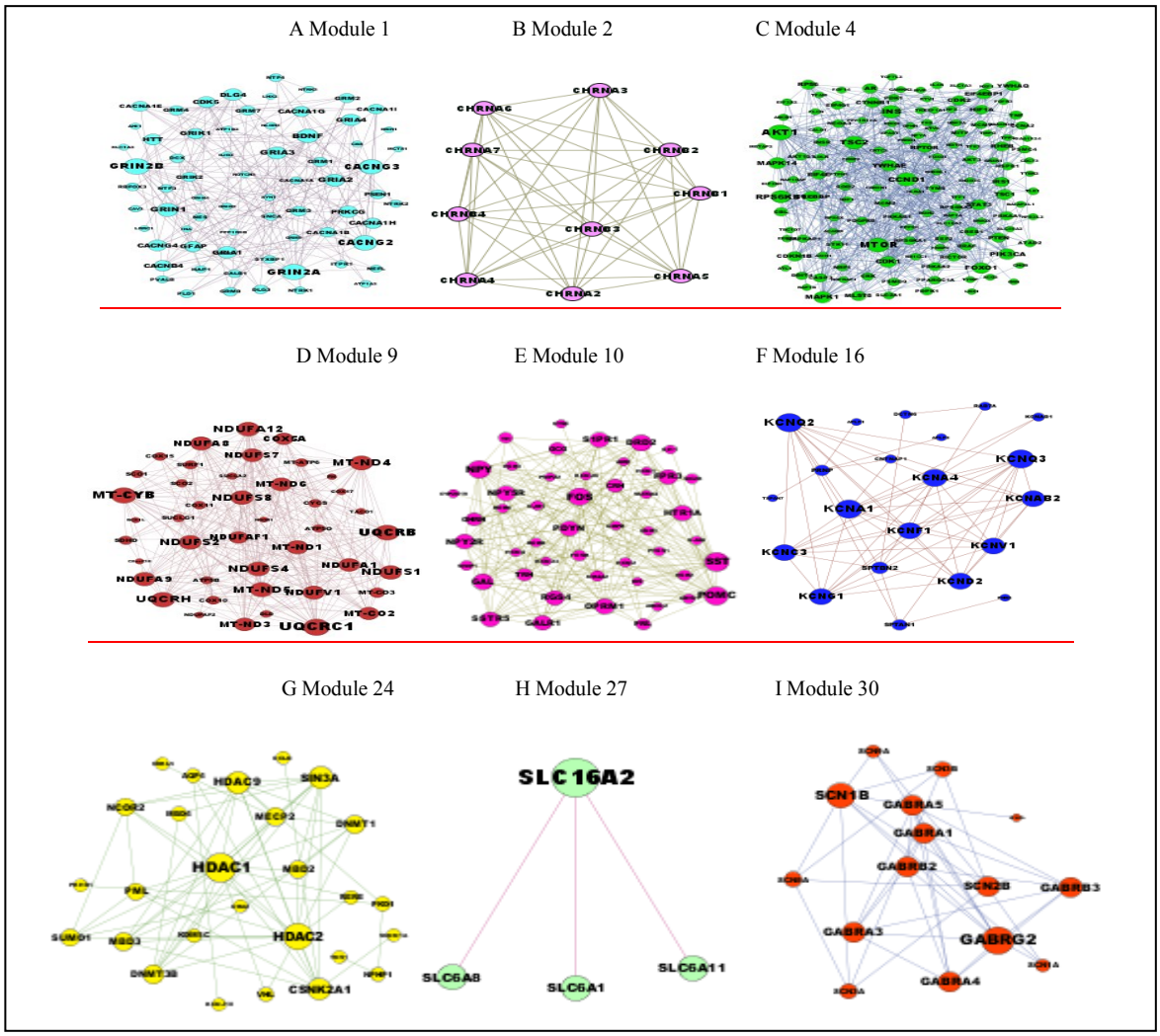

Fig. 1. PPIs of disease modules.

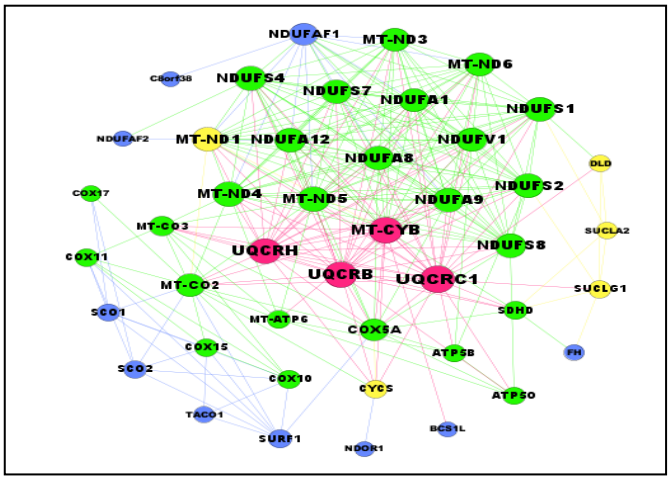

Fig.2. PPIs and drug targets in Module 9. Green nodes represent the proteins that exist in the Oxidative phosphorylation pathway and also the drug targets for epilepsy in the DrugBank database. Red nodes represent the proteins that exist in the Oxidative phosphorylation pathway and may be the potential drug targets for epilepsy. Yellow nodes represent the proteins that do not exist in the Oxidative phosphorylation pathway but are the drug targets for epilepsy in the DrugBank database. Blue nodes represent the proteins that neither exist in the Oxidative phosphorylation pathway nor are the drug targets for epilepsy in the DrugBank database.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

G. Identification of potential drug targets

Using the detected disease modules for epilepsy, we confirmed and obtained new biological meaningful drug targets for epilepsy. For example, it shows that Module 9 contains 44 proteins (Fig. 2), 29 of which are in the Oxidative phosphorylation pathway. There are 23 (23/29=79.3%) proteins, such as ATP5B, COX5A, MT-CO2 and MT-CYB, have been targeted by drugs in the DrugBank database. Those 23 proteins are mostly related to cellular respiration processes (19/23=83%), mitochondrial envelope cellular components (22/23=96%) and NADH dehydrogenase (ubiquinone) activity (14/23=61%), which show strong function correlation to epilepsy. For example, NADH targets 14 of these 23 proteins, such as MT-ND3, NDUFA and NDUFS1, etc. It is well known that the pathogenesis of epilepsy is closely related to the decline of the GABAA receptor α1 methyl phosphorylation level. The cause of the decline of the phosphorylation level is probably due to the intracellular NADH/NAD + percentage changes and the weakened ability of glycolysis [35]. NADH can promote the function of glycolysis within the neurons, and improve the GABAA receptor α1 methyl phosphorylation level, repair the function of the GABAA receptor and related functions, and eventually reduce seizures. Other proteins like COX5A, MT-CO2 and MT-CO3 (common drug: cholic acid) [35], ATP5B (drug: quercetin) [36], [37] and SDHD (drug: succinic acid) [38], [39] have well-documented targeted drugs.

However, although participating in the respiratory chain in mitochondria, the remaining four proteins, namely MT-CYB, UQCRB, UQCRC1 and UQCRH, have no approved available epilepsy drugs. This means that currently there are no available drugs for epilepsy targeting these four proteins. However, Letm1 knock-down by a lent virus bearing LV-Letm1-sh resulted in mitochondrial swelling and decreased the expression of Letm1 target protein mitochondrial encoded MT-CYB. A behavior study revealed that inhibition of Letm1 caused early onset of the first seizure, and increased seizure frequency and duration; moreover, inhibition of Letm1 and mitochondrial dysfunctions contributes to the development of epileptic seizures [40]. Therefore, these four proteins associated with epilepsy would be good candidate potential drug targets for the treatment of epilepsy.

Except for Oxidative phosphorylation pathway, there are totally eight eligible pathways like Parkinson's disease, Alzheimer's disease, Huntington's disease, Citrate cycle (TCA cycle) and Carbon metabolism in Module 9, in which Parkinson's disease, Alzheimer's disease and Huntington's disease have some connection with epilepsy [41]. Five of the remaining 15 proteins of 44 proteins not in the Oxidative phosphorylation pathway, such as BCS1L, CYCS, SCO1 and SUCLA2 connect to these eight pathways. For example, CYCS is involved in Parkinson's disease, Alzheimer's disease and Huntington's disease. However, the remaining ten proteins namely, BCS1L, MT-ND1, NDUFAF1, NDUFAF2, NDOR1, SCO1, SCO2, SURF1, TACO1 and C8orf38 are not involved in any pathway. Furthermore, Ten (e.g. BCS1L, FH, SURF1, TACO1 and C8orf38.) out of the 15 proteins are not recorded as drug targets in the DrugBank, which could be further investigated in experimental research.

The other five proteins like DLD, MT-ND1 and CYCS have existed drugs. For example, CYCS is the target of heme with well-defined pharmacological mechanisms [42].

III. Discussion

Integration of multiple sources for the drug targets identification is a promising approach in the big data era of biomedical research. We integrated three kinds of databases and utilized the well-defined network analysis algorithms and software to realize the detection of disease modules and potential drug targets for IE. To obtain a reliable disease-gene relationship database, we conducted large amount of manual efforts to evaluate the relationships based on the data captured from Coremine system. To detect the topological modules relevant to IE, we choose BGL algorithm to find the subnetworks with relatively dense links. However, it extracted several modules with small sizes (i.e.2-3 nodes), so we did additional post-processing to evaluate these modules. However, after we manually evaluated these small size modules, we found that all the genes in the 6 modules (see Table 3, e.g. Module 7, Module 14 and Module 21) belong to some kind of protein complex, which means that they form substantial functional modules. For example, the proteins: TERF2IP and ACD in Module 14, are component of the shelterin complex (telosome). Therefore, we see that BGL method obtained rather reliable results for small size topological modules.

To further investigate the potent of the novel drug targets we detected, we used String 9.1 and Stitch 4.0 database to evaluate the possible drugs for MT-CYB. We finally found that MT-CYB has binding action with UQCRB, UQCRC1 and UQCRH in the String 9.1 database with high scores: 0.999, 0.995 and 0.980, respectively. The four proteins all participate in Complex III and Ubiquinol-cytochrome c reductase, which relate to the respiratory electron transport pathway. In the Stitch 4.0 database [47], we discovered that succinate, a compound, which has a central inhibitory effect, exerts inhibition action on MT-CYB. Succinic acid is one of the ingredients of amber, which is used in traditional Chinese medicine, while amber has a mind-tranquillizing and anti-epileptic effect. Experimental studies have shown that succinic acid has an anti-epileptic effect, and the chemical structure of succinic acid is similar to that of glutamic acid and GABA, which demonstrates that succinic acid may affect the neurotransmitter receptor system[43]. In conclusion, MT-CYB may be the target of succinic acid for the treatment of epilepsy and may provide insight into the mechanism of action for succinic acid. This result would contribute to our future experimental studies in laboratory settings.

However, in this preliminary study, there are complicated and labor-intensive tasks involved in several

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

main steps like disease-gene relationship acquisition and drug-target evaluation. We would adopt text mining methods[44] and drug-target detection algorithms[45] to automatically help discover the disease modules and drug targets for IE. Furthermore, we will develop web server applications for data sharing when the related data sets are finalized.

IV. METHOD

A. Data integration

We used Medical Subject Headings 2013 (MeSH) [9] to obtain terms for IE, then extracted disease-gene relationships from the CoreMine (http://www.coremine.com/medical/#search) PubMed [46] search engine system. The protein-protein interaction (PPI) data was downloaded from the String 9.1 database [10]. We got the result of Gene Ontology(GO) enrichment analysis from the BINGO2.44 plugin in CytoScape [47], and enriched KEGG Pathway from KOBAS2.0 [48]. Network module visualization was designed by Gephi [14]. We searched relevant drugs in the DrugBank database [11], specific action of PPI in the String 9.1 database and protein-chemical interactions in the Stitch 4.0 database [49].

B. Extraction of epilepsy disease PPI network

We mapped the 871 genes to the PPI. Only genes both in the PPI network and belonging to the 871 genes are used in the following study. We defined the weight of each interaction between two proteins as their correlation value. We just chose the interactions whose correlation value>700 as edges for the study. We extracted the maximum connected component of the network, which consists 640 nodes and 2483 edges [34].

C. Topological module identification

Topological module refers to the subnetwork in one whole network, which has relatively dense links in the subnetwork when compared with the links outside of the subnetwork. We took the BGL community detection method to get 33 disease-related topological PPI modules [13]. Modularity is an attribute of a network community [50].

D. Gene ontology enrichment analysis

There are a lot of online analysis platforms as well as relevant analysis software for conducting GO enrichment analysis [51]. We draw support from the Bingo2.44 plugin in the CytoScape0.8.2 software to get the results of GO enrichment analysis [52]. The Bingo2.44 plugin can map the genes to GO terms using a hypergeometric distribution relationship, then we can obtain the GO terms with corresponding genes. Through Bonferroni correction to control the false positive rate of analysis, this process will return a p-value and a CPV.

E. KEGG Pathway analysis

In this paper, we utilized the KEGG Pathway database to achieve one or more pathways containing proteins in one module by means of KOBAS2.0 software. By calculating the hypergeometric distribution relationship, we can get the statistically significant pathways. Through Bonferroni correction to control the false positive rate of analysis, this process will also return a p-value and a CPV.

ACKNOWLEDGEMENTS

This work is partially supported by the National S&T Major Special Project (2012ZX09503-001-003), National Science Foundation of China (61105055, 81230086), National Key Technology R&D Program (2013BAI02B01, 2013BAI13B04) and Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (CURE) (14032).

REFERENCES

[1]. Hu, K., et al., Protective effects of quercetin against status epilepticus induced hippocampal neuronal injury in rats: involvement of X-linked inhibitor of apoptosis protein. Acta Neurol Belg, 2011. 111(3): p. 205-12.

[2]. Rogawski, M.A., AMPA receptors as a molecular target in epilepsy therapy. Acta Neurologica Scandinavica, 2013. 127: p. 9-18.

[3]. Kapur, J., Role of Neuronal Loss in the Pathogenesis of Recurrent Spontaneous Seizures. Epilepsy Curr, 2003. 3(5): p. 166-167.

[4]. Wassenaar, M., et al., Prognostic factors for medically intractable epilepsy: a systematic review. Epilepsy Res, 2013. 106(3): p. 301-10.

[5]. Cilio, M.R., E.A. Thiele, and O. Devinsky, The case for assessing cannabidiol in epilepsy. Epilepsia, 2014. 55(6): p. 787-90.

[6]. Rogawski, M.A. and W. Loscher, The neurobiology of antiepileptic drugs. Nat Rev Neurosci, 2004. 5(7): p. 553-64.

[7]. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, Network medicine: a network-based approach to human disease. Nat Rev Genet, 2011. 12(1): p. 56-68.

[8]. Kim, Y.A., S. Wuchty, and T.M. Przytycka, Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput Biol, 2011. 7(3): p. e1001095.

[9]. Lowe, H.J. and G.O. Barnett, Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. Jama, 1994. 271(14): p. 1103-1108.

[10]. Franceschini, A., et al., STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res, 2013. 41(Database issue): p. D808-15.

[11]. Law, V., et al., DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res, 2014. 42(Database issue): p. D1091-7.

[12]. Hamosh, A., et al., Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res, 2005. 33(Database issue): p. D514-7.

[13]. Blondel, V.D., et al., Fast unfolding of communities in large networks. Journal of Statistical Mechanics-Theory and Experiment, 2008(10): P10008

[14]. Jacomy, M., et al., ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLoS One, 2014. 9(6): p. e98679.

[15]. Ashburner, M., et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000. 25(1): p. 25-9.

[16]. Liu, Q., et al., Comparative evaluation of gene-set analysis methods. BMC Bioinformatics, 2007. 8: p. 431.

[17]. The Gene Ontology (GO) project in 2006. Nucleic Acids Res, 2006. 34(Database issue): p. D322-6.

[18]. Goh, K.I., et al., The human disease network. Proc Natl Acad Sci U S A, 2007. 104(21): p. 8685-90.

[19]. Quincozes-Santos, A., et al., Oxidative stress mediated by NMDA, AMPA/KA channels in acute hippocampal slices: Neuroprotective effect of resveratrol. Toxicology in Vitro, 2014. 28(4): p. 544-551.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

[20]. Kann, O. and R. Kovacs, Mitochondria and neuronal activity. American Journal of Physiology-Cell Physiology, 2007. 292(2): p. C641-C657.

[21]. Zhao, X., et al., [A novel biological pathway expansion method based on the knowledge of protein-protein interactions]. Yi Chuan, 2014. 36(4): p. 387-94.

[22]. Clauset, A., M.E.J. Newman, and C. Moore, Finding community structure in very large networks. Physical Review E, 2004. 70(6) (6 Pt 2).

[23]. Nagy, A., et al., [Convulsions in neonatal period and infancy with rare etiology (neurogenetic disease)]. Ideggyogy Sz, 2008. 61(11-12): p. 417-22.

[24]. He, G., et al., [Effects of electrode on epileptogenic focus potential and expressions of the beta subunit of ATP synthase in rats with penicillin-induced epilepsy]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi, 2012. 29(2): p. 287-90.

[25]. Liu, Y., et al., DiME: a scalable disease module identification algorithm with application to glioma progression. PLoS One, 2014. 9(2): p. e86693.

[26]. Etemad, S., et al., The juvenile myoclonic epilepsy mutant of the calcium channel beta subunit displays normal nuclear targeting in nerve and muscle cells. Channels (Austin), 2014. 8(4). 8(4):334-43.

[27]. Conroy, J., et al., Towards the identification of a genetic basis for Landau-Kleffner syndrome. Epilepsia, 2014. 55(6): p. 858-65.

[28]. Meng, X.F., et al., Role of the mTOR signaling pathway in epilepsy. J Neurol Sci, 2013. 332(1-2): p. 4-15.

[29]. Lasarge, C.L. and S.C. Danzer, Mechanisms regulating neuronal excitability and seizure development following mTOR pathway hyperactivation. Front Mol Neurosci, 2014. 7: p. 18.

[30]. Kothare, S.V., et al., Severity of manifestations in tuberous sclerosis complex in relation to genotype. Epilepsia, 2014. 55(7):1025-9.

[31]. Sha, L., et al., Notch signaling activation promotes seizure activity in temporal lobe epilepsy. Mol Neurobiol, 2014. 49(2): p. 633-44.

[32]. Huang, Y., et al., Increased expression of histone deacetylases 2 in temporal lobe epilepsy: a study of epileptic patients and rat models. Synapse, 2012. 66(2): p. 151-9.

[33]. Dejanovic, B., et al., Exonic microdeletions of the gephyrin gene impair GABAergic synaptic inhibition in patients with idiopathic generalized epilepsy. Neurobiol Dis, 2014. 67: p. 88-96.

[34]. He, D., Z.P. Liu, and L. Chen, Identification of dysfunctional modules and disease genes in congenital heart disease by a network-based approach. BMC Genomics, 2011. 12: p. 592.

[35]. Waterreus, R.J., et al., Cerebrotendinous xanthomatosis (CTX): a clinical survey of the patient population in The Netherlands. Clin Neurol Neurosurg, 1987. 89(3): p. 169-75.

[36]. Nieoczym, D., et al., Effect of quercetin and rutin in some acute seizure models in mice. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2014. 3; 54:50-8.

[37]. Nassiri-Asl, M., et al., Effects of quercetin on oxidative stress and memory retrieval in kindled rats. Epilepsy Behav, 2013. 28(2): p. 151-5.

[38]. Pearl, P.L., et al., Epilepsy in succinic semialdehyde dehydrogenase deficiency, a disorder of GABA metabolism. Brain Dev, 2011. 33(9): p. 796-805.

[39]. Vanadia, E., et al., Therapeutic efficacy of magnesium valproate in succinic semialdehyde dehydrogenase deficiency. JIMD Rep, 2013. 8: p. 133-7.

[40]. Zhang, X., et al., Association of Mitochondrial Letm1 with Epileptic Seizures. Cereb Cortex, 2013: bht118.

[41]. Quarrell, O.W., et al., Managing juvenile Huntington's disease. Neurodegener Dis Manag, 2013. 3(3): p. 267-76.

[42]. Quincozes-Santos, A., et al., Oxidative stress mediated by NMDA, AMPA/KA channels in acute hippocampal slices: neuroprotective effect of resveratrol. Toxicology in Vitro, 2014. 28(4): p. 544-51.

[43]. Wong, C.G., T. Bottiglieri, and O.C. Snead, 3rd, GABA, gamma-hydroxybutyric acid, and neurological disease. Ann Neurol, 2003. 54 Suppl 6: p. S3-12.

[44]. Fluck, J. and M. Hofmann-Apitius, Text mining for systems biology. Drug Discov Today, 2014. 19(2): p. 140-4.

[45]. Yamanishi, Y., et al., DINIES: drug-target interaction network inference engine based on supervised analysis. Nucleic Acids Res, 2014. 42(Web Server issue): p. W39-45.

[46]. Lu, Z., PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford), 2011. 2011: p. baq036.

[47]. Maere, S., K. Heymans, and M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics, 2005. 21(16): p. 3448-9.

[48]. Xie, C., et al., KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res, 2011. 39(Web Server issue): p. W316-22.

[49]. Kuhn, M., et al., STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res, 2014. 42(Database issue): p. D401-7.

[50]. Krawczyk, M.J., Differential equations as a tool for community identification. Physical Review E, 2008. 77(6): 065701.

[51]. Khatri, P. and S. Draghici, Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics, 2005. 21(18): p. 3587-3595.

[52]. Wang, J.G., et al., NOA: a novel Network Ontology Analysis method. Nucleic Acids Res, 2011. 39(13):e87.