Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2009, Article ID 726728, 6 pages
doi:10.1155/2009/726728

Research Article

Integrative Decomposition Procedure and Kappa Statistics for the Distinguished Single Molecular Network Construction and Analysis

Lin Wang,[1] Ying Sun,[1] Minghu Jiang,[2] and Xiguang Zheng[1]

[1] *Biomedical Center, School of Electronics Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2] *Lab of Computational Linguistics, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Lin Wang, wanglin98@tsinghua.org.cn

Received 27 October 2008; Accepted 19 February 2009

Recommended by Zhenqiu Liu

Our method concentrates on and constructs the distinguished single gene network. An integrated method was proposed based on linear programming and a decomposition procedure with integrated analysis of the significant function cluster using Kappa statistics and fuzzy heuristic clustering. We tested this method to identify ATF2 regulatory network module using data of 45 samples from the same GEO dataset. The results demonstrate the effectiveness of such integrated way in terms of developing novel prognostic markers and therapeutic targets.

1. Introduction

In the postgenomic era, with microarray technologies producing great deal of gene expression data, mining these data to get insight into biological processes at system-wide level has become a challenge for bioinformatics. On one hand, due to the complex and distribute nature of biological research, there is a great deal of methods for inferring gene regulatory networks. But all these methods focused on constructing the complicated entire network calculated from the given microarray data. The tremendous amounts of genes in those networks distribute analysts' attention, so it is hard to get any clear perception of valuable knowledge from such complicated networks, let alone further study of each single gene. On the other hand, the wide spread of knowledge over independent databases aggravates the hardness of integrating comprehensive annotation information for genes and lowers the study effectiveness. Thus, a novel method integrating both single molecular network construction and highly centralized gene-functional-annotation analysis is in demand for gene network and functional analysis.

This paper proposed an integrated method based on linear programming and a decomposition procedure with integrated analysis of the significant function cluster using Kappa statistics and fuzzy heuristic clustering. Our method concentrates on and constructs the distinguished single gene network integrated with function prediction analysis by DAVID. For the distinguished single molecular network, we did (1) control and experiment comparison, (2) identification of activation and inhibition networks, (3) construction of upstream and downstream feedback networks, and (4) functional module construction. We tested this method to identify ATF2 regulation network module using data of 45 samples from one and the same GEO dataset. The results demonstrate the effectiveness of such integrated way in terms of developing novel prognostic markers and therapeutic targets.

2. Methods

2.1. Distinguished Single Molecular Network Construction. The entire network was constructed using GRNInfer [1] and GVedit tools. GRNInfer is a novel mathematic method called gene network reconstruction (GNR) tool based on linear programming and a decomposition procedure that is used for inferring gene networks. The method theoretically ensures the derivation of the most consistent network

structure with respect to all of the datasets, thereby not only significantly alleviating the problem of data scarcity but also remarkably improving the reconstruction reliability. The general solution for a single dataset is the following (1), which represents all of the possible networks:

$$J = (X' - A)U\Lambda^{-1}V^T + YV^T = \hat{J} + YV^T, \quad (1)$$

where $J = (J_{ij})_{n\times n} = \partial f(x)/\partial x$ is an $n \times n$ Jacobian matrix or connectivity matrix, $X = (x(t_1),\ldots,x(t_m))$, $A = (a(t_1),\ldots,a(t_m))$, and $X' = (x'(t_1),\ldots,x'(t_m))$ are all $n \times m$ matrices with $x_i'(t_j) = [x_i(t_{j+1}) - x_i(t_j)]/[t_{j+1} - t_j]$ for $i = 1,\ldots,n$; $j = 1,\ldots,m$. $X(t) = (x_1(t),\ldots,x_n(t))^T \in R^n$, $a = (a_1,\ldots,a_n)^T \in R^n$, $x_i(t)$ is the expression level (mRNA concentrations) of gene i at time instance t. $y = (y_{ij})$ is an $n \times n$ matrix, where y_{ij} is zero if $e_j \neq 0$ and is otherwise an arbitrary scalar coefficient. $\wedge^{-1} = \text{diag}\,(1/e_i)$ and $1/e$ is set to be zero if $e_i = 0$. U is a unitary $m \times n$ matrix of left eigenvectors, $\wedge = \text{diag}\,(e_1,\ldots,e_n)$ is a diagonal $n \times n$ matrix containing the n eigenvalues, and V^T is the transpose of a unitary $n \times n$ matrix of right eigenvectors.

But the entire network is too complex to get any clear perception of such complicated relationships among those genes, let alone further study of each single gene. We constructed the distinguished single molecular network by selecting the centered gene and its directly related genes based on the entire network for further study. We take into account the effectiveness of biology study in order to concentrate on single molecular network rather than the intricate entire network. It is helpful to get intensive and deep insight of the whole network. For the distinguished single molecular network, we did (1) control and experiment comparison, (2) identification of activation and inhibition networks, (3) construction of upstream and downstream feedback networks, and (4) functional module construction.

2.2. Functional Annotation Clustering. For the function of genes that is neither determined by their sequence nor by the protein families they belong to [2], the function of those genes included in the same single molecular network should not be interpreted separately, but should be analyzed together according to the whole single molecular network. This method takes into account the network nature of biological annotation contents in order to concentrate on the larger biological picture rather than an individual gene. We used DAVID to do functional annotation clustering. It changes functional annotation analysis from term- or gene-centric to biological module-centric [2] in accordance with our network analysis aim.

The DAVID gene functional clustering tool provides typical batch annotation and gene-GO term enrichment analysis for highly throughput genes by classifying them into gene groups based on their annotation term co-occurrence [3]. DAVID uses a novel algorithm to measure relationships among the annotation terms based on the degrees of their coassociation genes to group similar annotation contents from the same or different resources into annotation groups. The grouping algorithm is based on the hypothesis that similar annotations should have similar gene members. The functional annotation clustering integrates the same techniques of Kappa statistics to measure the degree of the common genes between two annotations, and fuzzy heuristic clustering to classify the groups of similar annotations according kappa values [4, 5]. The tool also allows observation of the internal relationships of the clustered terms by comparing it to the typical linear, redundant term report, over which similar annotation terms may be distributed among many other terms.

3. Results and Discussion

We tested this method using microarrays containing 22215 genes in 40 MPM tumors and 5 normal pleural tissues from one and the same GEO datasets. We identified potential tumor molecular markers and chose the top 51 significant positive genes with normalization of log2, the minimum fold change = 3.5, delta = 1.59, and a false-discovery rate of 0% using SAM [6]. We selected activating transcription factor (ATF)-2 because it is one of the most distinguished genes in MPM. It is a member of the ATF/cyclic AMP-responsive element binding protein family of transcription factors.

3.1. Normal Tissues and Tumor Comparisons of Distinguished Single Molecular Network. We, respectively, constructed the interaction network of the above 51 genes in healthy tissues and that in tumor using GRNInfer [1] and GVedit tools and selected the ATF2-centered downstream subnetworks. With comparison of these ATF2-centered subnetworks, we can get a more clear perception of the notable differences between normal tissues and tumor, as shown in Figure 1. It appeared that ATF2 inhibits C11orf9, C18orf10, C20orf31, CALD1, CAMK2G, DDX3X, FALZ, GLS, GOLGA2, ID2, NME2, NMU, NONO, PAWR, PLOD2, PSMF1, RBMS1, RIC8A, RNF10, TEAD4, TIA1, TNPO1, unknown2, unknown3, WBSCR20C, and ZF in normal tissues, as shown in Figure 1(a). It appeared that ATF2 inhibits C11orf9, C15orf5, C18orf10, C20orf31, CAMK2G, CDR2, DDX3X, FALZ, FLJ10707, GLS, GOLGA2, ID2, KRT18, LRRC1, NME2, NMU, NONO, NSUN5, OBSL1_2, PLOD2, PLXNA1, PTOV1, RBMS1, RIC8A, RNASEH1, RNF10, TEAD4, TIA1, UCK2, USP11, and ZF, while it activates CALD1 and TFAP2C in tumor, as shown in Figure 1(b).

With comparison between the two results, notable differences can be shown clearly in order to get further perception of pathological changes in MPM. For example, ATF2 target genes appeared in ATF2 activation to CALD1, TFAP2C in MPM, as only shown in Figure 2(b). Caldesmon (CALD1) is a potential actomyosin regulatory protein found in smooth muscle and nonmuscle cells [7]. Transcription factor AP2-gamma (TFAP2C) is alternatively titled AP2. Families of related transcription factors are often expressed in the same cell lineages but at different times or sites in the developing embryo. The AP2 family appears to regulate the expression of genes required for development of tissues of ectodermal origin such as neural crest and skin [8]. AP2 may also be

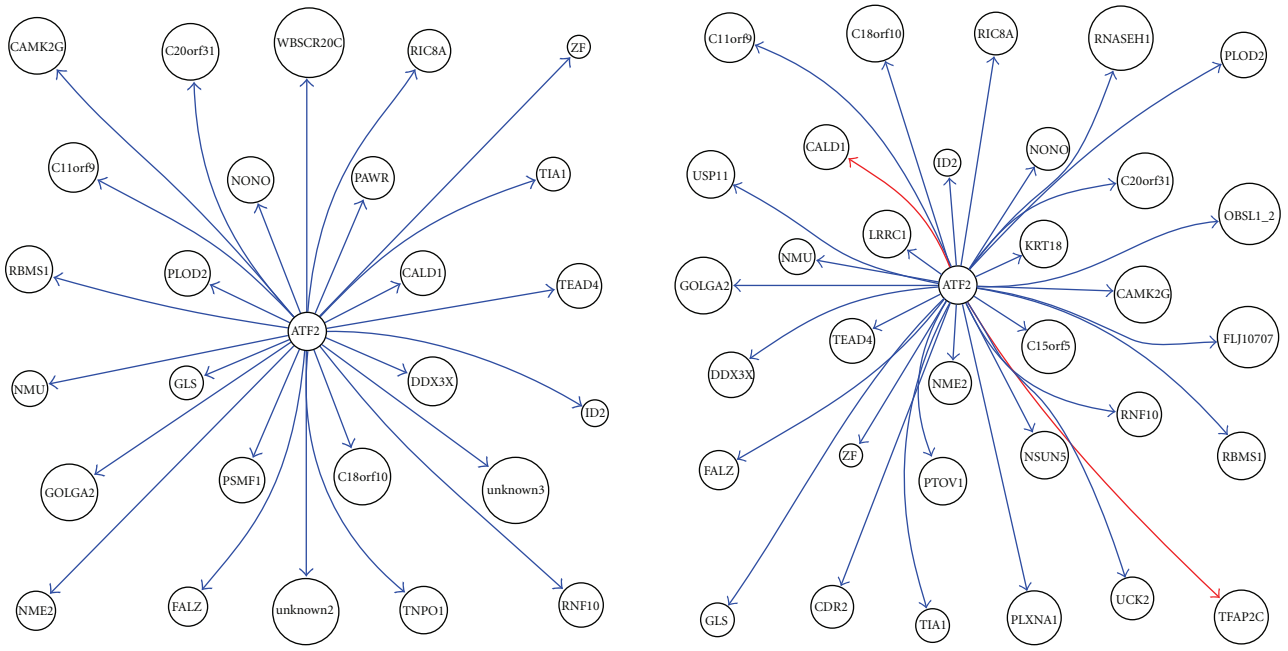

Figure 1: ATF2 downstream network in (a) normal tissue and (b) MPM tissue.

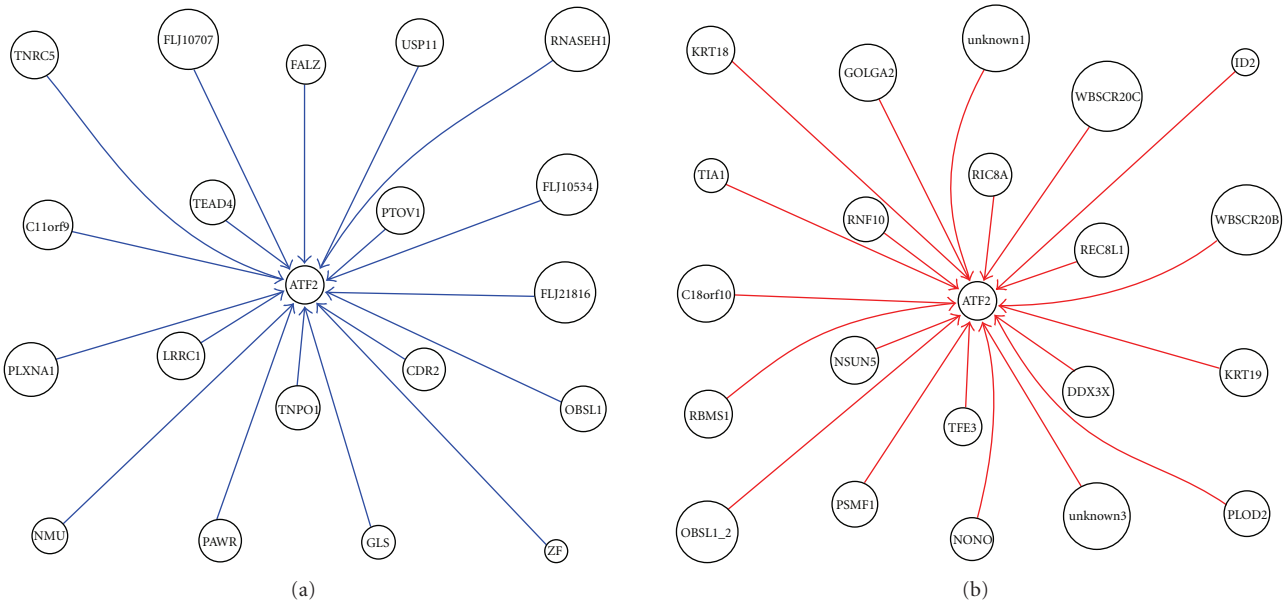

Figure 2: (a) ATF2 upstream inhibition network of MPM; (b) ATF2 upstream activation network of MPM.

involved in the overexpression of c-erbB-2 in human breast cancer cells [9].

3.2. Identification of Activation and Inhibition Networks for the Distinguished Single Molecule. We also identified the activation and inhibition networks, respectively, in order to simplify and intensify the analysis process. For example, in ATF2 upstream network of MPM, as shown in Figure 2, it appeared that C11orf9, CDR2, FALZ, FLJ10534, FLJ10707, FLJ21816, GLS, LRRC1, NMU, OBSL1, PAWR, PLXNA1, PTOV1, RNASEH1, TEAD4, TNPO1, TNRC5, USP11, and ZF inhibit ATF2, as shown in Figure 2(a), whereas C18orf10, DDX3X, GOLGA2, ID2, KRT18, KRT19, NONO, NSUN5, OBSL1_2, PLOD2, PSMF1, RBMS1, REC8L1, RIC8A, RNF10, TFE3, TIA1, unknown1, unknown3, WBSCR20B, and WBSCR20C activate ATF2, as shown in Figure 2(b).

ATF2 upstream genes TFE3, REC8L1 showed activation to ATF2. TFE3 is a member of the helix-loop-helix family

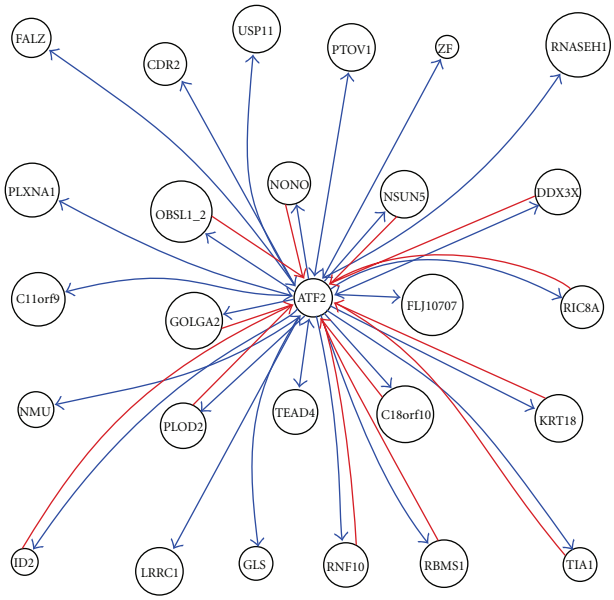

Figure 3: ATF2 feedback subnetwork of MPM.

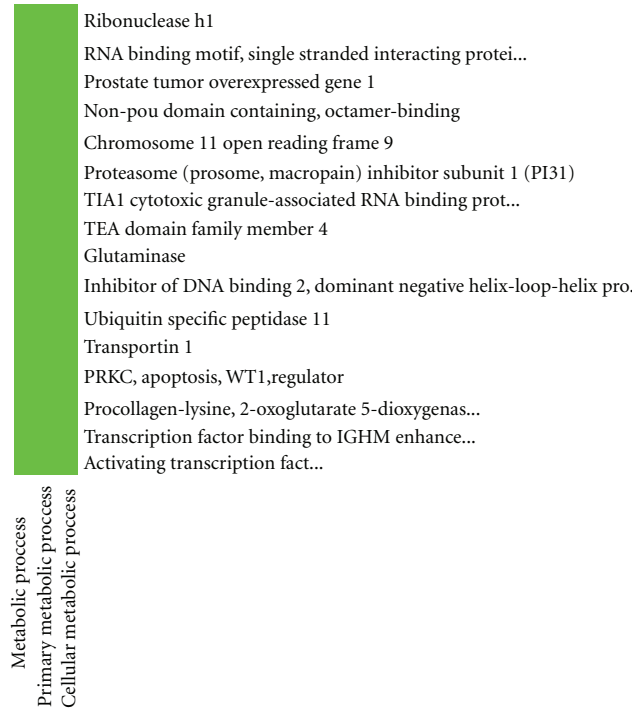

Figure 4: One ATF2 upstream gene metabolic network including RBMS1, RNASEH1, PTOV1, NONO, C11orf9, PSMF1, TIA1, TEAD4, GLS, ID2, USP11, TNPO1, PAWR, PLOD2, and TFE3.

of transcription factors and binds to the mu-E3 motif of the immunoglobulin heavy-chain enhancer and is expressed in many cell types [10]. Nakagawa et al. [11] identified TFE3 as a transactivator of metabolic genes that are regulated through an E box in their promoters which led to metabolic consequences such as activation of glycogen and protein synthesis, but not lipogenesis, in liver [11]. REC8L1 is the human homolog of yeast Rec8, a meiosis-specific phosphoprotein involved in recombination events [12]. Brar et al. (2006) showed that phosphorylation of the cohesin subunit REC8 contributes to stepwise cohesin removal [13].

3.3. Constructing Feedback Network of the Distinguished Single Upstream and Downstream Gene. We took into account the feedback relationship and setup ATF2 feedback network, as shown in Figure 3. ATF2 target genes appeared in ATF2 inhibition to CDR2, GLS, and USP11, consistently, its upstream genes also appeared in CDR2, GLS, and USP11 inhibition to ATF2. CDR2 is also called CDR62, where CDR means cerebellar degeneration-related. On Western blot analysis of Purkinje cells and tumor tissue, the anti-Yo sera react with at least 2 antigens, a major species of 62 kD called CDR62 and a minor species of 34 kD called CDR34 [14]. Sahai (1983) demonstrated phosphate-activated glutaminase (GLS) in human platelets [15]. It is the major enzyme yielding glutamate from glutamine. Significance of the enzyme derives from its possible implication in behavior disturbances in which glutamate acts as a neurotransmitter [16]. USP11 is also called UHX1. Swanson et al. (1996) cited evidence indicating that ubiquitin hydrolases play a role in oncogenesis (oncogenes and tumor suppressor gene products are degraded in ubiquitin-dependent pathways) [17]. The relationship of ATF2 with CDR2, GLS, and USP11 represents a negative feedback loop.

3.4. Functional Module Construction of the Distinguished Single Gene. According to ATF2 upstream network, we did DAVID analysis of function cluster, respectively. The DAVID functional annotation clustering results appeared that one ATF2 regulation network was identified as consisting of the ATF2 upstream genes including RBMS1, RNASEH1, PTOV1, NONO, C11orf9, PSMF1, TIA1, TEAD4, GLS, ID2, USP11, TNPO1, PAWR, PLOD2, and TFE3, as shown in Figure 4.

According to Figure 2, it appeared that RBMS1, NONO, PSMF1, TIA1, ID2, PLOD2, TFE3 activate ATF2; whereas RNASEH1, PTOV1, C11orf9, TEAD4, GLS, USP11, TNPO1, and PAWR inhibit ATF2.

RBMS1, NONO, TIA1, ID2, and TFE3 enhance nucleoside, nucleotide, and nucleic acid metabolism because RBMS1, NONO, TIA1, ID2, and TFE3 are involved in these metabolism; PSMF1 activation to ATF2 means the increase of Acyl-CoA metabolism and porphyrin metabolism; PLOD2 activation to ATF2 indicates the progress of cholesterol metabolism and other protein metabolism, as shown in Figure 5.

RNASEH1, PTOV1, and TEAD4 inhibition to ATF2 decreases nucleoside, nucleotide, and nucleic acid metabolism mediated by the three genes; C11orf9 inhibition to ATF2 means the decline of polysaccharide metabolism, whereas GLS represents the weakness of amino acid and cyclic nucleotides metabolism; USP11 inhibition to ATF2 indicates the fall-off in protein metabolism and modification, whereas PAWR in glycogen metabolism, as shown in Figure 5.

RBMS1	rna binding motif, single stranded interacting protein 1	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00039: Other transcription factor, MF00042: Nucleic acid binding, MF00053: Other RNA-binding protein,MF00057: DNA topoisomoerase, MF00068: mRNA splicing factor, MF0007: Chromatin/chromatin-binding protein, MF00076: Other nucleic acid binding, MF00085: Cation transporter, MF00101: Guanyl-nucleotide exchange factor, MF00131: Transferase,MF00175:Major histocompatibility complex antigen, MF00202: Other miscellaneous function protein, MF00208: Molecular function unclassified, MF00213: Non-receptor serine/threonine protein kinase, MF00224: KRAB box transcription factor, MF00232: Interleukin, MF00250: Serine protease inhibitor, MF00259: Cadherin,		
RNASEH1	ribonuclease h1	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00042: Nucleic acid binding, MF00053: Other RNA-binding protein, MF00072: Translation initiation factor, MF00212: Other G-protein modulator,		
PTOV1	prostate tumor overexpressed gene 1	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00031: Voltage-gated ion channel, MF00033: Voltage-gated calcuim channel,MF00036: Transcription factor, MF00075: Ribosomal protein, MF00086: Other transporter, MF00101: Guanyl-nucleotide exchange factor, MF00146: Deacetylase, MF00175: Major histocompatibility complex antigen, MF00202: Other miscellaneous function protein, MF00212: Other G-protein modulator, MF00222: Zinc finger transcription factor, MF00224: KRAB box transcription factor, MF00283: Ubiquitin-protein ligase,		
NONO	non-pou domain containing, octamer-binding	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00042: Nucleic acid binding, MF00065: mRNA processing factor, MF00068: mRNA splicing factor, MF00084: ATP-binding cassette (ABC) transporter, MF00208: Molecular function unclassified,		
C11orf9	chromosome 11 open reading frame 9	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00072: Translation initiation factor, MF00086: Other transporter, MF00101: Guanyl-nucleotide exchange factor, MF00135: Transaldolase, MF00150: Glycosidase, MF00154 : Metalloprotease, MF00174: Complement component, MF00189: Other select calcium binding proteins, MF00208: MF00208: Molescular function unclassified, MF00213: Non-receptor serine/threonine protein kinase, MF00224: KRAB box transcription factor, MF00250: Serine protease inhibitor, MF00279: Tumor necrosis factor receptor,		
PSMF1	proteasome (prosome, macropain) inhibitor subunit 1 (pi31)	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00002: G-protein coupled receptor, MF00006: Interleukin receptor, MF00068: mRNA splicing factor, MF00072: Translation initiation factor, MF00086: Other transporter, MF002101: Protease inhibitor, MF00175 : Major histocompatibility complex antigen, MF00208: Molecular function unclassified, MF00227: Basic helix-loop-helix transcription factor, MF00230 :Actin binding motor protein, MF00240: Immunoglobulin, MF00243: DNA helicase, MF00291: Other enzyme activator,		
TIA1	tia1 cytotoxic granule-associated rna binding protein	Related genes	Homo sapiens
PANTHER_MF_ALL	Mf00042: Nucleic acid binding, MF00053: Other RNA-binding protein, MF00055: Single-stranded DNA-binding protein, MF00212: Other G-protein modulator, MF00231: Microtubule binding motor protein, MF00243: DNA helicase,		
TEAD4	tea domain family member 4	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00036: Transcription factor, MF00039: Other transcription factor, MF00067: mRNA polyadenylation factor, MF00068: mRAN polyadenylation factor, MF00068: mRNA splicing factor, MF00088: Apolipoprotein ,MF00224: KRAB box transcription factor, MF00242: RNA helicase, MF00243: DNA helicase,		
GLS	glutaminase	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00002: G-protein coupled receptor, MF00023: Other signaling molecule, MF00034: Voltage-gated potassium channel, MF00083: Cation transporter, MF00100: G-protein modulator, MF00101: Guanyl-nucleotide exchange factor, MF0013 8: Transaminase, MF00141: Hydrolase, MF00148: Phosphodiesterase, MF00173: Defense/immunity protein, MF00180: Extracellular matrix glycoprotein, MF00231: Microtubule binding motor protein, MF00262: Non-motor actin binding protein,		
ID2	inhibitor of dna binding 2, dominant negative helix-loop-helix protein	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00021: Neuropeptide, MF00036: Transcription factor, MF00039 : Other transcription factor, MF00068: mRNA splicing factor, MF00074: Translation release factor, MF00258: CAM family adhesion molecule,		
USP11	ubiguitin specific peptidase 11	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00034: Voltage-gated potassium channel, MF00101: Guanyl-nucleotide exchange factor, MF00153: Protease, MF00215: Cysteine protease, MF00225: Other zinc finger transcription factor, MF00242: RNA helicase,		
TNPO1	transportin 1	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00087: Transfer/carrier protein, MF00230: Actin binding motor protein, MF00231: Microtubule binding motor protein, MF00261: Actin binding cytoskeletal protein, MF00264: Microtubule family cytoskeletal protein,		
PAWR	prkc, apoptosis, wt1, regulator	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00042: Nucleic acid binding, MF00096: Phosphatase modulator, MF00138: Transaminase, MF00208: Molecular function unclassified, MF00277: Other cell junction protein ,		
TFE3	transcription factor binding to ighm enhancer 3	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00036: Transcription factor, MF00042: Nucleic acid binding, MF00227: Basic helix-loop-helix transcription factor,		
PLOD2	procollagen-lvsine, 2-oxoglutarate 5-dioxyvgenase 2	Related genes	Homo sapiens
PANTHER_MF_ALL	MF00117: Other phosphatase, MF00123: Oxidoreductase, MF00124: Oxygenase, MF00130: Other oxidoreductase, MF00143: Phospholipase, MF00202: Other miscellaneous function protein, MF00208: Molecular function unclassified, MF00212: Other G-protein modulator, MF00213: Non-receptor serine/threonine protein kinase, MF00265: Tubulin,		
RBMS1	rna binding motif, single stranded interacting protein 1	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00040: mRNA transcription, BP00044: mRNA transcription regulation, BP00047: Pre-mRNA processing, BP00048:mRNA splicing, BP00071: Proteolysis, BP00077: Oxidative phosphorylation BP00142: Ion transport, BP00143: Cation transport, BP00149: T-cell mediated immunity, BP00150: MHCI-mediated immunity, BP00151: MHCII-mediated immunity, BP00193: Developmental processes, BP00216: Biological process unclassified, BP00273: Chromatin packaging and remodeling, BP00287: Cell motility,		
RNASEH1	ribonuclease h1	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00143: Cation transport, BP00197: Spermatogenesisand motility, BP00256: RNA catabolism,		
PTOV1	prostate tumor overexpressed gene 1	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00014: Amino acid biosynthesis, BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00044: mRNA transcription regulation, BP00061: Protein biosynthesis, BP00071: Proteolysis, BP00077: Oxisative phosphorylation, BP00104: G-protein mediated signaling, BP00142: Ion transport, BP00143: Cation transport, BP00149: T-cell mediated immunity, BP00150: MHCI-mediated immunity, BP00289: Other metabolism,		
NONO	non-pou domain containing, octamer-binding	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00047: Pre-mRNA processing, BP00048: mRNA splicing, BP00216: Biological process unclassified,		
C11orf9	chromosome 11 open reading frame 9	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00009: Other polysaccharide metabolism, BP00036: DNA repair, BP00044: mRNA transcription regulation, BP00071, Proteolysis, BP00077: Oxidative phosphorylation, BP00104: G-protein mediated signaling, BP00111: Intracellular signaling cascade, BP00112: Calcium mediated signaling, BP00153: Complement-mediated immunity, BP00216: Biological process unclassified, BP00273: Chromatin packaging and remodeling, BP00286: Cell structure,		
PSMF1	proteasome (prosome, macropain) inhibitor subunit 1 (pi31)	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00024: Acyl- CoA metabolism, BP00040: mRNA transcription, BP00044: mRNA transcription regulation, BP00071: Proteolysis, BP00087: Porphyrin metabolism, BP00103: Cell surface receptor mediated signal transduction, BP00104: G-protein mediated signaling, BP00119: Other intracellular signaling cascade, BP00122: Ligand-mediated signaling, BP00149: T-cell mediated immunity, BP00150: MHCI-mediated immunity, BP00151: MHCII-mediated immunity, BP00152: B-cell and antibody-mediated immunity, BP00216: Biological process unclassified, BP00274: Cell communication,		
TIA1	tia1 cytotoxic granule-associated rna binding protein	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00047: Pre-mRNA processig, BP00048: mRNA splicing,		
TEAD4	tea domain family member 4	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00040: mRNA transcription, BP00044: mRNA transcription, BP00044: mRNA transcription regulation,		
GLS	glutaminase	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00013: Amino acid metabolism, BP00014: Amino acid biosynthesis, BP00036: DNA repair, BP00042: mRNA transcription initiation, BP00047: Pre-mRNA processing, BP00049: mRNA polyadenylation, BP00056: Metabolism of cyclic nucleotides, BP00071: Proteolysis, BP00090: Nitrogen metabolism, BP00102: Signal transduction, BP00142: Ion transport, BP00143: Cation transport, BP00289: Other metabolism,		
ID2	inhibitor of dna binding 2, dominant negative helix-loop-helix protein	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00040: mRNA transcription, BP00044: mRNA transcription regulation, BP00048: mRNA splicing, BP00071: Proteolysis, BP00104: G-protein mediated signaling, BP00128: Constitutive exocytosis, BP00148: Immunity and defense, BP00273: Chromatin packaging and remodeling,		
USP11	ubiguitin specific peptidase 11	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00060: Protein metabolism and modification, BP00071: Proteolysis, BP00104: G-protein mediated signaling, BP00143: Cation transport, BP00179: Apoptosis, BP00250: Muscle development,		
TNPO1	transportin 1	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00063: Protein modification, BP00064: Protein phosphorylation, BP00125: Intracellular protein traffic, BP00194: Gametogenesis, BP00196: Oogensis,		
PAWR	prkc, apoptosis, wt1 regulator	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00040: mRNA trancription, BP00043: mRNA transcription elongation, BP00044: mRNA transcription regulation, BP00179: Apoptosis, BP00298 : Glycogen metabolism,		
TFE3	transcription factor binding to ighm enhancer 3	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00031: Nucleoside, nucleotide and nucleic acid metabolism, BP00040: mRNA transcription, BP00044: mRNA transcription regulation,		
PLOD2	procpllagen-lysine, 2-oxoglutarate 5-dioxygenase 2	Related genes	Homo sapiens
PANTHER_BP_ALL	BP00026: Cholesterol metabolism, BP00041: General mRNA transcription activities, BP00060: Protein metabolism and modification, BP00061: Protein biosynthesis, BP00075: Other protein metabolism, BP00104: G-protein mediated signaling, BP00142: Ion transport, BP00150: MHCI-mediated immunity, BP00216: Biological process unclassified, BP00268: Antioxiadation and free radical removal,		

Figure 5: Molecular function and biological process from DAVID.

4. Conclusions

Our method concentrates on and constructs the distinguished single gene network integrated with function prediction analysis by DAVID. For the distinguished single molecular network, we did (1) control and experiment comparison, (2) identification of activation and inhibition networks, (3) construction of upstream and downstream feedback networks, and (4) functional module construction. We tested this method to identify ATF2 regulation network module using data of 45 samples from one and the same GEO dataset. The results demonstrate the effectiveness of such integrated way in terms of developing novel prognostic markers and therapeutic targets.

Acknowledgments

This work was supported by the National Natural Science Foundation in China (no. 60673109 and no. 60871100) and the Teaching and Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, State Key Lab of Pattern Recognition Open Foundation.

References

[1] Y. Wang, T. Joshi, X.-S. Zhang, D. Xu, and L. Chen, "Inferring gene regulatory networks from multiple microarray datasets," *Bioinformatics*, vol. 22, no. 19, pp. 2413–2420, 2006.

[2] D. W. Huang, B. T. Sherman, Q. Tan, et al., "The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists," *Genome Biology*, vol. 8, no. 9, article R183, pp. 1–16, 2007.

[3] D. W. Huang, B. T. Sherman, Q. Tan, et al., "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists," *Nucleic Acids Research*, vol. 35, web server issue, pp. W169–W175, 2007.

[4] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[5] T. Byrt, J. Bishop, and J. B. Carlin, "Bias, prevalence and kappa," *Journal of Clinical Epidemiology*, vol. 46, no. 5, pp. 423–429, 1993.

[6] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 5116–5121, 2001.

[7] M. B. Humphrey, H. Herrera-Sosa, G. Gonzalez, R. Lee, and J. Bryan, "Cloning of cDNAs encoding human caldesmons," *Gene*, vol. 112, no. 2, pp. 197–204, 1992.

[8] J. A. Williamson, J. M. Bosher, A. Skinner, D. Sheer, T. Williams, and H. C. Hurst, "Chromosomal mapping of the human and mouse homologues of two new members of the AP-2 family of transcription factors," *Genomics*, vol. 35, no. 1, pp. 262–264, 1996.

[9] J. M. Bosher, T. Williams, and H. C. Hurst, "The developmentally regulated transcription factor AP-2 is involved in c-erbB-2 overexpression in human mammary carcinoma," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 3, pp. 744–747, 1995.

[10] P. S. Henthron, C. C. Stewart, T. Kadesch, and J. M. Puck, "The gene encoding human TFE3, a transcription factor that binds the immunoglobulin heavy-chain enhancer, maps to Xp11.22," *Genomics*, vol. 11, no. 2, pp. 374–378, 1991.

[11] Y. Nakagawa, H. Shimano, T. Yoshikawa, et al., "TFE3 transcriptionally activates hepatic IRS-2, participates in insulin signaling and ameliorates diabetes," *Nature Medicine*, vol. 12, no. 1, pp. 107–113, 2006.

[12] S. Parisi, M. J. McKay, M. Molnar, et al., "Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family conserved from fission yeast to humans," *Molecular and Cellular Biology*, vol. 19, no. 5, pp. 3515–3528, 1999.

[13] G. A. Brar, B. M. Kiburz, Y. Zhang, J.-E. Kim, F. White, and A. Amon, "Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis," *Nature*, vol. 441, no. 7092, pp. 532–536, 2006.

[14] H. Fathallah-Shaykh, S. Wolf, E. Wong, J. B. Posner, and H. M. Furneaux, "Cloning of a leucine-zipper protein recognized by the sera of patients with antibody-associated paraneoplastic cerebellar degeneration," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 8, pp. 3451–3454, 1991.

[15] S. Sahai, "Glutaminase in human platelets," *Clinica Chimica Acta*, vol. 127, no. 2, pp. 197–203, 1983.

[16] S. B. Prusiner, "Disorders of glutamate metabolism and neurological dysfunction," *Annual Review of Medicine*, vol. 32, pp. 521–542, 1981.

[17] D. A. Swanson, C. L. Freund, L. Ploder, R. R. McInnes, and D. Valle, "A ubiquitin C-terminal hydrolase gene on the proximal short arm of the X chromosome: implications for X-linked retinal disorders," *Human Molecular Genetics*, vol. 5, no. 4, pp. 533–538, 1996.