Construction of a gene-gene interaction network with a combined score across multiple approaches

A.M. Zhang, H. Song, Y.H. Shen and Y. Liu

Department of Genitourinary Surgery,
Jinan Military General Hospital of Chinese PLA, Jinan, Shandong, China

Corresponding author: A.M. Zhang
E-mail: aiminzhangmed@yeah.net

ABSTRACT. Recent progress in computational methods for investigating physical and functional gene interactions has provided new insights into the complexity of biological processes. An essential part of these methods is presented visually in the form of gene interaction networks that can be valuable in exploring the mechanisms of disease. Here, a combined network based on gene pairs with an extra layer of reliability was constructed after converting and combining the gene pair scores using a novel algorithm across multiple approaches. Four groups of kidney cancer data sets from ArrayExpress were downloaded and analyzed to identify differentially expressed genes using a rank products analysis tool. Gene co-expression network, protein-protein interaction, co-occurrence network and a combined network were constructed using empirical Bayesian meta-analysis approach, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database, an odds ratio formula of the cBioPortal for Cancer Genomics and a novel rank algorithm with combined score, respectively. The topological features of these networks were then compared to evaluate their performances. The results indicated that the gene pairs and their relationship rank-

ings were not uniform. The values of topological parameters, such as clustering coefficient and the fitting coefficient R^2 of interaction network constructed using our ranked based combination score, were much greater than the other networks. The combined network had a classic small world property which transferred information quickly and displayed great resilience to the dysfunction of low-degree hubs with high-clustering and short average path length. It also followed distinctly a scale-free network with a higher reliability.

Key words: Gene interaction; Empirical Bayesian model; cBioPortal; Protein-protein interaction; Scale-free network; Small word network

INTRODUCTION

A common goal of microarray and related high-throughput genomic experiments is to identify differentially expressed genes (DEGs) that provide insight into the understanding and treatment of disease. However, it does not accommodate other types of differential regulation (Braga-Neto and Dougherty, 2004), despite being useful to identify DEGs that vary across biological conditions. Meanwhile, diseases result from differential regulation (upregulation or downregulation) of genes that are frequently related and have a function in similar biological processes and pathways (Lee et al., 2004). It would be an important milestone towards a comprehensive description of disease mechanisms, knowing functional connections varying between stable complexes, metabolic pathways and a bewildering array of direct and indirect regulatory interactions between DEGs.

It is essential to correctly uncover and annotate all functional interactions between genes for any systems-level understanding of biological functions. With the development of bioinformatics, a large number of approaches have been proposed to evaluate gene interactions based on protein or nucleotide sequences on a large scale (Skrabanek et al., 2008; Zhang et al., 2014), both in terms of experimental measurements and computational prediction techniques. Among them, an empirical Bayesian (EB) approach has been applied, which can combine the results analyzed from our experimental data with meta-analysis findings. It effectively avoids the disadvantage of inconsistent co-expression between different studies by providing a false discovery rate (FDR) controlled list of significantly differentially co-expressed gene pairs (Dawson and Kendziorski, 2012). It is applicable within a single study as well as across multiple studies with considerable power by conducting an analysis under different conditions. Besides, it provides a useful tool to understand gene/protein interactions with PPI of the STRING database, which provides a global perspective for as many organisms as possible as an online resource. It can integrate and rank protein/gene associations by benchmarking them against a common reference set, and presents evidence in a consistent and intuitive web interface (Von Mering et al., 2005). In addition, it can also illustrate gene interactions in genetic alterations with an odds ratio (OR) formula which can describe the patterns of mutual exclusivity, co-occurrence or no association between multiple genes. cBioPortal is applied here, which facilitates the exploration of multidimensional cancer genomics data by allowing visualization and analysis across genes, samples, and data types (Gao et al., 2013).

Network-based approaches have become more powerful and informative for the study of disease mechanism (Bradley et al., 2008). Especially the analysis of gene interaction net-

works has proven effective for characterizing cellular processes in biologic research. In these networks, genes and their relationships are graphically represented as nodes and edges, respectively. On the other hand, at some extent, it sets up obstacles of useful information extracted from the various kinds of network data with the increase in large bioinformatic data. Currently, research on complex networks has attracted a great deal of attention since the discovery of the small-world phenomenon and scale-free property (Watts and Strogatz, 1998; Barabási and Albert, 1999). Compared to the random network, a scale-free network is a network whose node's degree follows a power law distribution, and the scale-free topology characteristics have a higher robustness to endure the random failure (Chen et al., 2009). The small-world network has two independent structural features with a small average shortest path length and a large clustering coefficient (Gitterman, 2000).

In this paper, we applied a novel algorithm to combine the interaction scores of each gene pair obtained using a variety of methods. The topological features, including clustering coefficient, average shortest path length and degree of distribution, were investigated and compared to see which tended to follow more a scale-free network and small word network after the construction of gene interaction networks of different methods. First, we focused on kidney cancer and identified DEGs using the RP package. To compare the approaches for analyzing gene interactions, the related scores of gene pairs were obtained using EB coexpress meta analysis, STRING database and OR algorithm. Considering the faults of predicted outcomes involving the EB model and STRING approach, we converted and united all scores from the three methods using a rank based model and obtained a combined score of each gene pairs. The gene co-expression network, PPI, co-occurrence and a combined network according to the above were then constructed, and their topological properties were further analyzed. This may provide a new tool to analyze gene interactions with a higher reliability strength and rapid information transmission combining the scores of each gene pair across multiple approaches.

MATERIAL AND METHODS

Subject samples and pretreatment process

Transcription expression profiles of kidney cancer biological data (E-GEOD-53757), (E-GEOD-36895) (Peña-Llopis et al., 2012), (E-GEOD-26574) and (E-GEOD-46699) were available from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). After pretreatment of these data by robust multi-array average (RMA) (Ma et al., 2006), quartile based algorithm (Rifai and Ridker, 2001), MAS5 algorithm (Zhang et al., 2003) and median polish summarization methods, unqualified chips were eliminated, leaving only qualified data into the next step through quality control. The gene expression values of all data were transformed to a comparable level, a digital expression profile for subsequent analysis. The data were then screened by a feature filter method of the genefilter package, and the probe-sets associated with one gene were analyzed by the maxim-based method to choose the most significant differentially expressed ones. Finally, an expression profile dataset including 20,107 common genes was obtained for further analysis.

Detection of DEGs

Considering that four kidney cancer data were from different origins, we detected

DEGs using rank products (RP) (Breitling et al., 2004), which is a powerful meta-analysis tool for integrating multiple array data sets from various experimental platforms. This method could be used to know how significant the changes were and how many of the selected genes were likely to be truly differentially expressed. Here, we let T and C stand for two experimental conditions (treatment versus control), and there are n_T and n_C, m_T and m_C, s_T and s_C, and w_T and w_C replicates in the first, second, third and fourth dataset, respectively.

Rank product for each gene was determined according to (1):

$$RP_g = (\prod_i r_{gi})^{1/K} \qquad \text{(Equation 1)}$$

where r_{gi} is rank of gth gene under ith comparison. $i = 1, \ldots, K$, $K = (n_T \times n_C) + (m_T \times m_C) + (s_T \times s_C) + (w_T \times w_C)$.

Rank product for each gene was presented in the form of -log2, and the percentage of false-positives (pfp) smaller than 0.05 was considered to be differentially expressed. The pfp was given by the expression:

$$E_{RP}(g) = c / p \qquad \text{(Equation 2)}$$

$$pfp(g) = E_{RP}(g) / rank(g) \qquad \text{(Equation 3)}$$

$E_{RP}(g)$ is the average expected value for the rank product; p is the permutations of K rank lists of length n; c value is the times that the rank products of the genes in the permutations are smaller or equal to the observed rank product; $pfp(g)$ is the percentage of false positives; $rank(g)$ is the rank of gene g in a list of all genes sorted by increasing RP.

Construction of gene-gene interaction network for DEGs

Identifying differential co-expression by empirical Bayesian approach

Currently, many approaches have been developed for co-expression analysis to identify differentially co-expressed (DC) gene pairs, but they are often prone to false discoveries under the conditions of large cardinality of the space to be interrogated (Cho et al., 2009). Here, we conducted an EB approach, which provided an FDR controlled list of interesting pairs along with pair-specific posterior probabilities to identify DC gene pairs (Dawson and Kendziorski, 2012). DC genes were distinguished from gene pairs having invariant expression by controlling the posterior expected FDR at 0.05, and the co-expression network was constructed to represent the correlation between each pair of genes in the study.

Scoring of gene associations using STRING database

At present, protein/gene interactions and associations are annotated at various levels

of detail, ranging from raw data repositories to highly formalized pathway databases in online resources. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) aims to simplify access to this information by providing a comprehensive, yet quality-controlled, collection of protein-protein associations for a large number of organisms with a global perspective. Most of the available information on protein/gene associations could be aggregated, scored and weighted with known and predicted interactions. Besides, protein interactions across diverse experimental conditions could be measured as a predictor of functional associations in STRING, which was used in our study. It employs two different strategies for transferring known and predicted associations between organisms (Von Mering et al., 2005). After assignment of association scores and transfer between species, a combined score between any pair of proteins was computed which increased confidence with a higher score than the individual sub-scores. This combined score took into account the prediction and known scores obtained from the STRING database from each protein interaction and was calculated according to the formula (4) shown below:

$$S_{AB} = 1 - \prod_i (1 - S_i)$$

(Equation 4)

where S_{AB} is the score for the interaction between proteins A and B, and S_i is the score normalized by the biggest value calculated for the method i.

Afterwards, a graphical PPI network was constructed and the topological features of the network were further analyzed.

Co-occurrence pattern of multiple genes in gene mutation

In addition to the above two methods of analyzing gene interactions, another strategy of determining gene co-occurrence pattern involving gene mutation has also been cited to identify functionally associated genes. In our study, DEGs were input in the cBioPortal for Cancer Genomics (http://www.cbioportal.org), which was used to store genomic data from large scale, integrated cancer genomic data sets, and renal clear cell carcinoma (Network, 2013) was selected as the cancer study in this research. With the genomic profiles filtered further, a total of 392 cases associated with kidney cancer were selected to explore whether gene pairs pertained to the co-occurrence model while genetic alterations occurred in multiple genes. The results indicated genes altered in 141 cases, or 36 of all samples. An odds ratio (OR) was determined, which indicated the likelihood that the events in the each pair of query genes G_A and G_B were co-occurring in 141 cases. The formula (5) is given as follows:

$$OR = (N_b \times N_n) / (N_a \times N_b)$$

(Equation 5)

where: N_b indicates the number of cases altered in both genes; N_n indicates number of cases altered in neither genes; N_a and N_b indicate number of cases altered in only one gene, G_A and G_B, respectively.

After the co-occurrence score (i.e., P value) of each pair of DEGs was computed, a

network was constructed via linking gene pairs with P < 0.05 using Cytoscape, a free software package for visualizing, modeling and analyzing the integration of biomolecular interaction networks with high-throughput expression data and other molecular states (Shannon et al., 2003).

Conversion and combination of gene association scores of the three methods

After gene interactions were analyzed using the above three methods, the score of each gene pair was obtained. Considering that the results were different by taking various approaches, all score values were processed further to keep uniform at the same standard. To reach the goal, a novel algorithm was implemented to convert the scores of all gene pairs in our study, using the expression (6) given as follows:

$$S_{com} = \frac{1}{n} \sum_{1 \leq N \leq M} (-2 \log N / M)$$

(Equation 6)

where S_{com} indicates the combined score of all gene pairs with integrating multiple results, n denotes the number of methods, N = 3 in this study. M represents the number of gene pairs of the DEGs, N denotes rank of a pair of genes. With combining the scores of each gene pair from the above three results, a new score was produced by calculating their averages and sorted again. The combined gene interaction network was then constructed via linking gene pairs while controlling the N/M at 10% (i.e., -2logN/M >= 6.643856).

After the calculation of scores using the above methods and construction of four networks, the clustering coefficient and their short average path length were also obtained and compared to investigate whether the networks constructed from the above four methods had the classic small word network property. Besides, since protein/gene interaction networks in general are modular (Ravasz et al., 2002) and scale-free (Barabási and Albert, 1999), which meant that they had power-law (or scale-free) degree distributions, the fitting coefficient R^2 of the power-law $y = ax^b$ of the four networks was also compared. The evaluation of topological parameters was done using Network Analyzer Version 2.7 (Assenov et al., 2008) plug-in at Cytoscape Version 3.1.0 (Morris et al., 2014).

RESULTS

Detection of DEGs between kidney cancer and normal

Four kidney cancer microarray data sets for normal and diseased subjects from different origins were subjected to integrated analysis to identify DEGs using the Rankprod method. The identification in the Rankprod package consists of two parts, the identification of upregulated and downregulated genes in class 2 compared to class 1, respectively. After combining the data of the four groups and analyzing them using RPadvance function, 90 DEGs were detected, including 28 upregulated genes and 62 downregulated genes with an estimated pfp < 0.05 (Figure 1).
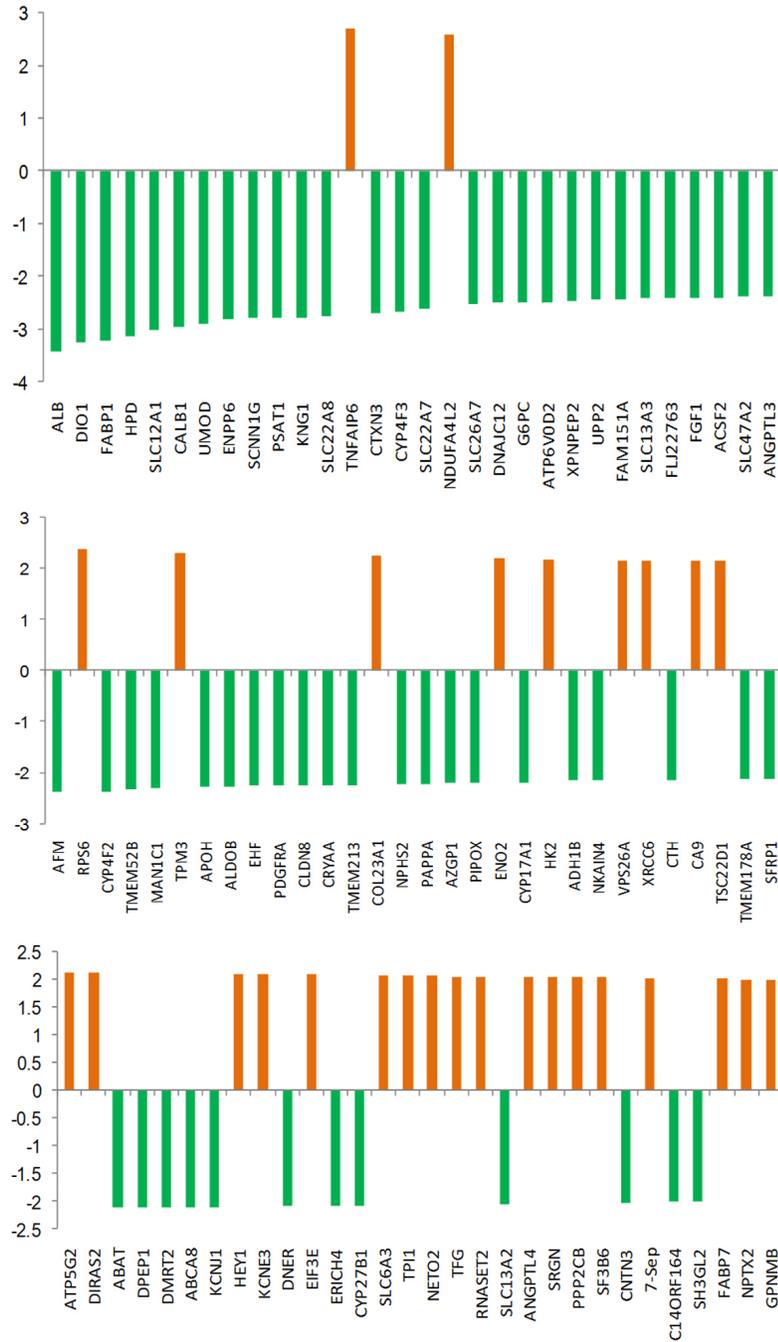
Figure 1. Column Chart of degree of differential expression (fold-change) of 90 DE genes in the four kidney cancer data. Plotted above the X-axis with red color for upregulation, while below the X-axis with red color for downregulation.

EBcoexpress approach and construction of gene co-expression network

Here, the empirical Bayesian approach was used to identify DC gene pairs separately for 90 DEGs. A list of 627 protein pairs with FDR of 5% and relational values of all pairs (i.e., EB values) were obtained after analyzing gene expression relationship using meta-analysis. We found that each pair of a total of 76 protein pairs showed a strong correlation, with EB score of 1. SCNN1G and ATP6V0D2, as the top two degree ranked of the 76 pairs, connected closely with the other 15 and 11 proteins, respectively. The degree indicated the number of interaction of a protein with the other proteins. To improve the comparability of the results with the four methods, the minimum edges were chosen as a benchmark to analyze. A gene interaction network of the top 39 protein pairs containing 26 nodes and 39 edges was constructed in our analysis (Figure 2A). The network was binary, where all interactions were unweighted and undirected. Besides, the clustering coefficient was 0.323 and mean shortest path 16.88. Degrees of proteins were presented, and a fitting coefficient R^2 of 0.804 of their degree distribution was obtained after nonlinear regression according to the power law ($y = ax^b$).
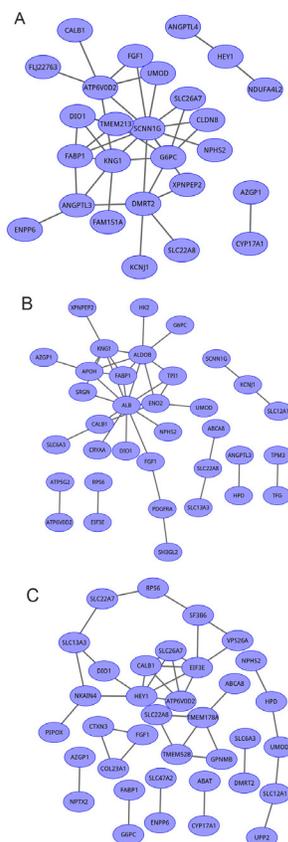

Figure 2. Graphical representation of (**A**) EB co-expression network, (**B**) PPI (**C**) co-occurrence network. Genes were denoted as nodes in the graph and interactions between gene pairs were presented as edges. To facilitate comparisons, there were 26, 35, and 36 nodes in these three networks, respectively at the same benchmark of the top ranked 39 gene pairs.

Scoring of protein associations and construction of PPI network

Considering the various major sources of association, data in STRING are benchmarked independently. In our study, a combined score was computed with the known and predicted associations of STRING data. The combined score indicated higher confidence when more than one type of information supported a given association. Each score represented a rough estimate of how likely a given association described a regulation between two proteins. A graphical protein-protein interaction (PPI) network (Figure 2B) of interaction partners was then presented with 35 nodes and 39 edges, and relational values of all pairs (i.e., STRING values) were obtained in the context of inputting 90 DEGs. The clustering coefficient was 0.151 and mean shortest path 31.34. After degree distribution was conducted by nonlinear curve fit according to the power law ($y = ax^b$), a fitting coefficient ($R^2 = 0.810$) was obtained. The top three degree ranked proteins were ALB, ALDOB and KNG1 by associating with the other 12, 8 and 6 proteins, respectively. ALB and KNG1 (score = 0.982), ALB and SRGN (score = 0.969), TPI1 and ALDOB (score = 0.954) were the top three protein pairs with high interactions.

Detection of gene co-occurrence and gene interaction network construction

The need for similar regulation is often reflected in a tendency of functionally associated genes to occur in a specific disease. The co-occurrence pattern of gene mutations was also applied to evaluate gene interaction. With the above 90 DEGs inputted in cBioPortal, co-occurrence scores of all 4006 pairs were obtained. The five most significant gene pairs were COL23A1 and FGF1, COL23A1 and CTXN3, FGF1 and CTXN3, ATP6V0D2 and SLC26A7, and TMEM178A and TMEM52B with the same score ($P = 1.00 \times 10^{-6}$). The gene interaction network was constructed with 36 nodes and 39 edges by choosing the top 39 pairs after the protein pairs with $P < 0.05$ were loaded into the Cytoscape software (Figure 2C). Likewise, the degrees of all nodes were presented and a fitting coefficient ($R^2 = 0.891$) of their degree distribution was obtained after nonlinear regression.

Combination of all gene pair scored and construction of gene interaction network

After all score values of the above three methods were unified, the scores of each gene pair were combined and an integrated score rank was presented with their averages. The top three ranked gene pairs were FABP1 and KNG1, KCNJ1 and SLC12A1, and APOH and FABP1 with a mean of 11.66, 11.63, and 11.14, respectively.

The gene interaction network was constructed containing 42 nodes and 57 edges from the top 57 gene pairs with N/M <= 10% (Figure 3A). SCNN1G, KNG1, and ATP6V0D2 were viewed as hub genes with a degree of 12, 7, and 6, respectively. According to further analysis, we found that the combined interaction network conformed to the scale-free network whose degree distribution followed the power law $y = ax^b$ (where a = 12.62, b = -1.194) (Figure 3B). Besides, the clustering coefficient, the mean shortest path (Figure 4) and the fitting coefficient R^2 were compared (Table 1) by aggregating the above four methods. It showed that the values of the clustering coefficient and fitting coefficient R^2 of the combined network were much

greater than the other three kinds of networks. It tended to be much more a scale-free network, adding an extra layer of reliability, and had the small world network feature with a short average path length between nodes, which could transfer information quickly at "low cost." *A priori*, it could also display great resilience to the dysfunction of low-degree hubs in this kind of network.
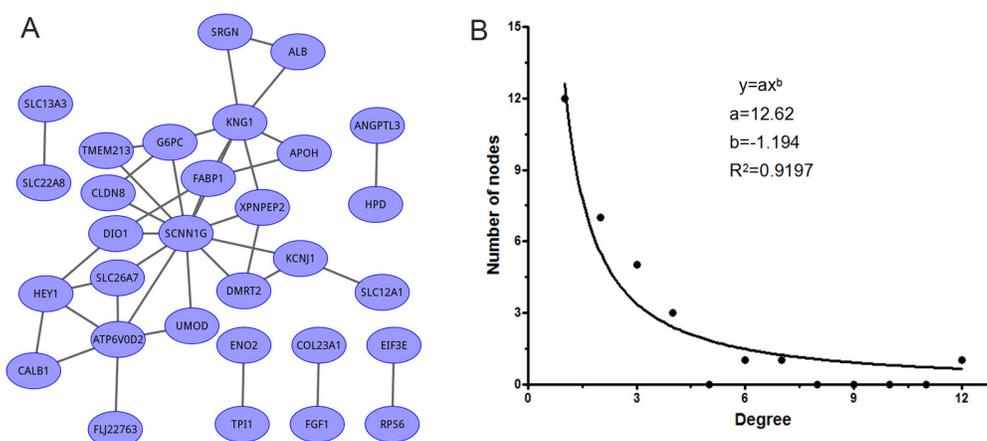

Figure 3. Gene interactions across multiple approaches based on 90 DE genes **A:** Gene interactions network map of the combined scores of each gene pairs with the three methods. A total of 35 nodes and 39 edges were composed in this network. **B:** Scatter Gram of DE genes in gene interactions network. The combined gene interactions network was a scale-free network whose gene degree distribution followed a power law $y = ax^b$ (where a = 12.62, b = -1.194).
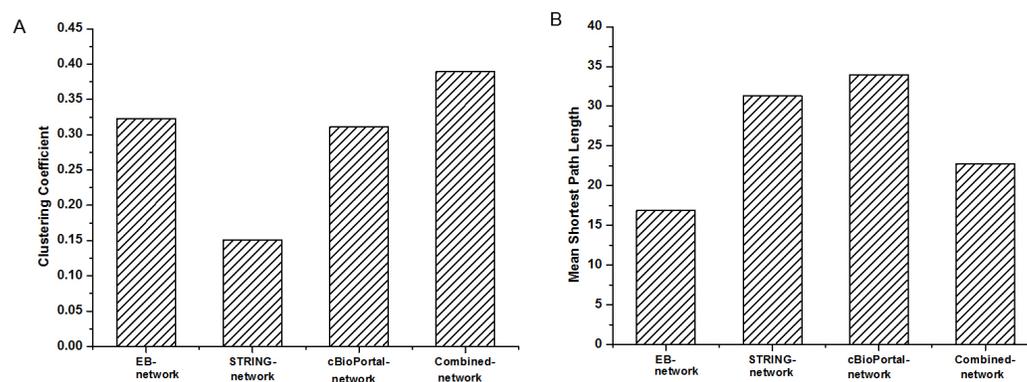

Figure 4. Comparison of clustering coefficient and mean shortest path length between the networks. The combined network presented a classic small word network property with the high-clustering and a short average path length.

Table 1. Fitting results and Clustering Coefficient, Mean Shortest Path Length for the gene co-expression network, PPI, co-occurrence network compared against the combined network after converting with our new algorithm.

Measure	Co-expression network	PPI network	Co-occurrence network	Combined network
R^2	0.804	0.810	0.891	0.920
Clustering Coefficient	0.323	0.151	0.311	0.389
Mean Shortest Path Length	16.88	31.34	33.97	22.73

DISCUSSION

Here, the gene interaction network of DEGs was constructed with a list of combined scores of gene pairs across multiple approaches which followed obviously a scale-free network and small word network. The combined network enables a topological characterization of the reliability strength of gene associations. Four sets of kidney cancer data were available for normal and diseased subjects and a total of 90 DEGs were identified using RP package. Gene interaction values of the 90 DEGs were computed with the EB coexpress meta-analysis approach, STRING database and OR algorithm with cBioPortal, and gene co-expression network, PPI, co-occurrence and combined networks were constructed visually. Here, to facilitate comparisons, the same benchmark of the top ranked 39 gene pairs were selected (input to network generation). New scores of each gene pair, combining the above three methods results of a novel algorithm, were produced in the form of -2logN/M, and the combined score gene interaction network was presented of the same 39 edges of the former. After the degree distribution, clustering coefficient and mean shortest path length of topological properties were compared. The combined gene interaction network was revealed evidently a scale-free network feature and held a classic small world network.

The capabilities of bioinformatic tools for DEG detection, network analysis, gene ontology and gene-disease relationships (Özgür et al., 2008; Tranchevent et al., 2011), together with all available data on protein/gene expression in cancer, provide an interesting and valuable opportunity for disease study. It is thought that the incidence of cancer is closely related to the abnormal expression of many genes. However, much more research is needed beyond only identifying DEGs to fully understand the complex mechanism of cancer. It is essential to explore the many gene interactions that participate in functionally biological processes. Many papers have reported gene interactions ranging from a single study (Jupiter and VanBuren, 2008) to combining pairwise gene interaction evidence across studies by the vote-counting method (Niida et al., 2009). However, these approaches are often underpowered and computationally intractable for even a moderately large number of pairs, while empirical Bayesian approach provides an FDR controlled list of interesting pairs from a high-throughput experiment measuring expression in two or more conditions without sacrificing power. Huttenhower et al. (2006) applied this method of Bayesian framework to combine studies for pairwise meta-correlation and predicted functional relationships. Besides, gene associations are also commonly analyzed with the STRING database. The co-regulation of genes across diverse experimental conditions, as measured by using microarray analysis, which is covered in the STRING database, can be a predictor of functional associations (Stuart et al., 2003). In addition, gene interaction analysis is also achieved in the co-occurrence pattern of gene alterations with the OR formula using the cBioPortal by storing cancer experimental results across multiple genomic data types. Here, we analyzed and compared DEG interactions using the EB

approach, and the resources of the STRING database and the cBioPortal software, which was comprehensive from the aspects of gene co-expression, mutation and protein associations.

Genetic interactions have proven highly effective for determining gene functions and identifying groups of genes that encode proteins in the same pathway or complex. Interactome network maps have been used to demonstrate gene interactions visually between biological components. Such gene association networks have allowed the generation of novel hypotheses about potential functional roles of genes or about their involvement in phenotype-specific cellular processes (Iancu et al., 2012). In this network, its analyses typically work with the gene-gene interaction matrix, which graphically represents the correlation between each pair of genes in the study. Here, gene co-expression, PPI and co-occurrence networks were constructed with three different methods, respectively. From the point of the flaw of EB model with weaker accuracy and STRING existing prediction factor, a new gene interaction network was produced by assembling the scores of gene pairs of the three methods. The four networks were all highly nonrandom, where the clustering coefficient, the mean shortest path and the fitting coefficient R^2 according to the power law ($y = ax^b$) were evaluated. The network constructed by combining the scores of three methods was prior to showing the scale-free and small world network, which may be a new approach to analyze gene interactions.

Interestingly, FGF1, CALB1, FABP1, UMOD, and DIO1 were identified as common genes in four networks, and three of them were closely associated with tumor. FGF1 is involved in pathological processes such as restenosis and tumor formation. It is frequently upregulated in hepatocellular carcinoma and its overexpression may be a consequence rather than contributor to hepatoma progression (Prudovsky et al., 2003; Huang et al., 2006). CALB1 is preferentially expressed in dopamine neurons that are spared from degeneration in Parkinson's disease (Airaksinen et al., 1997). It is also expressed in an insulin-producing tumor cell line derived from a radiation induced rat insulinoma (Pochet et al., 1989). DIO1 activity is an important determinant of thyroid status; it could help to better define a tumor signature for thyroid tumors (Arnaldi et al., 2005). All these genes may be useful for future investigations of molecular biomarker. Functional studies are necessary to elucidate the role of these genes in kidney cancer pathogenesis and determine their potential as molecular targets for the development of new kidney cancer therapeutic approaches.

ACKNOWLEDGMENTS

REFERENCES

Airaksinen MS, Thoenen H and Meyer M (1997). Vulnerability of Midbrain Dopaminergic Neurons in Calbindin-D28k-deficient Mice: Lack of Evidence for a Neuroprotective Role of Endogenous Calbindin in MPTPtreated and Weaver Mice. *Eur. J. Neurosci.* 9: 120-127.

Arnaldi L, Borra R, Maciel R and Cerutti J (2005). Gene expression profiles reveal that DCN, DIO1, and DIO2 are underexpressed in benign and malignant thyroid tumors. *Thyroid* 15: 210-221.

Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, et al. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24: 282-284.

Barabási AL and Albert R (1999). Emergence of scaling in random networks. *Science* 286: 509-512.

Bradley EW, Ruan MM, Vrable Aand Oursler MJ (2008). Pathway crosstalk between Ras/Raf and PI3K in promotion of M-CSF-induced MEK/ERK-mediated osteoclast survival. *J. Cell Biochem.* 104: 1439-1451.

Braga-Neto UM and Dougherty ER (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics* 20: 374-380.

Breitling R, Armengaud P, Amtmann A and Herzyk P (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573: 83-92.

Chen L-J, Liu M, Chen DX and Xie L (2009). Topology evolution of wireless sensor networks among cluster heads by random walkers. *Chin. J. Com.* 32: 69-76.

Cho SB, Kim J and Kim JH (2009). Identifying set-wise differential co-expression in gene expression microarray data. *BMC Bioinformatics* 10: 109.

Dawson JA and Kendziorski C (2012). An empirical Bayesian approach for identifying differential coexpression in high-throughput experiments. *Biometrics* 68: 455-465.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6: pl1.

Gitterman M (2000). Small-world phenomena in physics: the Ising model. *J. Phys. A: Math. Gen.* 33: 8373.

Huang X, Yu C, Jin C, Yang C, et al. (2006). Forced expression of hepatocyte-specific fibroblast growth factor 21 delays initiation of chemically induced hepatocarcinogenesis. *Mol. Carcinog.* 45: 934-942.

Huttenhower C, Hibbs M, Myers C and Troyanskaya OG (2006). A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* 22: 2890-2897.

Iancu OD, Kawane S, Bottomly D, Searles R, et al. (2012). Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics* 28: 1592-1597.

Jupiter DC and VanBuren V (2008). A visual data mining tool that facilitates reconstruction of transcription regulatory networks. *PLoS One* 3: e1717.

Lee HK, Hsu AK, Sajdak J, Qin J, et al. (2004). Coexpression analysis of human genes across many microarray data sets. *Genome Res.* 14: 1085-1094.

Ma L, Robinson LN and Towle HC (2006). ChREBP• Mlx Is the Principal Mediator of Glucose-induced Gene Expression in the Liver. *J. Biol. Chem.* 281: 28721-28730.

Morris JH, Wu A, Doncheva NT, Albrecht M, et al. (2014). setsapp: Set operations for cytoscape nodes and edges. Available at [http://f1000r.es/3ml].

Network CGAR (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499: 43-49.

Niida A, Imoto S, Nagasaki M, Yamaguchi R, et al. (2009). A novel meta-analysis approach of cancer transcriptomes reveals prevailing transcriptional networks in cancer cells. *Genome Inform.* 22: 121-131.

Özgür A, Vu T, Erkan G and Radev DR (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 24: i277-i285.

Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, et al. (2012). BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* 44: 751-759.

Pochet R, Blachier F, Malaisse W, Parmentier M, et al. (1989). Calbindin-D28 in mammalian brain, retina, and endocrine pancreas: immunohistochemical comparison with calretinin. *Adv. Exp.Med. Bio.* 255: 435-443.

Prudovsky I, Mandinova A, Soldi R, Bagala C, et al. (2003). The non-classical export routes: FGF1 and IL-1α point the way. *J. Cell Sci.* 116: 4871-4881.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, et al. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555.

Rifai N and Ridker PM (2001). Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. *Clini. Chem.* 47: 28-30.

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504.

Skrabanek L, Saini HK, Bader GD and Enright AJ (2008). Computational prediction of protein-protein interactions. *Mol. Biotechnol.* 38: 1-17.

Stuart JM, Segal E, Koller D and Kim SK (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302: 249-255.

Tranchevent L-C, Capdevila FB, Nitsch D, De Moor B, et al. (2011). A guide to web tools to prioritize candidate genes. *Brief. Bioinform.* 12: 22-32.

Von Mering C, Jensen LJ, Snel B, Hooper SD, et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33: D433-D437.

Watts DJ and Strogatz SH (1998). Collective dynamics of 'small-world' networks. *Nature* 393: 440-442.

Zhang L, Miles MF and Aldape KD (2003). A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.* 21: 818-821.

Zhang W, Zang Z, Song Y, Yang H, et al. (2014). Co-expression network analysis of differentially expressed genes associated with metastasis in prolactin pituitary tumors. *Mol. Med. Rep.* 10: 113-118.