Iran Red Crescent Med J. 2016 June; 18(6):e24373. doi: 10.5812/ircmj.24373.

Published online 2016 May 30. **Research Article**

Two-Way Gene Interaction From Microarray Data Based on Correlation Methods

Hamid Alavi Majd,[1] Atefeh Talebi,[2,*] Kambiz Gilany,[3] and Nasibeh Khayyer[4]

[1]Department of Biostatistics, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
[2]Department of Biostatistics, School of Paramedial Sciences, Students' Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
[3]Reproductive Biotechnology Research Center, Avicenna Research Institute, ACECR, Tehran, IR Iran
[4]Department of Proteomics, School of Paramedial Sciences, Students' Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran

[*]*Corresponding author*: Atefeh Talebi, Department of Biostatistics, School of Paramedial Sciences, Students' Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran. Tel: +98-2122707347, Fax: +98-2122721150, E-mail: a_talebi5855@yahoo.com

Received 2014 October 13; **Revised** 2015 March 26; **Accepted** 2015 April 21.

Abstract

Background: Gene networks have generated a massive explosion in the development of high-throughput techniques for monitoring various aspects of gene activity. Networks offer a natural way to model interactions between genes, and extracting gene network information from high-throughput genomic data is an important and difficult task.
Objectives: The purpose of this study is to construct a two-way gene network based on parametric and nonparametric correlation coefficients. The first step in constructing a Gene Co-expression Network is to score all pairs of gene vectors. The second step is to select a score threshold and connect all gene pairs whose scores exceed this value.
Materials and Methods: In the foundation-application study, we constructed two-way gene networks using nonparametric methods, such as Spearman's rank correlation coefficient and Blomqvist's measure, and compared them with Pearson's correlation coefficient. We surveyed six genes of venous thrombosis disease, made a matrix entry representing the score for the corresponding gene pair, and obtained two-way interactions using Pearson's correlation, Spearman's rank correlation, and Blomqvist's coefficient. Finally, these methods were compared with Cytoscape, based on BIND, and Gene Ontology, based on molecular function visual methods; R software version 3.2 and Bioconductor were used to perform these methods.
Results: Based on the Pearson and Spearman correlations, the results were the same and were confirmed by Cytoscape and GO visual methods; however, Blomqvist's coefficient was not confirmed by visual methods.
Conclusions: Some results of the correlation coefficients are not the same with visualization. The reason may be due to the small number of data.

Keywords: Gene Expression, Gene Regulatory Networks, Gene Ontology, Molecular Structure, Nonparametric

1. Background

In recent years, there has been a great explosion in the development of high-throughput techniques for globally monitoring various aspects of gene activity (1). High-throughput genomic data is a rich resource to explain how genes are joined (2-5). Until now, the study of the properties, activities, and roles of genes and proteins; the discovery of molecular processes within cells; and the tissues and molecular biological aspects of illnesses were assessed at one or several genes or proteins. Microarray technology has emerged as one way of simultaneously expressing the levels of thousands of genes, with the general approaches for the data being gene sets and cluster analyses (1). Likewise, several tools have been developed for the visualization and analysis of biological networks, such as Cytoscape (5), VisAnt (6), and tYNA (7). Clustering is the classification of a heterogeneous population into a number of homogeneous subsets, which are then referred to as clusters. This method attempts to find groups that are significantly different from each other, as members of these groups are extremely similar (8). Gene set enrichment analysis (GSE) is designed to find differences in gene expression between phenotypes by incorporating uses, biological knowledge, and statistical analysis (9). Clustering techniques cannot recognize molecular networks, nor can clustering methods show direct or indirect connections between the genes inside the clusters. Furthermore, clustering methods assign a gene to one cluster, while the tumor protein p53 can cooperate in several physiological pathways. Thus, we need to represent gene interaction methods based on different algorithms (8). Information about interactions improves our understanding of the disease and could provide a basis for new treatment methods (10, 11). There are several gene network constructions, such as Boolean network (12, 13), mutual information, and Bayesian network (14), to discover the more complex interactions and to detect in-

teraction networks within the gene expression data. One disadvantage of these methods is the large samples of expression data. Networks offer a natural way to model interactions between genes, with nodes representing genes and with edges representing various interactions inferred from different data sources (15).

2. Objectives

This study's purpose is to construct gene networks using nonparametric Spearman's correlation and Blomqvist's coefficient and then to compare them with Pearson's correlation. The first step in creating a gene co-expression network (GCN) is to score all pairs of gene vectors. The second step is to select a score threshold and connect all gene pairs whose scores exceed this value. Finally, the results were compared with Cytoscape, based on BIND, and GO visualization, based on molecular function methods.

3. Materials and Methods

Venous thrombosis is defined as a blood clot that forms in a vein, and it is a common reason for morbidity and mortality. A classical venous thrombosis is deep vein thrombosis (DVT), which can break off and cause a life-threatening pulmonary embolism (16). The VTE microarray dataset includes 70 adults with one or more prior VTE on warfarin and 63 healthy controls (17). Blood was gathered in PAX gene tubes, RNA was separated, and gene expression profiles were achieved using the Affymetrix human genome U133A 2.0 array. In the data, a set of six genes, such as CYP2A6, NAT2, CYP1A2, CYP2A13, XDH, and NAT1, was selected. The KEGG pathway was used for performing gene set analysis (18). In the foundation-application study, we used correlation algorithms to construct gene networks. The first stage in constructing a GCN is to score all pairs of gene vectors using correlation coefficients. The second stage is to select a score threshold and to connect all gene pairs whose scores exceed this value, focusing on undirected networks, which indicate pairwise relationships of co-expression without necessarily representing causality. There are several methods to survey the expression profiles of gene pairs.

The Pearson's correlation coefficient, r, is a measure of the degree of linear relationship between two gene vectors, X and Y, and it is calculated as (19):

$$r(X,Y) = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \quad (1)$$

The Spearman's rank correlation is like the Pearson's correlation coefficient except that it acts on the ranks of the data rather than the normal raw data (20). The Spearman's rank correlation coefficient, r_s, between two gene vectors, $X = (X_1, ..., X_N)$ and $Y = (Y_1, ..., Y_N)$ with the respective ranks $(R_1, ..., R_N)$ and $(S_1, ..., S_N)$, is calculated as:

$$r_s(X,Y) = 1 - \frac{\sum_{i=1}^{N}(R_i - S_i)^2}{\mathrm{N}(N^2 - 1)} \quad (2)$$

Blomqvist's coefficient is a nonparametric correlation method between two random variables. The coefficient is asymmetric and focuses on the difference of observed values among the first ranks in the orderings induced by the variables. Let $(x_1, y_1),....,(x_n,y_n)$ denote a sample from a continuous bivariate population, and let $\tilde{x}$, $\tilde{y}$ denote sample medians. It is separated into the (x, y)-plane by four quadrants with the lines $x = \tilde{x}$; and $y = \tilde{y}$. Then Blomqvist's B is defined as (21, 22):

$$B = \frac{n_1 + n_2}{n_1 - n_2} = \frac{2n_1}{n_1 + n_2} - 1 - 1 \leq B \leq 1 \quad (3)$$

The next step is to choose a score threshold and to create a GCN linking all gene pairs with scores exceeding this threshold. Let $Z_1, Z_2, ..., Z_P$ be p genes for the pair (Z_i, Z_j), $i \neq j$, i, j = 1, 2, ..., p the P value associated to the index $K \in [r, r_s, B]$ and to each fixed pair correlation (Z^*_a, Z^*_b) is defined by

$$P\,\mathrm{Value}(Z_i, Z_j) = \frac{\sum_{a=1}^{p}\sum_{b=a+1}^{p} \mathrm{I}[k(Z_a^*, Z_b^*) \leq k(Z_i, Z_j)]}{C(p,2)} \quad (4)$$

where I(A) denotes the indicator function of the set A. The P value was created for all pair genes in Pearson's correlation, Spearman's rank correlation, and Blomqvist's coefficient method. If P value is more than 0.95, it means that the genes are linked together to construct a network.

Cytoscape is an open source software platform to imagine interaction gene networks and to combine these interactions with gene expression and functional genomics data. Cytoscape is constructed of a gene network graph, with genes displayed as nodes and with interactions between nodes displayed as edges. The Cytoscape program is written in Java and has been released under an LGPL Open Source license; graph structures and some layout algorithms (hierarchical and circular) are implemented using the yFiles Graph Library (23).

The GO ontology is structured as a directed acyclic graph, and each term has defined relationships to one or more other terms in the same domain, and sometimes to other domains. GO is improved based on a cooperative

project and includes three structured control words (Ontologies) that depict gene products in terms of their cellular parts, biological procedures, and molecular functions in a species-independent way (24).

4. Results

In the study, the VTE dataset is a microarray dataset including 70 adults with one or more prior VTE on warfarin and 63 healthy controls. The descriptive statistics of variables are given in Table 1.

The scatter plots drawn in Figure 1 show the positive or negative linear relationships between two genes. A Kolmorov-Smirnov test was also conducted to examine the normality distribution of the genes. The results showed that all of them were insignificant ($P > 0.05$).

The dataset's Pearson's correlation, Spearman's rank, Blomqvist's coefficient, and calculated P value are presented in Table 2. We compared these relationships with GO and Cytoscape.

As shown, XDH-CYP2A6 has a strong relationship with Pearson and Spearman correlations; NAT2-CYP1A2 has a relationship with Blomqvist's coefficient; and the other pair genes have weak relationships at the 0.05 significance level. The Cytoscape visualization method, based on BIND, has been drawn for the genes in Figure 2.

As shown, the XDH gene has a relationship with CYP1A2, and this gene is related to CYP2A6 through OXY. Pearson and Spearman correlations confirmed the relationship; however, Blomqvist's beta does not show the relation.

Table 3 shows the GO method, based on molecular function, for six genes. The XDH gene has a relationship with CYP2A6, which confirms our algorithm. There are other relations in the GO method, as well.

The results showed that Pearson and Spearman correlation coefficients revealed better conclusions than the Blomqvist's coefficient. The reason may be due to the small number of data.

5. Discussion

In 2006, Kim et al. presented a new distance measure that is applied for both linear trends and fold-changes of expression in a mouse (25). They compared performances of different distance measures on seven experiments that consisted of 288 mouse oligonucleotide microarrays. They showed that the proposed distance measurement for comparing expression profiles recognizes genes with several numbers of common regulatory components since it considered the inherent regulatory knowledge better than previous distance measures. In the present study, we surveyed three correlation coefficients and two visualization methods that confirmed the relations. Although Blomqvist's coefficient does not have similar results, the findings of the Pearson and Spearman correlation coefficients are the same.

In 2002, Kue et al. surveyed mRNA measurement comparisons between matched measurements and calculated concordance between clusters from two DNA microarray technologies, Stanford type cDNA microarrays and Affymetrix oligonucleotide microarrays (26). They compared Pearson correlation and Spearman's rank correlation coefficient for genes, cell lines, and across all 162, 120 matched pairs of measurements. They hypothesized that the data had normal distribution and used Student's t-distribution. Hierarchical clustering was done using Euclidean distance, as the measure of similarity, and average linkage clustering using Matlab software. There were poor correlations between the two platforms. In the study, we presented several methods of correlation coefficients using R-3.1.1 and a visualization method using the Cytoscape and GO methods. Pearson and Spearman correlation coefficients showed the same results.

In 2000, Butte and Kohane used three methods not categories to cluster RNA expression data (27). The simple criterion for clusters was based on a fold-difference greater than a given threshold. They applied the Euclidean method for connecting all genes computing the extensive pair-wise mutual information, removed the connections under the threshold, isolated clusters of genes or related networks, and then detected related clusters biologically. Each gene was thus completely connected to every other gene with the calculated mutual information. In our study, we displayed the relationship between genes by correlation coefficients and compared them with visualization methods. Using Pearson and Spearman correlation coefficients, the results were the same.

In 2012, Bergen et al. expressed that the metabolic enzyme included in nicotine and cotinine metabolism is CYP2A6 (28). Other variables in the study were age, gender, BMI, smoking situation, and hormonal status. They carried out a hierarchical linear model for DMET SNPs and adjusted NMR, and then continued by adjusting for related tests (PACT) within genes with > 1 common SNP with $\geq$ 1 SNP with nominal $P < 0.05$. They recognized SNPs at 13 genes with PACT < 0.05 in ≥ 2 transmission models in a large twin dataset. In their article, they investigated the importance of CYP2A6 in tobacco smoking. However, we considered CYP2A6 and five other genes in order to draw a gene network based on the correlation method and on comparing them with Cytoscape and GO in venous thromboembolism. By comparing the two studies, it was concluded that CYP2A6 and five other genes are effective in venous

Table 1. Descriptive Statistics for Six Genes Related to VTE Dataset

Name of Genes	Minimum - Maximum	Mean ± SD	Median
CYP2A6	5.75 - 7.11	6.36 ± 0.28	6.33
NAT2	3.72 - 4.89	4.17 ± 0.19	4.16
CYP1A2	5.41 - 7.004	6.04 ± 0.301	6.037
CYP2A13	5.79 - 7.28	6.32 ± 0.26	6.28
XDH	4.95 - 6.71	5.77 ± 0.342	5.76
NAT1	5.03 - 8.82	6.404 ± 0.83	6.24

Table 2. Spearman's Rank Coefficient, Pearson's Correlation, Blomqvist's Coefficient, and Associated P Value for 15 Pairs of Genes CYP2A6, NAT2, CYP1A2, CYP2A13, XDH, and NAT1

Number	Gene - Gene	Spearman Corr.		Pearson Corr.		Blomqvist	
		Value	P Value	Value	P Value	Value	P Value
1	CYP2A6 - NAT2	0.573	0.73	0.594	0.73	1	1
2	CYP2A6 - CYP1A2	0.562	0.67	0.555	0.67	1	0.93
3	CYP2A6 - CYP2A13	0.643	0.8	0.678	0.93	0.33	0.4
4	CYP2A6 - XDH	0.765	1	0.785	1	0.5	0.6
5	CYP2A6 - NAT1	-0.648	0.87	-0.629	0.8	0.2	0.33
6	NAT2 - CYP1A2	0.357	0.13	0.367	0.13	1	0.86
7	NAT2 - CYP2A13	0.43	0.27	0.393	0.2	0	0.066
8	NAT2 - XDH	0.528	0.6	0.524	0.6	0.33	0.46
9	NAT2 - NAT1	-0.401	0.2	-0.409	0.27	0	0.133
10	CYP1A2 - CYP2A13	0.337	0.07	0.332	0.07	-1	0.8
11	CYP1A2 - XDH	0.515	0.53	0.513	0.4	0	0.2
12	CYP1A2 - NAT1	-0.435	0.33	-0.419	0.33	-0.33	0.53
13	CYP2A13 - XDH	0.468	0.4	0.523	0.53	-1	0.73
14	CYP2A13 - NAT1	-0.511	0.47	-0.51	0.4	0	0.27
15	XDH - NAT1	-0.658	0.93	-0.631	0.87	-1	0.66

thromboembolism disease, and they inferred that CYP2A6 is the predominant metabolic enzyme involved in nicotine and cotinine metabolism.

In 2012, Neal et al. surveyed the Cytochrome p450 (CYP) family of 60 genes in the metabolism and combination of different chemicals and lipid cellular molecules involving vitamin D (29). In genotyped NHANES III participators, they researched genetic deviation in CYP (33 SNPs in 9 genes), vitamin D receptor genes (2 SNPs), and additional variables connected to sufficiency in previous studies, such as body mass index (BMI), season of sample collection (SSC), sex, supplementation habit, income, and age for associations with vitamin D sufficiency. They applied chi square tests and multiple logistic regression to determine relations with Vitamin D sufficiency. There were important relationships between vitamin D sufficiency and SSC, BMI, sex, and age across RE level. Several CYP SNPs were associated with vitamin D sufficiency in general models. CYP2A6 (rs1801272) was meaningfully related to vitamin D sufficiency in several groups in adjusted and crude models. The article is the first report of CYP2A6's connection with vitamin D sufficiency, and there is also biological plausibility because of its wide range of potential metabolic targets. In their article, they surveyed the relation between a gene and vitamin D. Our study surveyed relationships between CYP2A6 and five other genes in venous thromboembolism, and it drew a gene network based on correlation methods in the data and on comparisons with Cytoscape and GO. Comparing the two studies, we concluded that CYP2A6 can cause vitamin D deficiency and skeletal, cardiovascular, autoimmune, and metabolic disease, as well as venous thromboembolism. Garcia-Closas et al. surveyed the

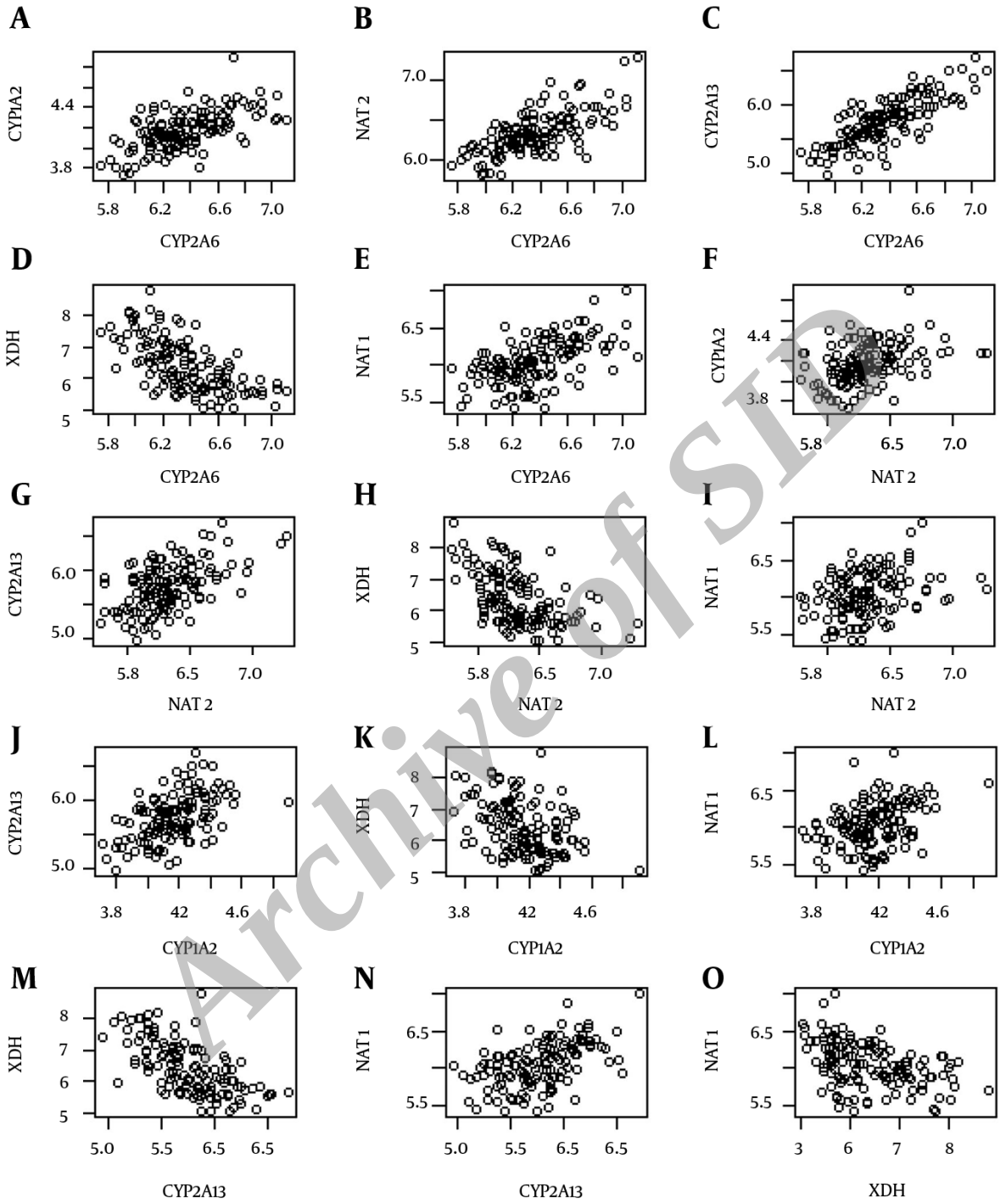

Figure 1. Scatter Plots for Pairs of Genes Related to VTE Dataset

association between NAT2 slow acetylation and GSTM1null genotype in the risk of bladder cancer (30). They studied polymorphisms in NAT2, GSTM1, NAT1, GSTT1, GSTM3, and GSTP1, and there were 1,150 patients with transitional

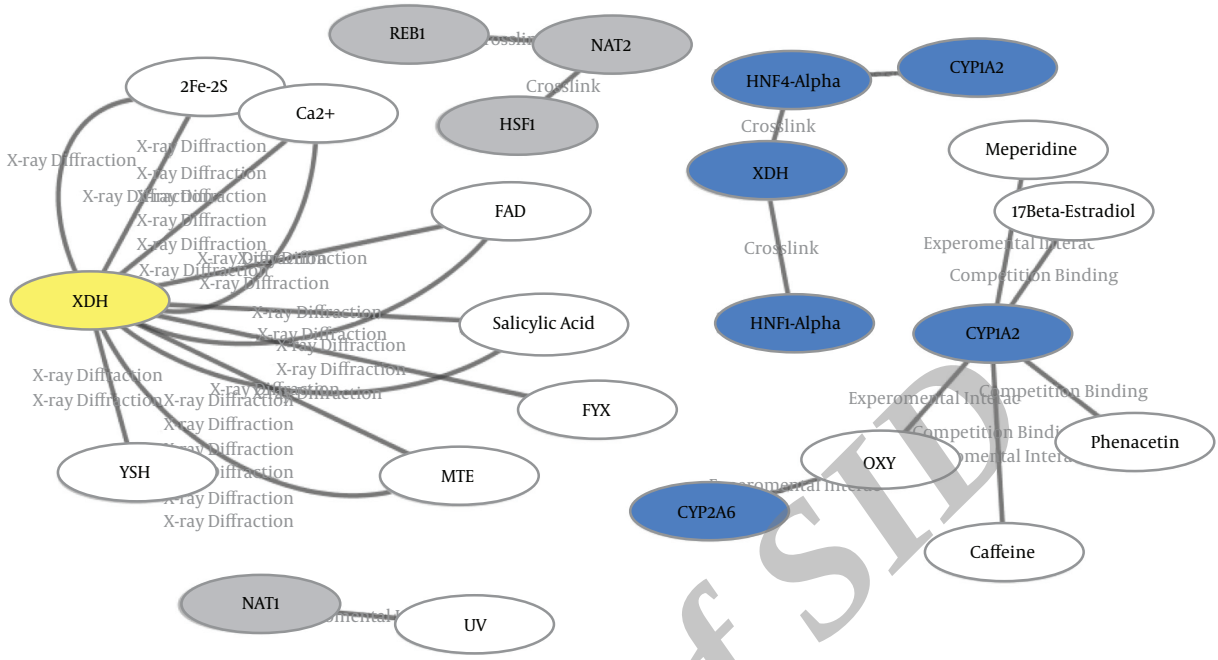

Figure 2. Cytoscape Visualization Method, Based on BIND, for Six Genes

cell carcinoma of the urinary bladder and 1,149 members of the control group in Spain. They also performed meta-analyses of GSTM1, NAT2, and bladder cancer that involved more than twice that of other studies. In bladder cancer, they compared the odds ratios for persons with an absence of one or two copies of the GSTM1gene with NAT2 rapid or intermediate acetylators. NAT2 slow acetylators had an increased overall risk of bladder cancer that was stronger in cigarette smokers than in nonsmokers. They concluded that the GSTM1null genotype increases the risk of bladder cancer, and the NAT2 slow acetylator genotype enhances the risk among cigarette smokers. In the current study, we investigated NAT2 and five other genes in venous thromboembolism. We also drew a gene network with a correlation-based algorithm and compared it with Cytoscape and GO visualization methods. Comparing the two studies, we concluded that NAT2 can cause bladder cancer and venous thromboembolism.

In 2000, Bartsch et al. studied several genes, such as CYP1A1, 1A2, 1B1, 2A6, 2D6, 2E1, 2C9, 2C19, 17, and 19, singularly or as a mixture with detoxifying enzymes as adjusters for the risk for tobacco-interconnected cancers (31). They expressed the important actions by which the compounds are metabolized and caused DNA adducts in the bladder epithelium, including N-hydroxylation (CYP1A2) and N-acetylation (NAT1 and NAT2). These aromatic amines are the main components of smoke and seem to be an important reason for urinary bladder cancer in smokers. They also stated that deleting the CYP2A6 region leads to an inactive enzyme or lack of protein synthesis, differences in the polyadenylation signal of NAT1 that affects transcript half-life, the quantity of the enzyme, and interactions of the CYP1A2 gene and its enzyme catalysis products. In this study, we surveyed NAT1, CYP2A6, CYP2A2, and three other genes in venous thromboembolism. We also constructed a gene network using a correlation-based algorithm and compared it with GO and Cytoscape methods. Comparing the two studies, it was concluded that NAT2, CYP2A6, and CYP2A2 are effective in tobacco-related cancers and venous thromboembolism.

Acknowledgments

We would like to express our sincere thanks to the referees for carefully reading our manuscript and for giving such constructive comments, which substantially helped in improving the paper's quality.

Footnotes

Authors' Contribution: Hamid Alavi Majd edited the study; Atefeh Talebi wrote and ran the programs; Nasibeh Khayyer and Kambiz Gilany ran the gene software.

Table 3. GO Method Base on Molecular Function

Gene Name	GO Term
CYP2A6	
Oxidoreductase activity	GO:0016705
Iron ion binding	GO:0005506
Heme binding	GO:0020037
NAT2	
Arylamine N-acetyltransferase activity	GO:0004060
CYP1A2	
Enzyme binding	GO:0019899
Aromatase activity	GO:0070330
Oxidoreductase activity	GO:0016712
Demethylase activity	GO:0032451
Electron carrier activity	GO:0009055
Oxidoreductase activity	GO:0016491
Iron ion binding	GO:0005506
Monooxygenase activity	GO:0004497
Heme binding	GO:0020037
Caffeine oxidase activity	GO:0034875
CYP2A13	
Iron ion binding	GO:0005506
Heme binding	GO:0020037
Aromatase activity	GO:0070330
XDH	
UDP-N-acetylmuramate dehydrogenase activity	GO:0008762
2 iron, 2 sulfur cluster binding	GO:0051537
Molybdopterin cofactor binding	GO:0043546
Electron carrier activity	GO:0009055
Xanthine oxidase activity	GO:0004855
Iron ion binding	GO:0005506
Protein homodimerization activity	GO:0042803
Flavin adenine dinucleotide binding	GO:0050660
NAT1	
Arylamine N-acetyltransferase activity	GO:0004060

Funding/Support: Shahid Beheshti University of Medical Sciences.

References

1. Dehmer M, Emmert-Streib F, Graber A, Salvador A. Applied statistics for network biology: methods in systems biology. John Wiley and Sons; 2011. p. 478.
2. Lander ES. Array of hope. *Nat Genet.* 1999;**21**(1 Suppl):3–4. doi: 10.1038/4427. [PubMed: 9915492].
3. Quackenbush J. Genomics. Microarrays–guilt by association. *Science.* 2003;**302**(5643):240–1. doi: 10.1126/science.1090887. [PubMed: 14551426].
4. Zhang MQ. Extracting functional information from microarrays: a challenge for functional genomics. *Proc Natl Acad Sci U S A.* 2002;**99**(20):12509–11. doi: 10.1073/pnas.212532499. [PubMed: 12271149].
5. Killcoyne S, Carter GW, Smith J, Boyle J. Cytoscape: a community-based framework for network modeling. *Methods Mol Biol.*

2009;**563**:219–39. doi: 10.1007/978-1-60761-175-2_12. [PubMed: 19597788].
6. Hu Z, Snitkin ES, DeLisi C. VisANT: an integrative framework for networks in systems biology. *Brief Bioinform.* 2008;**9**(4):317–25. doi: 10.1093/bib/bbn020. [PubMed: 18463131].
7. Yip KY, Yu H, Kim PM, Schultz M, Gerstein M. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. *Bioinformatics.* 2006;**22**(23):2968–70. doi: 10.1093/bioinformatics/btl488. [PubMed: 17021160].
8. Wu X, Ye Y, Subramanian KR, Zhang L. Interactive Analysis of Gene Interactions Using Graphical gaussian model. *Biol Knowl Discov Data Min.* 2003;**3**:63–9.
9. Lists of Software for Bioinformatics: Pathway Analysis Tool Available from: http://bioinformatics.ai.sri.com/ptools/.
10. Mendes P, editor. Advanced visualization of metabolic pathways in PathDB. Proceedings of the 8th Conference on Plant and Animal Genome. 2000; San Diego. .
11. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. *Microbiol Rev.* 1995;**59**(1):94–123. [PubMed: 7708014].
12. Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput.* 1998:18–29. [PubMed: 9697168].
13. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics.* 2002;**18**(2):261–74. [PubMed: 11847074].
14. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol.* 2000;**7**(3-4):601–20. doi: 10.1089/106652700750050961. [PubMed: 11108481].
15. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;**5**(2):101–13. doi: 10.1038/nrg1272. [PubMed: 14735121].
16. Saha P, Humphries J, Modarai B, Mattock K, Waltham M, Evans CE, et al. Leukocytes and the natural history of deep vein thrombosis: current concepts and future directions. *Arterioscler Thromb Vasc Biol.* 2011;**31**(3):506–12. doi: 10.1161/ATVBAHA.110.213405. [PubMed: 21325673].
17. Lewis DA, Stashenko GJ, Akay OM, Price LI, Owzar K, Ginsburg GS, et al. Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res.* 2011;**128**(6):536–40. doi: 10.1016/j.thromres.2011.06.003. [PubMed: 21737128].
18. Alavi-Majd H, Khodakarim S, Zayeri F, Rezaei-Tavirani M, Tabatabaei SM, Heydarpour-Meymeh M. Assessment of gene set analysis methods based on microarray data. *Gene.* 2014;**534**(2):383–9. doi: 10.1016/j.gene.2013.08.063. [PubMed: 24012817].
19. Taylor R. Interpretation of the Correlation Coefficient: A Basic Review. *J Diagn Med Sonogr.* 1990;**6**(1):35–9. doi: 10.1177/875647939000600106.
20. Gauthier T. Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environ Forensics.* 2001;**2**(4):359–62. doi: 10.1006/enfo.2001.0061.
21. Blomqvist N. On a Measure of Dependence Between two Random Variables. *Ann Math Stat.* 1950;**21**(4):593–600. doi: 10.1214/aoms/1177729754.
22. Genest C, Plante JF. On blest's measure of rank correlation. *Can J Stat.* 2003;**31**(1):35–52. doi: 10.2307/3315902.
23. yWorks - The Diagramming Company . Files Graph Library Available from: www.yworks.
24. Slimani T. Description and Evaluation of Semantic Similarity Measures Approaches. *Int J Comput Appl.* 2013;**80**(10):25–33. doi: 10.5120/13897-1851.
25. Kim RS, Ji H, Wong WH. An improved distance measure between the expression profiles linking co-expression and co-regulation in mouse. *BMC Bioinformatics.* 2006;**7**:44. doi: 10.1186/1471-2105-7-44. [PubMed: 16438730].
26. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics.* 2002;**18**(3):405–12. [PubMed: 11934739].
27. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput.* 2000:418–29. [PubMed: 10902190].
28. Bergen A, Javitz H, Michel M, Krasnow R, Nishita D, Lessov-Schlaggar C, et al, editors. Drug Metabolizing Enzyme Genes and Nicotine and Cotinine Metabolism. American Society of Human Genetics 62nd Annual Meeting. 2012; San Francisco, California. .
29. Neal C, Jackson J, Crider K, editors. Genetic variation and vitamin D sufficiency in the U.S. population (NHANES III). American Society of Human Genetics 62nd Annual Meeting. 2012; San Francisco, California. .
30. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet.* 2005;**366**(9486):649–59. doi: 10.1016/S0140-6736(05)67137-1. [PubMed: 16112301].
31. Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol Biomarkers Prev.* 2000;**9**(1):3–28. [PubMed: 10667460].