

# Identifying key genes in glaucoma based on a benchmarked dataset and the gene regulatory network

XI CHEN<sup>1</sup>, QIAO-LING WANG<sup>2</sup> and MENG-HUI ZHANG<sup>3</sup>

<sup>1</sup>Department of Ophthalmology, The Ninth Hospital of Chongqing, Chongqing 400700; <sup>2</sup>Department of Ophthalmology, The Second Hospital of Jinan, Jinan, Shandong 250022; <sup>3</sup>Department of General Surgery, The Fourth Hospital of Jinan, Jinan, Shandong 250031, P.R. China

Received August 3, 2016; Accepted April 28, 2017

DOI: 10.3892/etm.2017.4931

**Abstract.** The current study aimed to identify key genes in glaucoma based on a benchmarked dataset and gene regulatory network (GRN). Local and global noise was added to the gene expression dataset to produce a benchmarked dataset. Differentially-expressed genes (DEGs) between patients with glaucoma and normal controls were identified utilizing the Linear Models for Microarray Data (Limma) package based on benchmarked dataset. A total of 5 GRN inference methods, including Zscore, GeneNet, context likelihood of relatedness (CLR) algorithm, Partial Correlation coefficient with Information Theory (PCIT) and GENE Network Inference with Ensemble of Trees (Genie3) were evaluated using receiver operating characteristic (ROC) and precision and recall (PR) curves. The interference method with the best performance was selected to construct the GRN. Subsequently, topological centrality (degree, closeness and betweenness) was conducted to identify key genes in the GRN of glaucoma. Finally, the key genes were validated by performing reverse transcription-quantitative polymerase chain reaction (RT-qPCR). A total of 176 DEGs were detected from the benchmarked dataset. The ROC and PR curves of the 5 methods were analyzed and it was determined that Genie3 had a clear advantage over the other methods; thus, Genie3 was used to construct the GRN. Following topological centrality analysis, 14 key genes for glaucoma were identified, including *IL6*, *EPHA2* and *GSTT1* and 5 of these 14 key genes were validated by RT-qPCR. Therefore, the current study identified 14 key genes in glaucoma, which may be potential biomarkers to use in the diagnosis of glaucoma and aid in identifying the molecular mechanism of this disease.

## Introduction

Glaucoma describes a group of eye diseases that damage the optic nerve and cause vision loss (1). Glaucoma affects >70 million people worldwide and ~10% of patients develop bilateral blindness (2); thus, it is the second leading cause of blindness in the world (3). Glaucoma is usually asymptomatic until it becomes severe, meaning that the number of affected individuals may be much higher than the number formally diagnosed with glaucoma (4). Population-level surveys suggest that only 10-50% of people with glaucoma are aware that they have the condition (4). Therefore, it is important to develop appropriate techniques to diagnose glaucoma early and enable prompt treatment of patients with glaucoma.

At present, the majority of studies aim to develop treatments for late stage glaucoma, including surgery and pharmaceuticals, and do not investi-

gate methods of early detection and prevention (5,6). With the development of large-scale gene expression studies, gene-level therapies have become more powerful and informative, facilitating investigations into glaucoma's mechanism of action (7). Additionally, gene expression studies can be used to detect the target genes of glaucoma, which may provide an effective method of detecting glaucoma early (1). A mutation of the cytochrome P450 family 1 subfamily B member 1 has been identified in the majority of glaucoma cases (8). Furthermore, sequence variants in the ankyrin repeat and SOCS box containing 10 gene may be a risk factor for glaucoma (9). However, few target genes and biomarkers of glaucoma have been identified to date.

Therefore, the objective of the current study was to determine the key genes in glaucoma. Firstly, a dataset of glaucoma (E-GEOD-9963) was recruited from the ArrayExpress database, and the dataset was benchmarked by adding local and global noise. Then, differentially-expressed genes (DEGs) were determined from the benchmarked dataset. Subsequently, 5 different types of gene regulatory network (GRN) inference approaches comprising Zscore, GeneNet, context likelihood of relatedness (CLR) algorithm, Partial Correlation coefficient with Information Theory (PCIT) and GENE Network Inference with Ensemble of trees (Genie3) were evaluated by Receiver Operating Characteristic (ROC) and Precision and Recall (PR) curves, and the GRN on the basis of DEGs

---

*Correspondence to:* Dr Xi Chen, Department of Ophthalmology, The Ninth Hospital of Chongqing, 69 Jialing Road, Beibei, Chongqing 400700, P.R. China  
E-mail: xichenmedi@yeah.net

**Key words:** glaucoma, benchmarked dataset, gene regulatory network, key genes, reverse transcription-polymerase chain reaction

was constructed using the method with the best performance. Topological centrality (degree, closeness and betweenness) was then conducted to identify key genes in the GRN of glaucoma. Finally, the key genes were validated by performing reverse transcription-quantitative polymerase chain reaction (RT-qPCR). These key genes may be potential biomarkers and provide methods of improving the diagnosis and treatment of glaucoma.

## Materials and methods

**Benchmarked dataset.** In the current study, a gene expression dataset involving noise-free experiments for patients with glaucoma was taken from the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/>). This dataset had the accession number E-GEOD-9963. E-GEOD-9963 was then deposited on the Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133\_Plus\_2] Platform, and consisted of 6 glaucoma samples and 54 control samples. The dataset on the probe level was converted into gene symbols based on the annotate package (10) and duplicated ones were removed using the feature filter method (11).

To eliminate the bench effects of the real microarray generation process within the same laboratory or among different ones, the dataset was contaminated with a mixture of Gaussian noise and lognormal noise (12). Gaussian noise is defined as local noise with a zero mean (13). Standard deviation of the 'local' noise,  $\sigma_{Local(g)}$ , was around a percentage ( $\theta\%$ ) of the gene standard deviation ( $\sigma_g$ ) and was calculated as follows:

$$\sigma_{Local(g)} = \sigma_g \frac{\Gamma(0.8\theta, 1.2\theta)}{100}$$

Here,  $\Gamma(a, b)$  denoted the uniform distribution between  $a$  and  $b$ . Subsequently, an independent lognormal noise termed 'global' noise, was added (14). The standard deviation of this noise ( $\sigma_{Global}$ ) was the same for the whole dataset and was a percentage ( $\theta_g\%$ ) of the mean variance of all the genes in the dataset ( $\sigma_g$ ). It was calculated as follows:

$$\sigma_{Global; \theta_g\%} = \sigma_g \frac{\Gamma(0.8\theta_g, 1.2\theta_g)}{100}$$

A range of 20% around  $\theta$  and  $\theta_g$  was selected to add variability to the different generated datasets (15). This allowed the various datasets to have some heterogeneity in noise but at the same time ensured that they did not differ too much from the originally specified values. Ultimately, a benchmarked dataset of 12,442 genes was obtained for further investigation.

**DEGs.** It has been determined that the propensity of numerous diseases can be identified by the differences in their gene expression (16). Therefore, DEGs between glaucoma and healthy controls were detected from the benchmarked dataset using on the Linear Models for Microarray Data (Limma) package (Bioconductor; <http://www.bioconductor.org/>) (17). The lmFit function implemented in Limma was utilized to perform empirical Bayes statistics and false discovery rate calibration of the p-values on the data (18,19). Only genes which met to the thresholds of  $P < 0.05$  and  $\log_2 \text{FoldChange} > 2$  were regarded as DEGs between patients with glaucoma and healthy controls.

**GRN inference methods.** A total of 5 common GRN inference approaches (CLR, Zscore, Genie3, GeneNet and PCIT) were compared to construct a reliable and stable GRN based on the DEGs of glaucoma.

**CLR.** The CLR algorithm, an information-theoretic approach, makes use of a generalization of the pairwise correlation coefficient called mutual information ( $M_{ij}$ ) for each pair of genes (20). It derived a score ( $S_{ij}$ ) between gene  $i$  and gene  $j$  associated with the empirical distribution of the  $M_{ij}$  values (21).

$$S_{ij} = \sqrt{S_i^2 + S_j^2}$$

Of which:

$$S_j = \max\left(0, \frac{M_{ij} - \mu M_j}{\sigma M_j}\right)$$

$\mu M_i$  and  $\sigma M_i$  represented the sample mean and standard deviation of the empirical distribution of  $M_{ij}$  respectively.  $S_j$  was calculated using the following formula:

$$S_i = \max\left(0, \frac{M_{ij} - \mu M_i}{\sigma M_i}\right)$$

Where  $\mu M_j$  and  $\sigma M_j$  represented the sample mean and standard deviation of the empirical distribution of  $M_{ij}$  respectively.

**Zscore.** Zscore executes knockout experiments more concretely and results in changes in other genes (22). The knocked-out gene  $i$  in the experiment  $k$  may more strongly affect the genes that it regulates than any others. The effect of gene  $i$  over gene  $j$  was determined using the Zscore,  $Z_{ij}$ :

$$Z_{ij} = \left| \frac{x_{jk} - \mu X_j}{\sigma X_j} \right|$$

$X_j$  represented the expression of the  $j$ th gene in every experiment,  $x_{jk}$  denoted the particular gene expression level of the  $k$ th experiment of the  $j$ th gene,  $\mu X_j$  and  $\sigma X_j$  represented the mean and standard deviation of the empirical distribution for the gene  $j$ , respectively. If the same gene was detected to be knocked-out in various experiments, then the final Zscore was the mean of the individual Zscore values.

**Genie3.** The Genie3 algorithm implements the random forest feature selection technique to solve a regression problem for each of the genes included in the network (23,24).  $X_k^{-i}$  describes the vector containing the expression values in the  $k$ th experiment of  $n$  genes apart from gene  $i$ .

$$X_k^{-i} = (x_k^1, \dots, x_k^{i-1}, x_k^{i+1}, \dots, x_k^n)^T$$

Therefore,

$$x_k^i = f_i(X_k^{-i}) + \varepsilon_k$$

$\varepsilon_k$  represented random noise with zero mean. The function  $f_i$  only exploited the expression in  $X_i$  of the genes that were direct regulators of gene  $i$ , i.e. genes that were directly connected to gene  $i$  in the targeted network.

**GeneNet.** GeneNet is a simple heuristic for the statistical learning of a high dimensional causal network (25). A

correlation network was converted into a partial correlation that remained between two variables if the effect of the other variables had been regressed away. Assuming that  $A_m$  and  $B$  are random variables with known variances  $\text{var}(B)$  and  $\text{var}(A_m)$  and with covariance  $\text{cov}(B, A_m)$ , the best linear predictor of  $B$  in terms of the  $A_m$  could be calculated using the following formula:

$$B_m^k = C_{km} \sqrt{\frac{\tau_k^2}{\tau_m^2}}$$

$C_{km}$  was the partial correlation between  $B$  and  $A_m$  in  $k$  experiments;  $\tau_k^2$  and  $\tau_m^2$  were the respective partial variances. Subsequently, partial ordering of the nodes was established by multiple testing of the log-ratio of standardized partial variances, which facilitated the identification of a directed acyclic causal network as a sub-graph of the partial correlation network.

*PCIT.* The PCIT algorithm combines the concept of PCIT to identify significant gene-to-gene associations (26). For every trio of genes in  $x$ ,  $y$  and  $z$ , the three first-order partial correlation coefficients ( $r_{xyz}$ ,  $r_{xzy}$  and  $r_{yzx}$ ) were computed. In order to determine the tolerance level ( $T$ ) used as the local threshold for capturing significant associations, the average ratio of partial to direct correlation was defined as follows:

$$T = \frac{1}{3} \left( \frac{r_{xyz}}{r_{xy}} + \frac{r_{xzy}}{r_{xz}} + \frac{r_{yzx}}{r_{yz}} \right)$$

An association between genes  $x$  and  $y$  was disregarded if:  $|r_{xy}| \leq |Tr_{xz}|$  and  $|r_{xy}| \leq |Tr_{yz}|$ . Otherwise, the association was defined as significant and an association between the pair of genes was established during construction of the GRN.

*GRN construction.* The GRN of glaucoma was constructed using the GRN inference approach with the best performance. To evaluate the performance of the five inference methods in glaucoma, the usual metrics of machine learning, ROC and PR curves, were utilized. ROC curves present the relative frequencies of true positive fraction (TPF) to false positive fraction (FPF) for every predicted link of the edge list (27), of which TPF measures the fraction of correctly labeled positive examples and FPF denotes the partition of misclassified negative examples. The PR identifies the relative precision (fraction of correct predictions) vs. recall that is equivalent to the true positive ratio (TPR) (28). Recall is the same as TPR and precision assessments that classifies a specific fraction of examples as positive that are truly positive. The calculations for TPR, false positive ratio (FPR), recall, and precision were performed using the following formula:

$$\begin{aligned} \text{TPR or Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}} \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \end{aligned}$$

TP was defined as a true positive sample, FP stood for false positive sample, TN represented a true negative sample and FN indicated a false negative sample.

*Key genes.* To evaluate the biological importance of genes in GRN, topological centrality (degree, closeness and betweenness) analyses were conducted. Specifically, degree quantifies the local topology of each gene, by summing up the number of its adjacent genes (29). Closeness centrality is a measure of the average length of the shortest paths to all other proteins in the network (30). Betweenness considers the ratio of a node in the shortest path between two other nodes and scales with the number of pairs of nodes as implied by the summation indices (31). The different methods may lead to different results; therefore, the three types of results were integrated using the Rank Product (RP) algorithm (32) and avoided the side effects caused by single one approach, including a false discovery rate. The genes at the  $\geq 92\%$  merged quantile distribution in the significantly perturbed networks were defined as key genes.

*Reverse transcription-quantitative polymerase chain reaction (RT-qPCR).* In the present study, RT-qPCR was performed to assess the expression of key genes in patients with glaucoma compared with normal controls. For convenience, RT-qPCR was performed on 5 of the 14 key genes, including *IL6*, *EPHA2*, *GSTT*, *SPTBN1* and *ERBB2*. Total RNA was extracted from 10 patients with glaucoma and 10 healthy controls from optic nerve head astrocytes using TRIzol (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA). The the mean age was  $(63.90 \pm 7.12)$ , and the sex ration was 4 male patients: 6 female patients. Patients were recruited from the Ninth Hospital of Chongqing (Chongqing, China) between May 2015 and February 2016. All patients provided informed consent for inclusion in the current study and ethical approval was obtained from the Ninth Hospital of Chongqing Ethics Committee (Chongqing, China). Subsequently, total RNA underwent first-strand synthesis with an oligo (dT18) primer and treated with RNasin, reverse transcriptase buffer, dNTPs and AMV reverse transcriptase according to the manufacturer's instructions (Invitrogen; Thermo Fisher Scientific, Inc.). For qPCR, cDNA was used as a template and  $\beta$ -actin as an internal reference.

For PCR amplification, the reaction mixture contained PCR Buffer I, Taq DNA Polymerase High Fidelity (both from Invitrogen; Thermo Fisher Scientific, Inc.), dNTPs and forward and reverse sequences of *IL6*, *EPHA2*, *GSTT*, *SPTBN1* and *ERBB2* primers (sequences not provided). Conditions for qPCR were as follows: 2 min at 94°C for predenaturation, followed by 35 cycles of 20 sec at 94°C, 15 sec at 60°C and 1 min at 68°C, and a final 7 min extension at 72°C. Products were analyzed by 1.5% agarose gel electrophoresis and Quantity One 1-D Analysis software of gel imaging analyzer (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Three replicates of the assay were performed to assess reproducibility and the  $2^{-\Delta\Delta Cq}$  method was applied (33).

*Statistical analysis.* The results were analyzed using SPSS 19.0 software (SPSS, Inc., Chicago, IL, USA) (34). Data were expressed as the mean  $\pm$  standard deviation. Differences between groups were assessed by an unpaired, two-tailed Student's t-test (35) and  $P < 0.05$  was considered to indicate a statistically significant difference.

Table I. Evaluation of the five-gene regulatory network inference methods.

| Index | CLR    | Zscore  | Genie3 | GeneNet   | PCIT    |
|-------|--------|---------|--------|-----------|---------|
| ROC   | 0.204  | 0.0117  | 0.520  | 0.0000195 | 0.169   |
| PR    | 0.0119 | 0.00233 | 0.0192 | 0.000394  | 0.00859 |

ROC, receiver operating characteristic; PR, precision and recall; CLR, context likelihood of relatedness; PCIT, Partial Correlation coefficient with Information Theory; Genie3, GENE Network Inference with Ensemble of Trees.

## Results

**Benchmarked dataset.** In the current study, to ensure that the preprocessed dataset was more representative of the real expression data and to reproduce the variability in the real gene expression dataset generation process, the dataset was benchmarked by addition of 20% local noise ( $\sigma_{Local(g)}$ ) and 20% global noise ( $\sigma_{Global}$ ). A total of 12,442 genes were obtained in the benchmarked dataset for subsequent analysis.

**DEGs.** Using the benchmarked dataset, DEGs between patients with glaucoma and normal controls were identified using the Limma package. When setting the thresholds of  $P < 0.05$  and  $\log_2(\text{FoldChange}) > 2$ , a total of 176 DEGs were detected. *ATP6V0D1* ( $P = 2.37 \times 10^{-16}$ ), *COPG1* ( $P = 1.24 \times 10^{-15}$ ), *PKNI* ( $P = 4.55 \times 10^{-14}$ ), *SAFB* ( $P = 6.371 \times 10^{-10}$ ) and *BAD* ( $P = 4.04 \times 10^{-09}$ ) were the top five ranked DEGs.

**GRN construction.** To construct a more stable and reliable GRN for glaucoma, five GRN inference methods (CLR, Zscore, Genie3, GeneNet and PCIT) were compared using ROC and PR curves. The method with the best performance was then selected. The results for the ROC and PR curves are illustrated in Fig. 1. The cross points of any two ROC curves indicated that the TPF and FPF of the two methods were the same. Furthermore, for each method, the comparison of TPF against FPF referred to the slope of the curve; a steeper curve indicated a better method. There were differences among the curves of the five methods. For PR curves, the meaning of cross point was the same as the ROC curve. The curves of the five approaches were very similar, however GeneNet was more precise than the others at the initial point of Recall. Genie3 exhibited a slight advantage over the other four methods as it had the largest ROC (0.520) and PR (0.0192; Table I).

Therefore, the Genie3 method was utilized to construct a GRN for glaucoma on the basis of DEGs (Fig. 2). The primary features of Genie3 with respect to existing techniques is that it makes very few assumptions about the nature of associations between the variables (which may therefore be non-linear) and may potentially identify high-order conditional dependencies between expression patterns (36). All DEGs were mapped to the network and the GRN contained 176 nodes and 15,400 edges.

**Key genes.** By accessing topological centrality (degree, closeness and betweenness) analyses for genes in GRN, the rank

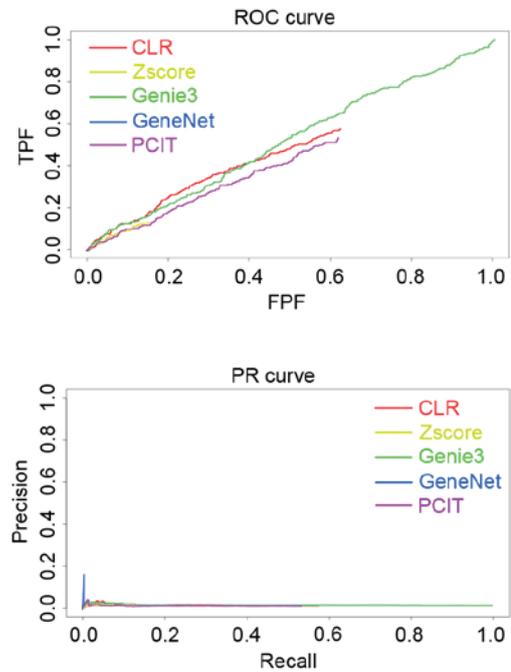


Figure 1. ROC and PR curves for evaluating the gene regulatory network inference approaches. ROC, receiver operating characteristic; PR, precision and recall; CLR, context likelihood of relatedness; PCIT, Partial Correlation coefficient with Information Theory; Genie3, GENE Network Inference with Ensemble of Trees; TPF, true positive fraction; FPF, false positive fraction.

of each gene could be obtained. Due to inconsistent results caused by different methods, the RP algorithm was applied to integrate these results. The top 8% of the ranked genes were defined as key genes for glaucoma. A total of 14 key genes were identified: *IL6*, *EPHA2*, *GSTT1*, *SPTBN1*, *ERBB2*, *TOM1*, *BAD*, *ALDH4A1*, *BSG*, *MRPL2*, *NOL3*, *SLC25A22*, *MAP3K5* and *ITGA6*. These are presented as the yellow nodes in GRN (Fig. 2). To determine interactions among the key genes, a sub-network that was closely correlated to these key genes from the GRN was extracted and termed as the 'key sub-network' (Fig. 3). There were 14 nodes and 91 interactions in this key sub-network, indicating that these key genes interacted with each other and participated in similar biological processes together.

RT-qPCR was subsequently performed to determine the expression of the key genes in patients with glaucoma compared with normal controls. *IL6*, *EPHA2*, *GSTT1*, *SPTBN1* and *ERBB2* were used as examples (Fig. 4). Compared with normal controls, the levels of *IL6*, *EPHA2*, *GSTT1* and *ERBB2* mRNA were significantly higher in samples from patients with glaucoma compared with healthy controls ( $P < 0.05$ ; Fig. 4). Additionally, the expression of *SPTBN1* was significantly lower in the glaucoma samples compared with healthy controls ( $P < 0.05$ ; Fig. 4). Thus, the results from RT-qPCR validate that the results of GRN indicating that these are key genes in glaucoma.

## Discussion

The current study benchmarked the gene dataset by adding local and global noise to eliminate batch or other effects on the dataset, and obtained a benchmarked dataset of 12,442 genes.

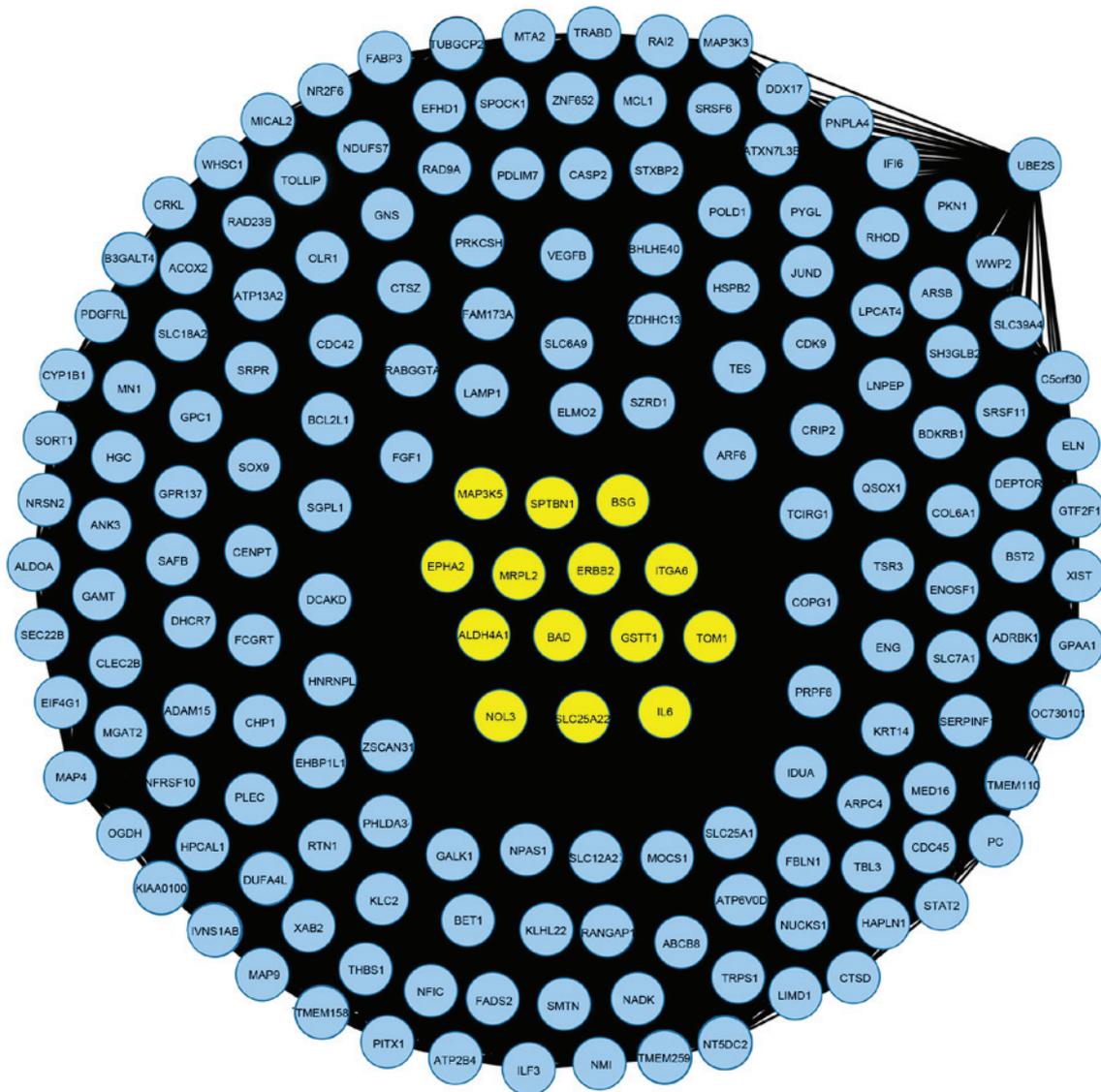


Figure 2. Gene regulatory network for glaucoma constructed using the GENE Network Inference with Ensemble of Trees. Nodes represent the genes and the edges stand for the interactions among nodes. The yellow nodes represent the 14 key genes.

Subsequently, the effectiveness of five GRN inference methods (CLR, Zscore, Genie3, GeneNet and PCIT) was analyzed on the systemic level. For the five GRN inference methods, Zscore and GeneNet used co-expression algorithms, CLR and PCIT represented information-theoretic methods and Genie3 was one of the feature selection approaches. Co-expression algorithms usually construct a network by computing a similarity score for each pair of genes (37). Information-theoretic approaches made use of a generalization of the pairwise correlation coefficient (20). Genie3 assumes that the expression of each gene in a given condition is a function of the expression of the other genes in the network (38). The results demonstrated that Genie3 exhibited the best performance following evaluation of the ROC and PR curves, suggesting that the feature selection approach was more suitable for revealing key genes in glaucoma than the other approaches.

Therefore, GRN was constructed using the Genie3 method and key genes were identified using GRN following topological centrality analyses. A total of 14 key genes were identified by integrating degree, closeness and betweenness

centrality. RT-qPCR was then used to assess the expression of 5 key genes (*IL6*, *EPHA2*, *GSTT1*, *SPTBN1* and *ERBB2*) in patients with glaucoma compared with healthy controls. It was determined that there were significant differences between the expression of all 5 genes in patients with glaucoma compared with controls.

*IL6* is an inflammatory cytokine that serves a role in the inflammatory response, which may occur following monocyte recruitment and nucleophilic apoptosis and phagocytosis in the inflamed area (39,40). It has been suggested that *IL6* signaling via classical and trans-signaling pathways may be important in determining retinal ganglion cell survival in glaucoma (41) and that cleavage of the *IL6* receptor  $\alpha$  may be a potential mechanism for *IL6* trans-signaling (42). It has been demonstrated that *IL6* increases the genetic susceptibility to primary open angle glaucoma and these genetic variants may be used as molecular markers to predict the risk of the disease (43). In addition, Micera *et al* (44) demonstrated that the expression of *IL6* differed in the primary open angle trabecular meshwork, which was consistent with the results of the current study.

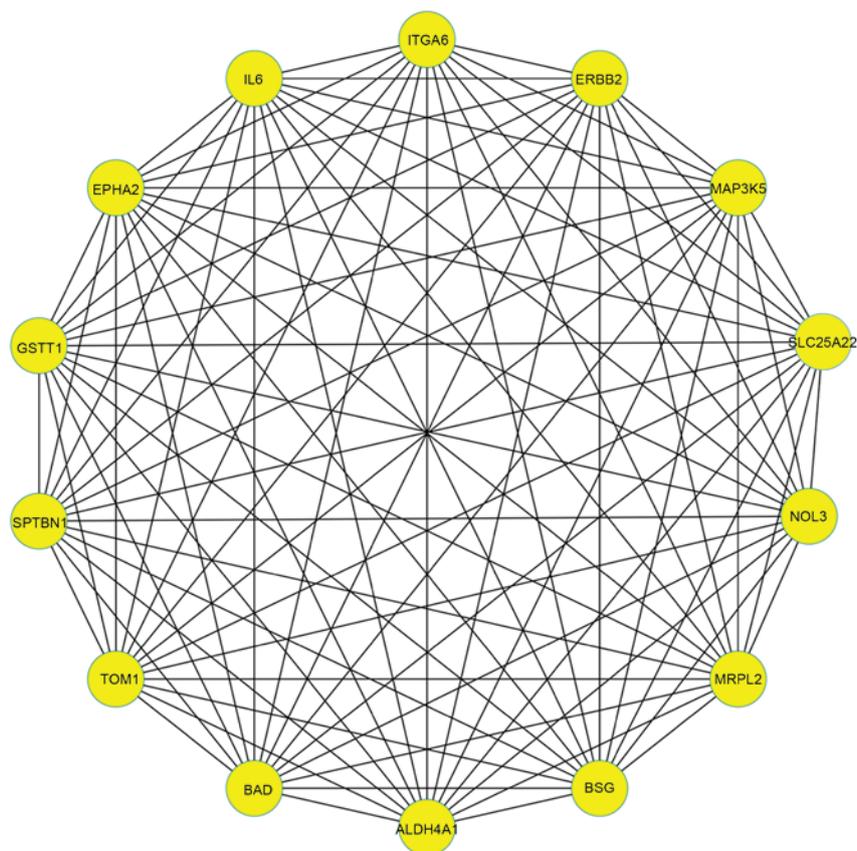


Figure 3. Key sub-network extracted from gene regulatory network of glaucoma. Nodes were key genes, and the edge indicated that two key genes had interaction.

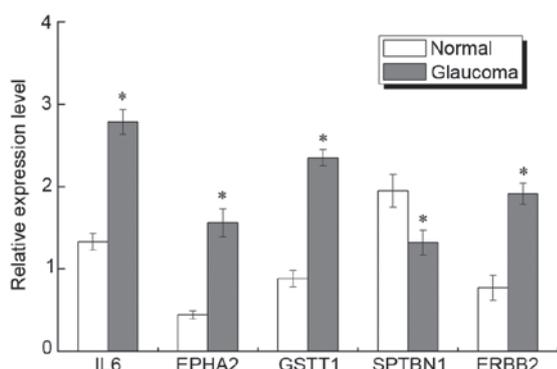


Figure 4. The relative expression of 5 key genes (*IL6*, *EPHA2*, *GSTT1*, *SPTBN1* and *ERBB2*) in patients with glaucoma compared with normal controls. \* $P < 0.05$  vs. normal control.

*EPHA2*, an Eph receptor tyrosine kinase, has been implicated in cell-cell and cell-matrix interactions, as well as cell growth and survival (45). It has been confirmed that *EPHA2* mutations affect the intracellular region of the protein, the sterile  $\alpha$  motif, tyrosine kinase domain, the juxtamembrane regions that interact with signaling genes and other important functions (45). Immunohistochemical analysis of the tumor microarray demonstrated that the expression of *EPHA2* is significantly greater in tumors compared with healthy adjacent tissues (46). The results of the RT-qPCR performed in the current study determined that *EPHA2* is more highly expressed in patients with glaucoma

compared with healthy controls. This indicates that early prevention of *EPHA2* alteration may be a novel method of decreasing the cancer incidence in patients with glaucoma. In addition, Reis *et al* (47) demonstrated that the congenital cataract phenotype is accompanied by a mutation in *EPHA2*, however two individuals in the family with the *EPHA2* mutation assessed in this study were also affected with juvenile glaucoma. Thus, the altered expression of *EPHA2* may occur in glaucoma.

In conclusion, the current study identified 14 key genes including *IL6*, *EPHA2* and *GSTT1* in glaucoma using a benchmarked dataset, the topological centrality of GRN and RT-qPCR. These genes may be potential biomarkers for the early detection, diagnosis and treatment of glaucoma, and may facilitate identification of the molecular mechanism underlying this disease.

## References

1. Casson RJ, Chidlow G, Wood JP, Crowston JG and Goldberg I: Definition of glaucoma: Clinical and experimental concepts. *Clin Exp Ophthalmol* 40: 341-349, 2012.
2. Quigley HA and Broman AT: The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 90: 262-267, 2006.
3. Weinreb RN and Khaw PT: Primary open-angle glaucoma. *Lancet* 363: 1711-1720, 2004.
4. Budenz DL, Barton K, Whiteside-de Vos J, Schiffman J, Bandi J, Nolan W, Herndon L, Kim H, Hay-Smith G and Tielsch JM; Tema Eye Survey Study Group: Prevalence of glaucoma in an urban West African population: The Tema Eye Survey. *JAMA Ophthalmol* 131: 651-658, 2013.

5. de Juan E Jr, Boyd S, Deem ME, Gifford HS III and Rosenman D: Glaucoma treatment device. US Patent 8721656. Filed December 22, 2006; issued May 13, 2014.
6. De Moraes CG, Liebmann JM, Greenfield DS, Gardiner SK, Ritch R and Krupin T: Low-pressure Glaucoma Treatment Study Group: Risk factors for visual field progression in the low-pressure glaucoma treatment study. *Am J Ophthalmol* 154: 702-711, 2012.
7. Bradley EW, Ruan MM, Vrable A and Oursler MJ: Pathway cross-talk between Ras/Raf and PI3K in promotion of M-CSF-induced MEK/ERK-mediated osteoclast survival. *J Cell Biochem* 104: 1439-1451, 2008.
8. Chen L, Huang L, Zeng A and He J: CYP1B1 gene mutations with incomplete penetrance in a Chinese pedigree with primary congenital glaucoma: A case report and review of literatures. *Int J Clin Exp Med* 8: 14538-14541, 2015.
9. Micheal S, Ayub H, Islam F, Siddiqui SN, Khan WA, Akhtar F, Qamar R, Khan MI and den Hollander AI: Variants in the ASB10 gene are associated with primary open angle glaucoma. *PLoS One* 10: e0145005, 2015.
10. Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS and Green MR: ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11: 237, 2010.
11. Gentleman R, Carey V, Huber W and Hahne F: GeneFilter: Methods for filtering genes from microarray experiments. R package version 1, 2011.
12. Stolovitzky GA, Kundaje A, Held GA, Duggar KH, Haudenschild CD, Zhou D, Vasicek TJ, Smith KD, Aderem A and Roach JC: Statistical analysis of MPSS measurements: Application to the study of LPS-activated macrophage gene expression. *Proc Natl Acad Sci USA* 102: 1402-1407, 2005.
13. Smolin JA, Gambetta JM and Smith G: Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Phys Rev Lett* 108: 70502, 2012.
14. Ben Slimane S: Bounds on the distribution of a sum of independent lognormal random variables. *Browes J magazines* 49: 975-978, 2001.
15. Bellot P, Olsen C, Salembier P, Oliveras-Vergés A and Meyer PE: NetBenchmark: A bioconductor package for reproducible benchmarks of gene regulatory network inference. *BMC Bioinformatics* 16: 312, 2015.
16. Zhao J, Yang TH, Huang Y and Holme P: Ranking candidate disease genes from gene expression and protein interaction: A Katz-centrality based approach. *PLoS One* 6: e24306, 2011.
17. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3, 2004.
18. Datta S, Satten GA, Benos DJ, Xia J, Heslin MJ and Datta S: An empirical bayes adjustment to increase the sensitivity of detecting differentially expressed genes in microarray experiments. *Bioinformatics* 20: 235-242, 2004.
19. Reiner A, Yekutieli D and Benjamini Y: Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19: 368-375, 2003.
20. Cover TM and Thomas JA (eds): *Elements of Information Theory*. 2nd edition. John Wiley & Sons, 2012.
21. Meyer PE, Lafitte F and Bontempi G: Minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* 9: 461, 2008.
22. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, Clarke ND, Altan-Bonnet G and Stolovitzky G: Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS One* 5: e9202, 2010.
23. Huynh-Thu VA, Irrthum A, Wehenkel L and Geurts P: Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5: pii: e12776, 2010.
24. Breiman L: Random forests. *Machine Learning* 45: 5-32, 2001.
25. Opgen-Rhein R and Strimmer K: From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *Bmc Sys Biol* 1: 37, 2007.
26. Reverter A and Chan EK: Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24: 2491-2497, 2008.
27. Hajian-Tilaki K: Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 4: 627-635, 2013.
28. Davis J and Goadrich M: The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. ACM, pp233-240, 2006.
29. Haythornthwaite C: Social network analysis: An approach and technique for the study of information exchange. *Lib Inf Sci Res* 18: 323-342, 1996.
30. Wasserman S and Faust K: *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
31. Barthélemy M: Betweenness centrality in large complex networks. *Eur Phy J B* 38: 163-168, 2004.
32. Yang TY: A simple rank product approach for analyzing two classes. *Bioinform Biol Insights* 9: 119-123, 2015.
33. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25: 402-408, 2001.
34. Bryman A and Cramer D: *Quantitative Data Analysis with SPSS 12 and 13. Analyzing Qualitative Data*, 2005.
35. Haynes W: Student's t-test. In: *Encyclopedia of Systems Biology*. Springer, pp2023-2025, 2013.
36. Marbach D, Schaffter T, Floreano D, Prill RJ and Stolovitzky G: The DREAM4 In Silico Network Challenge. Draft Version 0.3, 2009.
37. Ma S, Shah S, Bohnert HJ, Snyder M and Dinesh-Kumar SP: Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways. *PLoS Genet* 9: e1003840, 2013.
38. Chandrashekar G and Sahin F: A survey on feature selection methods. *Com Ele Eng* 40: 16-28, 2014.
39. Carey AL, Steinberg GR, Macaulay SL, Thomas WG, Holmes AG, Ramm G, Prelovsek O, Hohnen-Behrens C, Watt MJ, James DE, *et al*: Interleukin-6 increases insulin-stimulated glucose disposal in humans and glucose uptake and fatty acid oxidation in vitro via AMP-activated protein kinase. *Diabetes* 55: 2688-2697, 2006.
40. Gabay C: Interleukin-6 and chronic inflammation. *Arthritis Res Ther* 8 (Suppl 2): S3, 2006.
41. Lin KH, Feng SC, Shen YC, Wei LC, Liang CY, Chang CJ, Yang YY, Chiu CH and Wang CY: Interleukin-6(-174) locus polymorphism and serum IL-6 levels in normal tension glaucoma. *Ophthalmic Genet* 35: 255-257, 2014.
42. Russom P, Formichella C and Sappington R: Cleavage of IL-6 receptor alpha as a potential mechanism for IL-6 transsignaling in Glaucoma. *Investigative Ophthalmology Visual Sci* 54: 786, 2013.
43. Wu L, Chen Z, Huang G and Lu X: Association between IL-6 genetic polymorphisms and primary open-angle glaucoma risk in Chinese population. *Int J Clin Exp Med* 9: 9539-9544, 2016.
44. Micera A, Quaranta L, Esposito G, Floriani I, Pocobelli A, Saccà SC, Riva I, Manni G and Oddone F: Differential protein expression profiles in glaucomatous trabecular meshwork: An evaluation study on a small primary open angle glaucoma population. *Adv The* 33: 252-267, 2016.
45. Park JE, Son AI and Zhou R: Roles of EphA2 in development and disease. *Genes (Basel)* 4: 334-357, 2013.
46. Faoro L, Singleton PA, Cervantes GM, Lennon FE, Choong NW, Kanteti R, Ferguson BD, Husain AN, Tretiakova MS, Ramnath N, *et al*: EphA2 mutation in lung squamous cell carcinoma promotes increased cell survival, cell invasion, focal adhesions, and mammalian target of rapamycin activation. *J Biol Chem* 285: 18575-18585, 2010.
47. Reis LM, Tyler RC and Semina EV: Identification of a novel C-terminal extension mutation in EPHA2 in a family affected with congenital cataract. *Mol Vis* 20: 836-842, 2014.