Expression profile analysis to predict potential biomarkers for glaucoma: BMP1, DMD and GEM

Dao wei Zhang[1], Shenghai Zhang[1,2,3,4] and Jihong Wu[1,2,3,4]

[1] Eye Institute, Eye and ENT Hospital, College of Medicine, Fudan University, Shanghai, China
[2] Shanghai Key Laboratory of Visual Impairment and Restoration, Science and Technology Commission of Shanghai Municipality, Shanghai, China
[3] State Key Laboratory of Medical Neurobiology, Institutes of Brain Science and Collaborative Innovation Center for Brain Science, Shanghai, China
[4] Key Laboratory of Myopia, Ministry of Health, Shanghai, China

ABSTRACT

Purpose. Glaucoma is the second commonest cause of blindness. We assessed the gene expression profile of astrocytes in the optic nerve head to identify possible prognostic biomarkers for glaucoma.

Method. A total of 20 patient and nine normal control subject samples were derived from the GSE9944 (six normal samples and 13 patient samples) and GSE2378 (three normal samples and seven patient samples) datasets, screened by microarray-tested optic nerve head tissues, were obtained from the Gene Expression Omnibus (GEO) database. We used a weighted gene coexpression network analysis (WGCNA) to identify coexpressed gene modules. We also performed a functional enrichment analysis and least absolute shrinkage and selection operator (LASSO) regression analysis. Genes expression was represented by boxplots, functional geneset enrichment analyses (GSEA) were used to profile the expression patterns of all the key genes. Then the key genes were validated by the external dataset.

Results. A total 8,606 genes and 19 human optic nerve head samples taken from glaucoma patients in the GSE9944 were compared with normal control samples to construct the co-expression gene modules. After selecting the most common clinical traits of glaucoma, their association with gene expression was established, which sorted two modules showing greatest correlations. One with the correlation coefficient is 0.56 ($P = 0.01$) and the other with the correlation coefficient is -0.56 ($P = 0.01$). Hub genes of these modules were identified using scatterplots of gene significance versus module membership. A functional enrichment analysis showed that the former module was mainly enriched in genes involved in cellular inflammation and injury, whereas the latter was mainly enriched in genes involved in tissue homeostasis and physiological processes. This suggests that genes in the green–yellow module may play critical roles in the onset and development of glaucoma. A LASSO regression analysis identified three hub genes: Recombinant Bone Morphogenetic Protein 1 gene (*BMP1*), Duchenne muscular dystrophy gene (*DMD*) and mitogens induced GTP-binding protein gene (*GEM*). The expression levels of the three genes in the glaucoma group were significantly lower than those in the normal group. GSEA further illuminated that *BMP1*, *DMD* and *GEM* participated in the occurrence and development of some important metabolic progresses. Using the GSE2378 dataset, we confirmed the high

Corresponding authors
Shenghai Zhang, zsheent@163.com
Jihong Wu, jihongwu@fudan.edu.cn

validity of the model, with an area under the receiver operator characteristic curve of 85%.

Conclusion. We identified several key genes, including *BMP1*, *DMD* and *GEM,* that may be involved in the pathogenesis of glaucoma. Our results may help to determine the prognosis of glaucoma and/or to design gene- or molecule-targeted drugs.

INTRODUCTION

Glaucoma is the leading cause of irreversible blindness worldwide (*Jonas et al., 2017*). Like tumors and cardiovascular disorders, it is strongly associated with the deterioration of quality of life, therefore it is an important medical problem that requires treatment (*Quaranta et al., 2016*). Gene expression aberrations have been widely recognized to participate in the pathological process of tumor (*Flavahan, Gaskell & Bernstein, 2017*). This abnormality is not limited to tumors, but has been linked to a variety of ophthalmopathy (*Heavner & Pevny, 2012*; *Solinis et al., 2015*), thus we assume that they also contribute to the development and progression of glaucoma. Systemic diseases, genetic factors, and environmental factors may also play critical roles in the pathogenesis of glaucoma. Long-term studies have repeatedly shown that elevated intraocular pressure (IOP), age, and genetics are the main risk factors for glaucoma (*Doucette et al., 2015*; *McMonnies, 2017*). However, the treatment of glaucoma is limited to drugs that lower IOP (*Lusthaus & Goldberg, 2016*) and surgical therapy (*Fenwick et al., 2019*), which may temporarily alleviate IOP.

Glaucoma is generally characterized by a reduction in the number of retinal ganglion cells, the thinning of the retinal nerve fiber layer, and the cupping of the optic disc. The optic nerve head is a key anatomical site that shows signs of early glaucomatous damage (*Minckler, Bunt & Johanson, 1977*). In addition to pathological changes of the retinal ganglion cells, the optic nerve head undergoes reactive astrocyte remodeling, particularly in response to stress or other pathological disorders. This can lead to a series of morphologic, gene expression, and functional changes (*Sofroniew & Vinters, 2010*). It has also been reported that astrocytes play a key role in the disease process, but there is limited evidence about the specific pathological events that occur in these cells.

In recent decades, advances in bioinformatics, high-throughput sequencing, and genetic association studies have greatly accelerated the discovery of genes and genomic regions involved in ophthalmic diseases, including glaucoma. These studies have demonstrated that glaucoma is a complex disease (*MacGregor et al., 2018*). Previous genome-wide association studies identified over 10 genes associated with primary open-angle glaucoma, including myocilin (*MYOC*), optineurin (*OPTN*) and WD repeat domain 36 (*WDR36*) (*Liu & Allingham, 2017*). Mutations in *MYOC* cause a cascade of abnormalities in the trabecular meshwork, including the intracellular retention of myocilin, reduced aqueous outflow, increased IOP, and glaucoma (*Fini, 2017*). Optineurin regulates many physiological

processes, including membrane trafficking, protein secretion, cell division, autophagy, and host defense against pathogens (*Kachaner et al., 2012*). WDR36 is a nucleolar protein involved in the maturation of 18S rRNA and the mutation of *WDR36* can delay the formation of 18S rRNA and the apoptosis of human trabecular meshwork cells. A number of other genes are also involved in secondary glaucoma, including *SBF2*, *EPHA2*, *TRPM3* and *TMEM98* (*Doucette et al., 2015*).

In this study, we sought to identify novel biomarkers associated with the pathogenesis of glaucoma. To achieve this, we used a weighted gene coexpression network analysis (WGCNA) to construct a network of coexpressed genes, in order to describe the correlations among genes across multiple samples. Although WGCNA may not be as effective as other techniques in identifying modules with high functional relevance and biological significance, it is the most widely used for this purpose (*Kakati et al., 2019*). We used this correlation-based gene screening method to identify candidate biomarkers and/or therapeutic targets for glaucoma.

MATERIAL AND METHODS

Data source

We searched the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) using the keyword "glaucoma". The datasets were yielded according to the following criteria: (1) datasets contain samples from both normal and glaucoma patients along with necessary clinical characteristics such as gender and age; (2) datasets exhibit original expression profiles derived from microarray which already had been background corrected and standardized; (3) datasets have a relatively sufficient sample size for further analysis. Two datasets were selected with GSE9944 (*Lukas et al., 2008*) using as the training dataset and GSE2378 (*Hernandez et al., 2002*) as the validation dataset.

WGCNA

WGCNA is a bioinformatics tool that we applied to construct the expression patterns of genes from multiple samples, generating clusters of genes with similar expression patterns, allowing the researcher to analyze the correlations between modules and specific traits or phenotypes (*Langfelder & Horvath, 2008*). To identify genes related to glaucoma, we assumed that the gene networks obeyed a scale-free distribution in the WGCNA. Pairwise Pearson's correlation coefficients were used to evaluate the co-expression relationships among all the genes, and were converted into a scale-free network by transforming the correlation matrix to an adjacency matrix with a soft threshold value, which specified the range of changes in the detected data. The soft threshold value was selected as the standard of a scale-free distribution. Based on the adjacency matrix, genes with absolute high correlation were clustered into the same module (*Feltrin et al., 2019*). The dissimilarity of the topological matrix was then incorporated into an unsupervised hierarchical cluster with a dynamic tree-cutting algorithm. The branches of the clusters were defined as modules (*Oros Klein et al., 2016*) and each module represented a specific gene expression profile that was generalized by the main component eigengene (*Zhang et al., 2017*). The scatterplots of gene significance and module membership were painted to define hub genes. All the above

processes of WGCNA were the realization by R program. Besides, the topological overlap of intramodules and adjacency modules was used for selecting the functional modules, and modules with low or no correlation to glaucoma (P value ≥ 0.01) were excluded.

Functional enrichment analysis

A functional enrichment analysis was performed on the modules most relevant to the glaucoma traits using Metascape (http://metascape.org/), which utilized the well-adopted hypergeometric test and Benjamini-Hochberg p-value correction algorithm to identify all the genes. We identified all the statistically significantly enriched terms by calculating the cumulative hypergeometric P values and enrichment factors, which were then used to filter the data. The significant terms were then hierarchically clustered into a tree based on the statistical similarities (derived from the κ values) among the member genes. A κ value of 0.3 was used as the threshold to divide the tree into term clusters. A P value of <0.01 was used as the screening threshold for significant pathways (*Zhou et al., 2019*).

LASSO regression analysis

We performed LASSO regression analysis with integration of the genes in the modules by using the glmnet R package (version 2.0.16, https://cran.r-project.org/web/packages/glmnet) (*Friedman, Hastie & Tibshirani, 2010*), as our purpose was to extract the precise expression quantity resorting dimensionality reduction algorithm. We regularized the LASSO regression coefficient to prevent over-fitting the results in order to screen the key genes.

Gene set enrichment analysis

In order to compare difference in glaucoma group and normal group, we first applied the point graphs to represent genes expression levels in the GSE9944. Geneset enrichment analyses (GSEA) was divided into high and low groups according to the median value of gene expression, and then analyzed with GSEA enrichment software (version 4.0.2, http://software.broadinstitute.org/gsea/index.jsp) (*Subramanian et al., 2005*). The significant pathway was presented by a table whose screening condition was P value <0.05. GSEA was performed to identify the pathways associated with key genes (*Ye et al., 2019*) in the training dataset.

External validation of key genes

The key genes identified with the training dataset were evaluated the accuracy of these genes with a logistic regression analysis using the pROC R package (version : 1.16.1, https://cran.r-project.org/web/packages/pROC) (*Robin et al., 2011*). Key genes were validated via the external dataset (GSE2378) with a receiver operating characteristic (ROC) curve analysis to determine the sensitivity and specificity of the analysis. The area under the curve (AUC) represents the likelihood that the eigengene can be considered as biomarker (*Yu et al., 2018*).

Statistical analysis

The two.sided wilcox test was used to determine the statistical significance when comparing the glaucoma tissue with the normal tissue. Statistical analysis was carried out using R software, version 3.6.2.

RESULTS

Sample traits and data

Two gene expression datasets of optic nerve head astrocytes (GSE9944 and GSE2378) obtained from donors with or without glaucoma were downloaded from GEO in October 2019, and the GPL version of two datasets with the largest disease samples size was selected for further analysis. The training dataset, GSE9944 dataset, comprised 63 samples, and 13 patient samples and 6 normal samples were finally selected for WGCNA to screen potential genes. The GSE2378 dataset comprised 15 samples in two GPLs, we chose the one with a larger sample size ($N = 13$). Among these 13 samples, however, there were three samples without clarification of disease or normal, so we finally kept 10 samples as the external validation dataset.

Glaucoma-related WGCNA modules and genes

All of the genes included in the GEO dataset were subjected to WGCNA. The top 25% of genes that showed the greatest variance were retained for subsequent WGCNA analysis, and ultimately 8,086 genes and 19 samples were included. For network analysis, the soft threshold power for matrix transformation was confirmed to be 6, the scale independence was 0.85, and the mean connectivity of the co-expression network was high to ensure a scale-free network (Fig. 1). We constructed the co-expression modules and identified 10 glaucoma-related modules, which were arbitrarily designated black (440 genes), blue (2,053 genes), brown (3,706 genes), cyan (65 genes), green (1,088 genes), green-yellow (674 genes), gray (5 genes), midnight-blue (59 genes), red (364 genes), and tan (125 genes) (Fig. 2). We selected the most common clinical traits (gender and age) of disease and then linked them to the gene expression modules based on the correlations between the eigengenes and the clinical traits of common expression pattern modules. As shown in Fig. 3, gender and age were included in the module–trait analyses. The green-yellow ($r = 0.56$, $P = 0.01$) and red modules ($r = -0.56$, $P = 0.01$) were strongly associated with age.

We constructed scatterplots of gene significance versus module membership for the green-yellow and red modules (Fig. 4). Genes with an absolute gene significance score of >0.5 and an absolute module membership score of >0.8 were regarded as the core genes in each module.

Enrichment analysis

An enrichment analysis of the core genes based on Gene Ontology (GO) revealed that the green–yellow module was significantly enriched in genes involved in the development of glaucoma, particularly the regulation of reactive oxygen species metabolic processes (GO: 2000377), acute inflammatory responses (GO: 0002526), the cyclooxygenase pathway (GO:
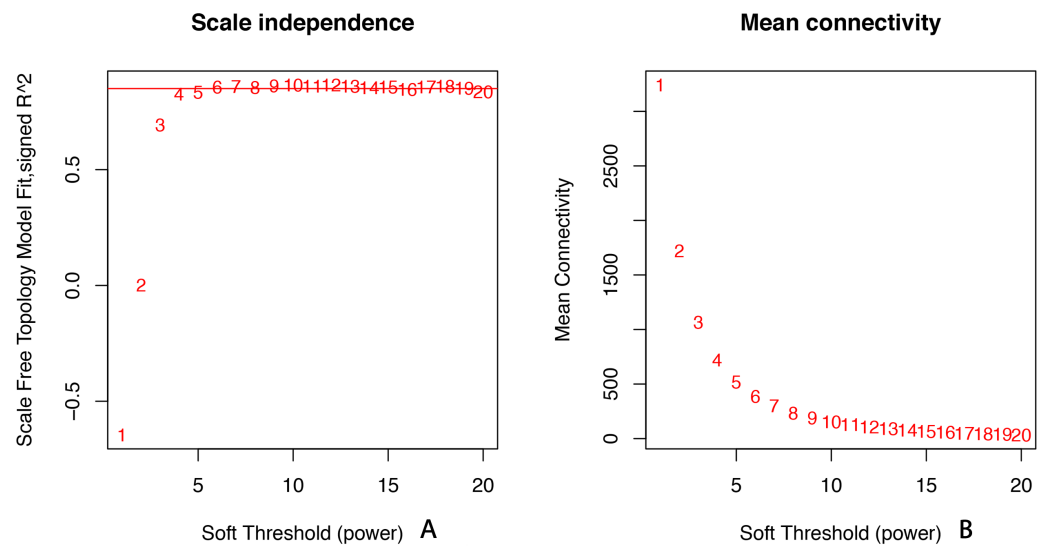
Figure 1 Selection of soft threshold powers. (A) Scale independence. Analysis of scale-free topology fit index for soft threshold powers (β). Red line indicates the correlation coefficient (0.85). (B) Mean connectivity of gene co-expression modules for different soft threshold powers.
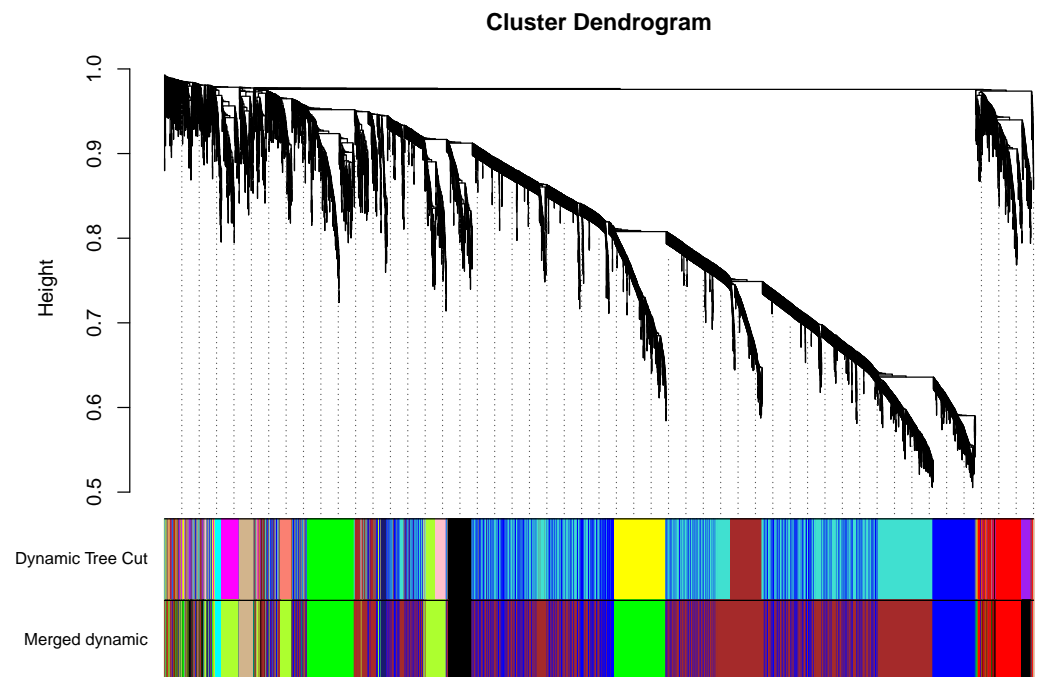
Figure 2 Identification of a co-expression module in glaucoma. The branches of the cluster dendrogram correspond to the 10 gene modules. Each piece of the leaves on the cluster dendrogram corresponds to a different gene module.
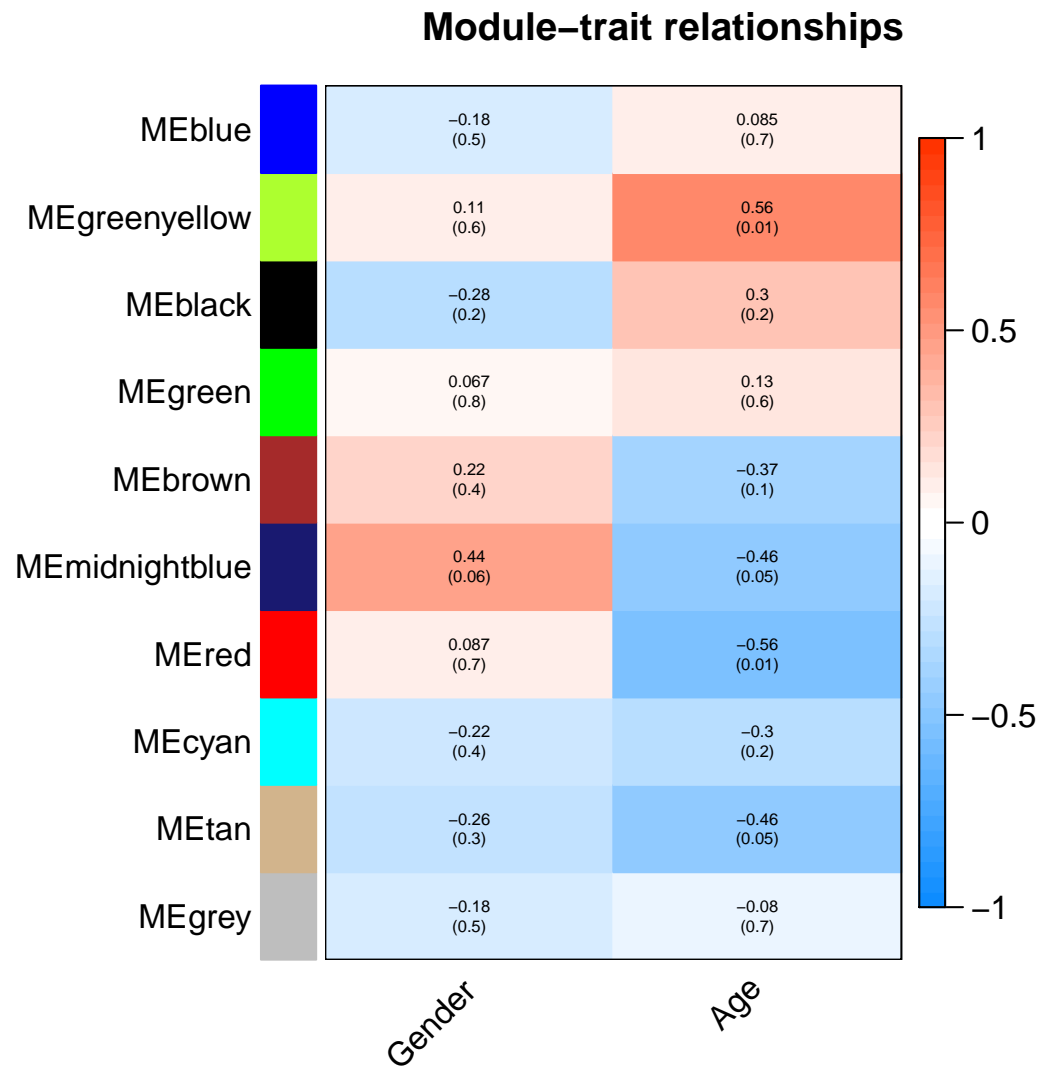
Figure 3 Correlation between the gene module and clinical traits. Gender and Age are included. The correlation coefficient in each cell represented the correlation between the gene module and the clinical traits. The corresponding *P*-value is also annotated. The green-yellow and the red were the most positively and negatively relevant modules to age, respectively.

0019371), blood vessel development (GO: 0001568), and the response to acidic chemicals (GO: 0001101) (Fig. 5A). The red module was significantly enriched in genes involved in the response to organophosphorus (GO: 0046683) and polyol metabolic processes (GO: 0019751) (Fig. 5B). These results demonstrated the marked differences in the genes present in each module, and that the green-yellow module was predominantly enriched in genes involved in cellular inflammation and injury. Based on gene enrichment, we speculated that the green–yellow module was the most important module in age-related glaucoma. By contrast, the red module was mainly enriched in genes involved in tissue physiological processes and cell cycles, which suggested that the encoded proteins may act in the regulation of pathological processes in these cells.
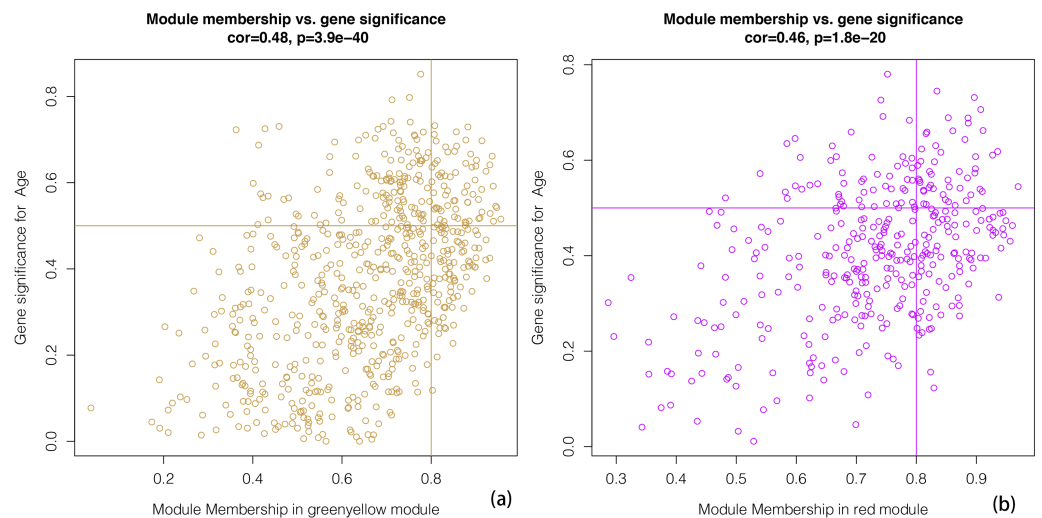
Figure 4 **A scatter plot of Gene Significance (GS) vs. Module Membership (MM).** A scatter plot of Gene Significance (GS) vs. Module Membership (MM) in the greenyellow (A). A scatter plot of Gene Significance (GS) vs. Module Membership (MM) in the red (B). We define cone genes as which absolute value of Gene Signicance is greater than 0.5 and the absolute value of Module Membership is greater than 0.8.

LASSO regression analysis

LASSO regression is widely used in tumor analysis (*He et al., 2019*; *Zhao et al., 2015*), but is rarely applied in other diseases, especially to use WGCNA and LASSO regression analysis to screen hub gene together, which is an innovative aspect of our research. We used a LASSO regression analysis to identify the key genes in the green-yellow and red modules. Then we screened three key genes: *BMP1*, *DMD* and *GEM* in green–yellow module as a consequence (Fig. 6).

GESA

Genes expression levels were shown in the Fig. 7. Interestingly, all these three genes were significantly expressed higher in the normal samples than in the glaucoma samples. Based on gene expression level, GESA was performed to identify the pathways associated with key genes. The enrichment scores of these differentially expressed genes in important terms were presented in Table 1. BMP1 was significantly enriched in TGF-β signaling pathway, ECM receptor interaction, natural killer cell mediated cytotoxicity (Fig. 8A). DMD was significantly enriched in focal adhesion, ECM receptor interaction, VEGF signaling pathway (Fig. 8B). GEM was significantly enriched in ECM receptor interaction, natural killer cell mediated cytotoxicity (Fig. 8C). Particularly worth noting is that all of the three key genes were enriched in the ECM receptor interaction, and both of *BMP1* and *GEM* were enriched in natural killer cell mediated cytotoxicity.

Validation of the key genes

Finally, we sought to validate the expression of the three key genes *BMP1*, *DMD* and *GEM* by using the GSE2378 dataset, and the high AUC value (85%) showed in Fig. 9 indicated the accurate discrimination of pathological and normal samples.
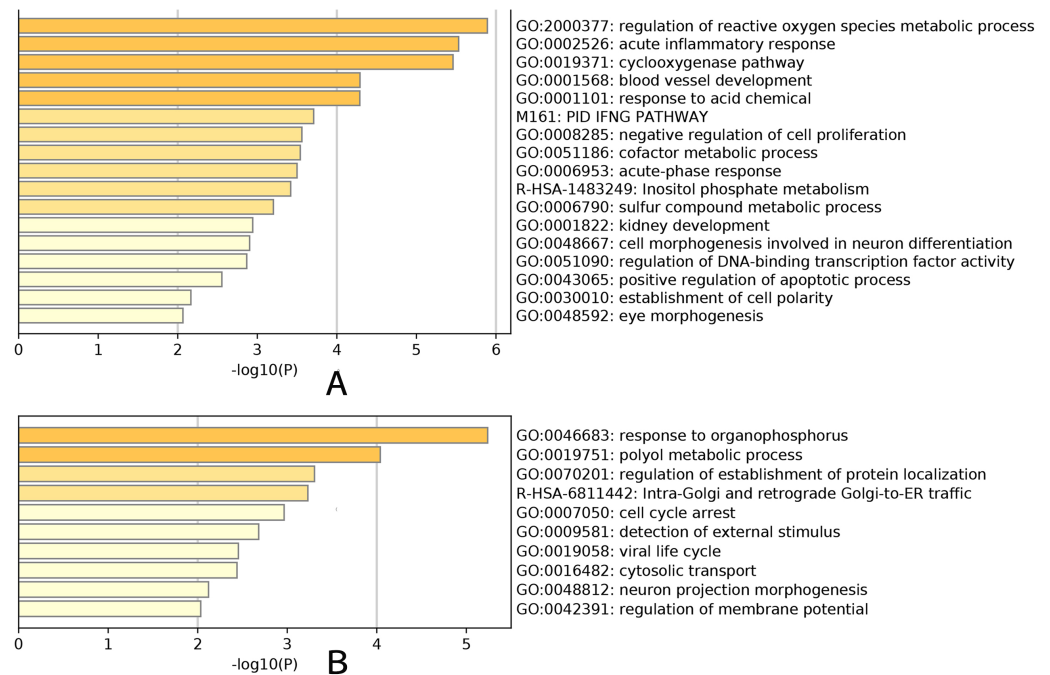
Figure 5 **GO analysis.** Enrichment analysis result for the yellow-green module (A). Enrichment analysis result for the red module (B). Yellow-green module was significantly enriched in development of glaucoma, especially in regulation of reactive oxygen species metabolic process (GO:2000377), acute inflammatory response (GO:0002526), cyclooxygenase pathway (GO:0019371), blood vessel development (GO:0001568) and response to acid chemical (GO:0001101). Genes in red module were significantly enriched in response to organophosphorus (GO:0046683) and polyol metabolic process (GO:0019751).

Our results indicate that *BMP1*, *DMD* and *GEM* are potential biomarkers of glaucoma. These genes were validated with a separate database. These findings may provide novel insight into the pathogenesis of glaucoma, and could be helpful for its early diagnosis, prevention, and treatment.

DISCUSSION

The occurrence and development of some types of glaucoma may be determined by several genes, especially in early-onset and adult-onset forms, such as developmental glaucoma, juvenile-onset primary open angle glaucoma, congenital glaucoma, and familial normal-tension glaucoma (*Wang & Wiggs, 2014*). Therefore, we tried to identify genes involved in glaucoma, instead of discuss existing therapies and known mechanisms. In this study, we used WGCNA to construct 10 coexpression modules containing 8,606 genes identified in 19 human optic nerve head samples in order to determine the relationships between the glaucoma transcriptome and clinic traits. As WGCNA is reportedly to be more reliable and more biologically significant than other methods (*Chou et al., 2014*), we used this method to form clusters of predictive functionally related genes. In this way, we identified modules and selected genes that might be considered as biomarkers for the diagnosis and/or treatment of glaucoma. Two coexpression modules were strongly
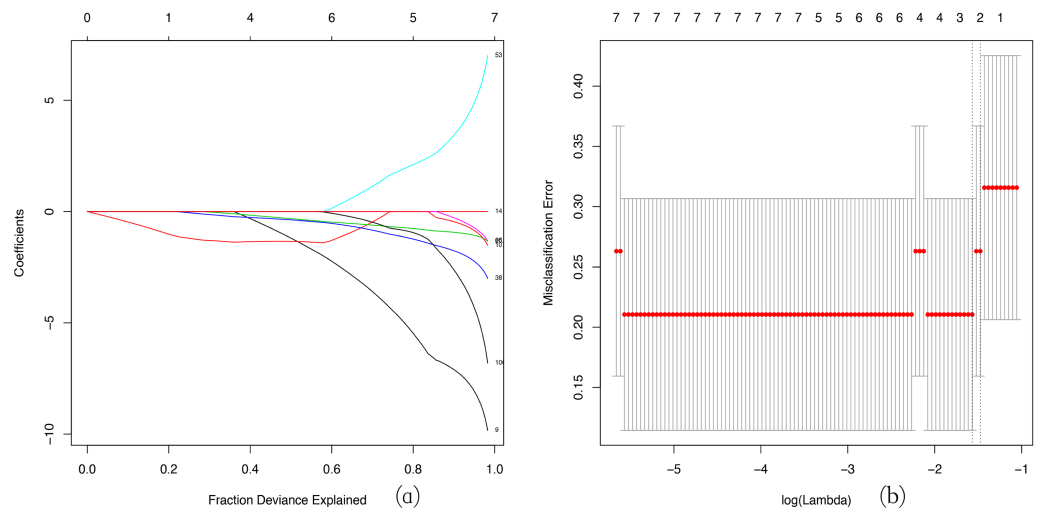
Figure 6 LASSO regression analysis. LASSO coefficient profiles of the module genes (A). Selection of the tuning parameter (λ) in the LASSO model through cross-validation procedure was plotted as a function of log (λ) (B). The y-axis represents Misclassification Error, and the lower x-axis represents the log (λ). Numbers along the upper x-axis represent the average number of predictors. Red dots indicate average deviance values for each model with a given λ, where the model provides its best fit to data.

associated with the clinical traits of glaucoma, particularly age. The green–yellow module was enriched in genes involved in cellular inflammation and injury, whereas the red module was mainly enriched in genes involved in tissue physiological processes and cell cycles.

After the two coexpression modules were constructed with WGCNA, we identified three key hub genes: *BMP1*, *DMD* and *GEM*. Based on genes expression presentation and results of GSEA analysis, we found that three key genes were significantly lower in glaucoma samples than in normal samples and they enriched in some of the same pathways. This further suggests that they may be potential biomarkers of the disease.

BMP1, also known as procollagen C-proteinase, is involved in the maturation of collagen, which is necessary for bone and cartilage growth and structural maintenance. It is synthesized throughout the human body, except in the brain (*Brown & Goldstein, 1986*). Loss-of-function mutations in *BMP1* result in abnormal collagen formation and occur in various autosomal recessive diseases associated with defective osteogenesis (*Syx et al., 2015*). BMP1 is also known to cleave extracellular macromolecules, including probiglycan (*Scott et al., 2000*) and prolaminin 5 (*Amano et al., 2000*), when the extracellular matrix is too excessive. Besides, BMP1 is necessary to generate functional high-density lipoprotein particles for reverse cholesterol transport (*Riggs & Rohatgi, 2019*), and contributes to renal fibrosis in chronic kidney disease by affecting the maturation and deposition of collagen and subsequent profibrotic responses and inflammation (*Bai et al., 2019*). Therefore, it is difficult to predict whether BMP1 has beneficial or deleterious effects in glaucoma.

DMD, with 79 exons and tightly regulated introns, is the largest protein-coding gene and encodes an important cytoskeletal protein. Mutations in *DMD* cause Duchenne muscular dystrophy (*Hoffman, Brown Jr & Kunkel, 1987*). DMD plays critical roles in the
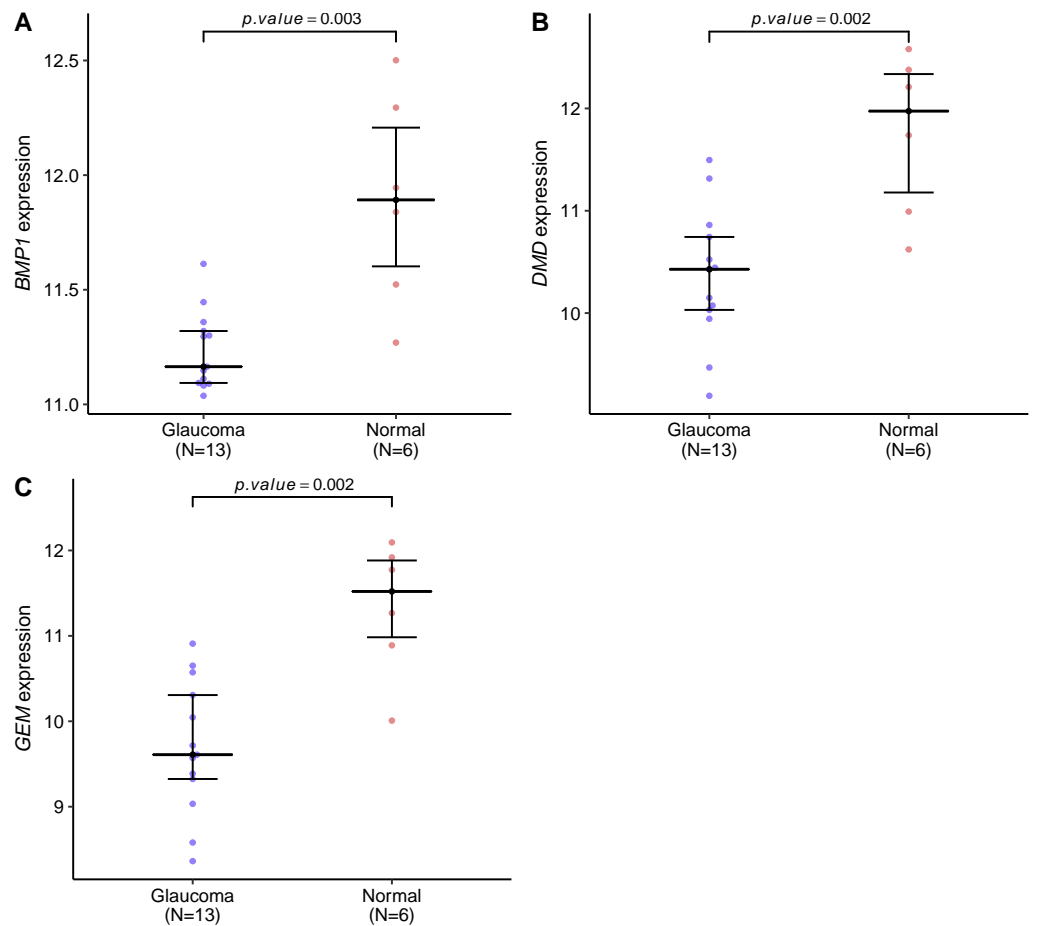
Figure 7 **BMP1, DMD and GEM expression levels in the glaucoma group and normal group.** The expression levels of all these key genes were significantly lower in the glaucoma group than in the normal group (A–C).

formation and maintenance of neuromuscular junctions (*Ahn & Kunkel, 1993*), and has been identified as a signaling protein involved in muscle contraction and other functions, as well as in dystrophin function (*Anderson et al., 2002*). It is prominently expressed in the eye and brain. Depending on the promoter used, *DMD* encodes both dystrophin and Dp71, which show different expression patterns during embryonic development in zebrafish, especially in the eye and germ-layer structures (*Bolaños-Jiménez et al., 2001*). Preliminary results have indicated that DP260 (and possibly DP71) is associated with abnormal b-wave amplitudes on dark-adapted electroretinography (ERG) in mice (*Pillers et al., 1999*). More-recent studies have demonstrated a strong association between *DMD* mutations that affect various DMD isoforms and abnormal scotopic ERGs and severe neurodevelopmental problems (*Ricotti et al., 2016*). Genetic studies have linked myocilin with open-angle glaucoma (*Stone et al., 1997*). Taken together, these findings provide strong evidence that *DMD* may be used as a biomarker of dystrophin function in glaucoma.

Table 1 Enrichment score s of differentially expressed genes in important terms.

GENE	NAME	ES	NES	NOM *p*-val
BMP1	KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	0.41486168	1.4794924	0.01814516
	KEGG_HEMATOPOIETIC_CELL_LINEAGE	0.5492907	1.4255947	0.02554028
	KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO _AND_NEOLACTO_SERIES	0.58440936	1.4215839	0.03406814
	KEGG_ECM_RECEPTOR_INTERACTION	0.39169973	1.3057723	0.0492126
DMD	KEGG_GLIOMA	0.47700384	1.8822955	0
	KEGG_RENAL_CELL_CARCINOMA	0.48509955	1.7634672	0.004
	KEGG_PANCREATIC_CANCER	0.46747884	1.6679943	0.00823045
	KEGG_FOCAL_ADHESION	0.42024386	1.6549476	0.00205339
	KEGG_NON_SMALL_CELL_LUNG_CANCER	0.39712295	1.6283306	0.02301255
	KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	0.57694465	1.6273366	0.00409836
	KEGG_VIBRIO_CHOLERAE_INFECTION	0.36380395	1.6116077	0.03846154
	KEGG_DILATED_CARDIOMYOPATHY	0.49189743	1.5051197	0.02414487
	KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR _CARDIOMYOPATHY_ARVC	0.4962006	1.5024999	0.02857143
	KEGG_VEGF_SIGNALING_PATHWAY	0.45013994	1.480682	0.03617021
	KEGG_SMALL_CELL_LUNG_CANCER	0.3820947	1.4728578	0.02794411
	KEGG_MELANOMA	0.40423834	1.4425186	0.03607215
	KEGG_GAP_JUNCTION	0.3834123	1.4365592	0.0359408
	KEGG_PATHWAYS_IN_CANCER	0.31513023	1.3521765	0.036
	KEGG_ECM_RECEPTOR_INTERACTION	0.40342882	1.3490801	0.02663934
GEM	KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO _AND_NEOLACTO_SERIES	0.62023604	1.5112221	0.0155642
	KEGG_JAK_STAT_SIGNALING_PATHWAY	0.4429717	1.423994	0.0130841
	KEGG_AUTOIMMUNE_THYROID_DISEASE	0.6393696	1.4187429	0.02808989
	KEGG_PRIMARY_IMMUNODEFICIENCY	0.5353427	1.4012718	0.01801802

Notes.
ES, enrichment score; NES, normalized enrichment score.

GEM is a small GTP-binding protein belonging to the RAS superfamily of monomeric G-proteins, with extensive biological functions. It is already widely accepted as a protein involved in rat obesity (*Bourne, Sanders & McCormick, 1991*). GEM has two key biological effects, depending on its site of expression: the inhibition of voltage-gated calcium channel activity and the inhibition of RHO-kinase-mediated cytoskeletal reorganization (*Ward et al., 2004*). Previous studies have shown that GEM is a regulatory protein that participates in receptor-mediated signal transduction at the plasma membrane (*Maguire et al., 1994*). It is widely expressed in many tissues and organs, including the spleen, lung and large intestine of the goat (*Xu et al., 2018*), and it is overexpressed in skeletal muscle in individuals with type 2 diabetes (*Maguire et al., 1994*). Here, we propose that the expression of *GEM* in the eye might also be clinically significant in glaucoma.

In conclusion, we have shown that the green–yellow module is the module most relevant to glaucoma. We also identified three key genes, *BMP1*, *DMD* and *GEM* that might be served as potential biomarkers of glaucoma. Although the biological and clinical relevance
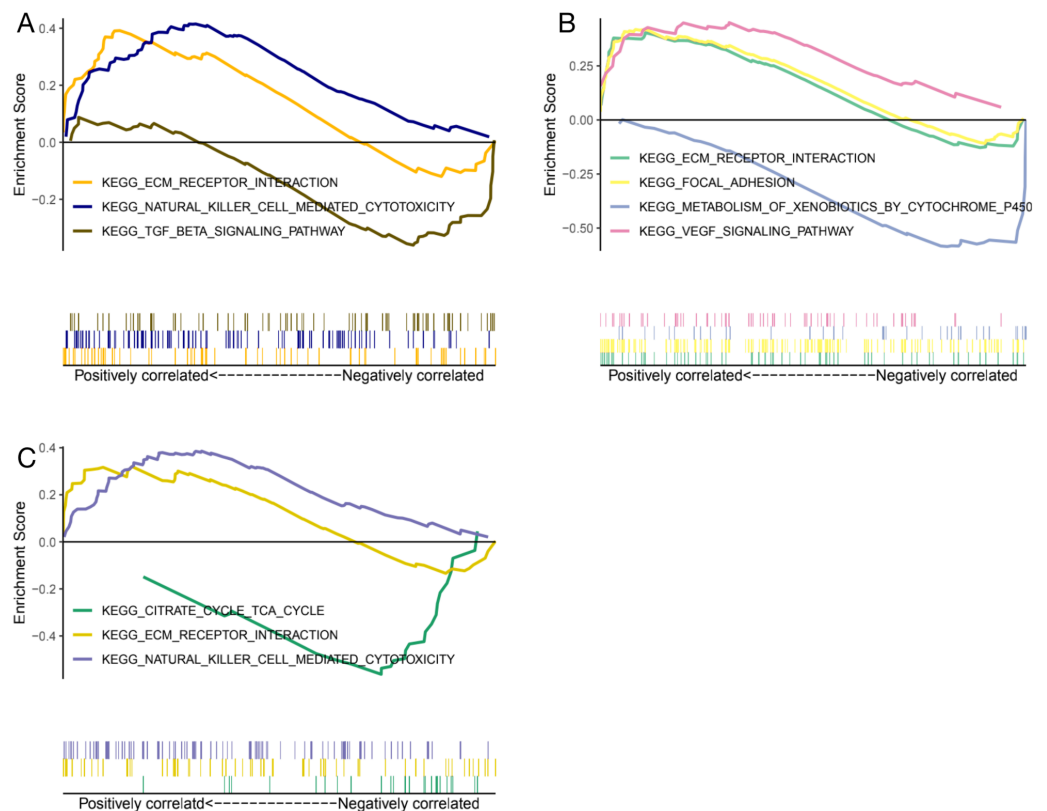
Figure 8 GESA enrichment of BMP1, DMD and GEM. BMP1 was significantly enriched in TGF-β signaling pathway, ECM receptor interaction, natural killer cell mediated cytotoxicity (A). DMD was significantly enriched in focal adhesion, ECM receptor interaction, VEGF signaling pathway (B). GEM was significantly enriched in ECM receptor interaction, natural killer cell mediated cytotoxicity (C).

of these genes are not completely understood, they may be involved in a number of cellular processes relevant to the pathogenesis of glaucoma, such as inflammation or oxidative stress. Our results may also contribute to the development of novel approaches with great efficacy to diagnose and treat glaucoma, as they are genes of particular interest in relation to the development or progression of the disease.

There were certain limitations to our study, for example, the sample size was relatively small with Chinese ethnicity included only. Although we screened three key genes through a combination of multiple analyses, this study still lack experimental verification. To further unravel the function of these newly genes in the pathogenesis of glaucoma, studies with a larger sample size should focus on confirming the relevance between biological and clinical changes, and validating the value of these potential genetic targets by experiments to prevent the onset or the progression of glaucoma.

ACKNOWLEDGEMENTS

The author would like to thank all the patients who provided samples to the database and the staff for their valuable contributions to this research.
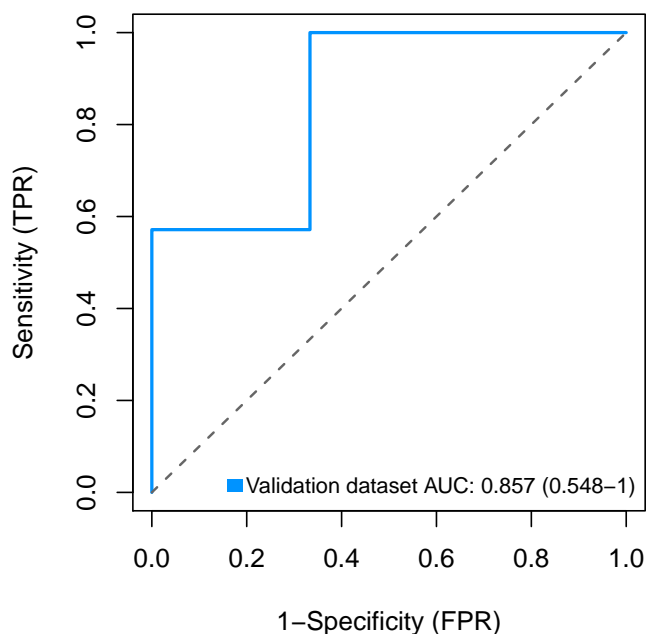
Figure 9 **Validation for the key genes using the GSE2378 dataset.** TPR, true positive rate, represent sensitivity. FPR, false positive rate, represent specificity.

Funding
The authors received no funding for this work.

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Dao wei Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Shenghai Zhang and Jihong Wu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability
    The data and code are available at GitHub: https://18211260008.github.io/glaucoma/. The data is also available at NCBI GEO: GSE9944, GSE2378.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.9462#supplemental-information.

REFERENCES

Ahn AH, Kunkel LM. 1993. The structural and functional diversity of dystrophin. *Nature Genetics* **3**:283–291 DOI 10.1038/ng0493-283.

Amano S, Scott IC, Takahara K, Koch M, Champliaud MF, Gerecke DR, Keene DR, Hudson DL, Nishiyama T, Lee S, Greenspan DS, Burgeson RE. 2000. Bone morphogenetic protein 1 is an extracellular processing enzyme of the laminin 5 gamma 2 chain. *Journal of Biological Chemistry* **275**:22728–22735 DOI 10.1074/jbc.M002345200.

Anderson JL, Head SI, Rae C, Morley JW. 2002. Brain function in Duchenne muscular dystrophy. *Brain* **125**:4–13 DOI 10.1093/brain/awf012.

Bai M, Lei J, Wang S, Ding D, Yu X, Guo Y, Chen S, Du Y, Li D, Zhang Y, Huang S, Jia Z, Zhang A. 2019. BMP1 inhibitor UK-383, 367 attenuates renal fibrosis and inflammation in CKD. *American Journal of Physiology: Renal Physiology* **317**:F1430–F1438 DOI 10.1152/ajprenal.00230.2019.

Bourne HR, Sanders DA, McCormick F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**:117–127 DOI 10.1038/349117a0.

Brown MS, Goldstein JL. 1986. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**:34–47 DOI 10.1126/science.3513311.

Chou WC, Cheng AL, Brotto M, Chuang CY. 2014. Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. *BMC Genomics* **15**:300 DOI 10.1186/1471-2164-15-300.

Doucette LP, Rasnitsyn A, Seifi M, Walter MA. 2015. The interactions of genes, age, and environment in glaucoma pathogenesis. *Survey of Ophthalmology* **60**:310–326 DOI 10.1016/j.survophthal.2015.01.004.

Feltrin AS, Tahira AC, Simoes SN, Brentani H, Martins Jr DC. 2019. Assessment of complementarity of WGCNA and NERI results for identification of modules associated to schizophrenia spectrum disorders. *PLOS ONE* **14**:e0210431 DOI 10.1371/journal.pone.0210431.

Fenwick E, Man R, Aung T, Ramulu P, Lamoureux E. 2019. Beyond intraocular pressure: optimizing patient-reported outcomes in glaucoma. *Progress in Retina and Eye Research* **76**:100801 DOI 10.1016/j.preteyeres.2019.100801.

Fini ME. 2017. Another piece of the puzzle: MYOC and Myocilin glaucoma. *Investigative Ophthalmology and Visual Science* **58**:5319 DOI 10.1167/iovs.17-23045.

Flavahan WA, Gaskell E, Bernstein BE. 2017. Epigenetic plasticity and the hallmarks of cancer. *Science* **357(6348)**:eaal2380 DOI 10.1126/science.aal2380.

Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**:1–22.

He A, He S, Peng D, Zhan Y, Li Y, Chen Z, Gong Y, Li X, Zhou L. 2019. Prognostic value of long non-coding RNA signatures in bladder cancer. *Aging* **11**:6237–6251 DOI 10.18632/aging.102185.

Heavner W, Pevny L. 2012. Eye development and retinogenesis. *Cold Spring Harbor Perspectives in Biology* **4(12)**:a008391 DOI 10.1101/cshperspect.a008391.

Hernandez MR, Agapova OA, Yang P, Salvador-Silva M, Ricard CS, Aoi S. 2002. Differential gene expression in astrocytes from human normal and glaucomatous optic nerve head analyzed by cDNA microarray. *Glia* **38**:45–64 DOI 10.1002/glia.10051.

Bolaños-Jiménez F, Bordais A, Behra M, Strähle U, Sahel J, Rendón A. 2001. Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish. *Mechanisms of Development* **102**:239–241 DOI 10.1016/S0925-4773(01)00310-0.

Jonas JB, Aung T, Bourne RR, Bron AM, Ritch R, Panda-Jonas S. 2017. Glaucoma. *Lancet* **390**:2183–2193 DOI 10.1016/s0140-6736(17)31469-1.

Hoffman EP, Brown Jr RH, Kunkel LM. 1987. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* **51**:919–928 DOI 10.1016/0092-8674(87)90579-4.

Kachaner D, Genin P, Laplantine E, Weil R. 2012. Toward an integrative view of Optineurin functions. *Cell Cycle* **11**:2808–2818 DOI 10.4161/cc.20946.

Kakati T, Bhattacharyya DK, Barah P, Kalita JK. 2019. Comparison of methods for differential co-expression analysis for disease biomarker prediction. *Computers in Biology and Medicine* **113**:103380 DOI 10.1016/j.compbiomed.2019.103380.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Liu Y, Allingham RR. 2017. Major review: molecular genetics of primary open-angle glaucoma. *Experimental Eye Research* **160**:62–84 DOI 10.1016/j.exer.2017.05.002.

Lukas TJ, Miao H, Chen L, Riordan SM, Li W, Crabb AM, Wise A, Du P, Lin SM, Hernandez MR. 2008. Susceptibility to glaucoma: differential comparison of the astrocyte transcriptome from glaucomatous African American and Caucasian American donors. *Genome Biology* **9**:R111 DOI 10.1186/gb-2008-9-7-r111.

Lusthaus JA, Goldberg I. 2016. Emerging drugs to treat glaucoma: targeting prostaglandin F and E receptors. *Expert Opinion on Emerging Drugs* **21**:117–128 DOI 10.1517/14728214.2016.1151001.

MacGregor S, Ong JS, An J, Han X, Zhou T, Siggs OM, Law MH, Souzeau E, Sharma S, Lynn DJ, Beesley J, Sheldrick B, Mills RA, Landers J, Ruddle JB, Graham SL, Healey PR, White AJR, Casson RJ, Best S, Grigg JR, Goldberg I, Powell JE, Whiteman DC, Radford-Smith GL, Martin NG, Montgomery GW, Burdon KP, Mackey DA, Gharahkhani P, Craig JE, Hewitt AW. 2018. Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. *Nature Genetics* **50**:1067–1071 DOI 10.1038/s41588-018-0176-y.

Maguire J, Santoro T, Jensen P, Siebenlist U, Yewdell J, Kelly K. 1994. Gem: an induced, immediate early protein belonging to the Ras family. *Science* **265**:241–244 DOI 10.1126/science.7912851.

McMonnies CW. 2017. Glaucoma history and risk factors. *Journal of Optometry* **10**:71–78 DOI 10.1016/j.optom.2016.02.003.

Minckler DS, Bunt AH, Johanson GW. 1977. Orthograde and retrograde axoplasmic transport during acute ocular hypertension in the monkey. *Investigative Ophthalmology and Visual Science* **16**:426–441.

Oros Klein K, Oualkacha K, Lafond MH, Bhatnagar S, Tonin PN, Greenwood CM. 2016. Gene coexpression analyses differentiate networks associated with diverse cancers harboring TP53 missense or null mutations. *Frontiers in Genetics* **7**:137 DOI 10.3389/fgene.2016.00137.

Pillers DA, Weleber RG, Green DG, Rash SM, Dally GY, Howard PL, Powers MR, Hood DC, Chapman VM, Ray PN, Woodward WR. 1999. Effects of dystrophin isoforms on signal transduction through neural retina: genotype-phenotype analysis of duchenne muscular dystrophy mouse mutants. *Molecular Genetics and Metabolism* **66**:100–110 DOI 10.1006/mgme.1998.2784.

Quaranta L, Riva I, Gerardi C, Oddone F, Floriani I, Konstas AG. 2016. Quality of life in glaucoma: a review of the literature. *Advances in Therapy* **33**:959–981 DOI 10.1007/s12325-016-0333-6.

Ricotti V, Jagle H, Theodorou M, Moore AT, Muntoni F, Thompson DA. 2016. Ocular and neurodevelopmental features of Duchenne muscular dystrophy: a signature of dystrophin function in the central nervous system. *European Journal of Human Genetics* **24**:562–568 DOI 10.1038/ejhg.2015.135.

Riggs KA, Rohatgi A. 2019. HDL and reverse cholesterol transport biomarkers. *Methodist DeBakey Cardiovascular Journal* **15**:39–46 DOI 10.14797/mdcj-15-1-39.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**:77 DOI 10.1186/1471-2105-12-77.

Scott IC, Imamura Y, Pappano WN, Troedel JM, Recklies AD, Roughley PJ, Greenspan DS. 2000. Bone morphogenetic protein-1 processes probiglycan. *Journal of Biological Chemistry* **275**:30504–30511 DOI 10.1074/jbc.M004846200.

Sofroniew MV, Vinters HV. 2010. Astrocytes: biology and pathology. *Acta Neuropathologica* **119**:7–35 DOI 10.1007/s00401-009-0619-8.

Solinis MA, Del Pozo-Rodriguez A, Apaolaza PS, Rodriguez-Gascon A. 2015. Treatment of ocular disorders by gene therapy. *European Journal of Pharmaceutics and Biopharmaceutics* **95**:331–342 DOI 10.1016/j.ejpb.2014.12.022.

Stone EM, Fingert JH, Alward WL, Nguyen TD, Polansky JR, Sunden SL, Nishimura D, Clark AF, Nystuen A, Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC. 1997. Identification of a gene that causes primary open angle glaucoma. *Science* **275**:668–670 DOI 10.1126/science.275.5300.668.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**:15545–15550 DOI 10.1073/pnas.0506580102.

Syx D, Guillemyn B, Symoens S, Sousa AB, Medeira A, Whiteford M, Hermanns-Le T, Coucke PJ, De Paepe A, Malfait F. 2015. Defective proteolytic processing of fibrillar procollagens and prodecorin due to biallelic BMP1 mutations results in a severe, progressive form of osteogenesis imperfecta. *Journal of Bone and Mineral Research* **30**:1445–1456 DOI 10.1002/jbmr.2473.

Wang R, Wiggs JL. 2014. Common and rare genetic risk factors for glaucoma. *Cold Spring Harbor Perspectives in Biology* **4**:a017244 DOI 10.1101/cshperspect.a017244.

Ward Y, Spinelli B, Quon MJ, Chen H, Ikeda SR, Kelly K. 2004. Phosphorylation of critical serine residues in Gem separates cytoskeletal reorganization from down-regulation of calcium channel activity. *Molecular and Cellular Biology* **24**:651–661 DOI 10.1128/mcb.24.2.651-661.2004.

Xu Q, Wang Y, Zhu J, Zhao Y, Lin Y. 2018. Molecular characterization of GTP binding protein overexpressed in skeletal muscle (GEM) and its role in promoting adipogenesis in goat intramuscular preadipocytes. *Animal Biotechnology* **31**:1–8 DOI 10.1080/10495398.2018.1523796.

Ye N, Jiang N, Feng C, Wang F, Zhang H, Bai HX, Yang L, Su Y, Huang C, Wanggou S, Li X. 2019. Combined therapy sensitivity index based on a 13-gene signature predicts prognosis for IDH wild-type and MGMT promoter unmethylated glioblastoma patients. *Journal of Cancer* **10**:5536–5548 DOI 10.7150/jca.30614.

Yu H, Guan Z, Cuk K, Brenner H, Zhang Y. 2018. Circulating microRNA biomarkers for lung cancer detection in Western populations. *Cancer Medicine* **7**:4849–4862 DOI 10.1002/cam4.1782.

Zhang C, Peng L, Zhang Y, Liu Z, Li W, Chen S, Li G. 2017. The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Medical Oncology* **34**:101 DOI 10.1007/s12032-017-0963-9.

Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. 2015. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Briefings in Bioinformatics* **16**:291–303 DOI 10.1093/bib/bbu003.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**:1523 DOI 10.1038/s41467-019-09234-6.