Functional Annotations of Single-Nucleotide Polymorphism (SNP)-Based and Gene-Based Genome-Wide Association Studies Show Genes Affecting Keratitis Susceptibility

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

BCDEF 1 **Yue Xu***
BCDEF 2 **Xiao-Lin Yang***
BCD 1 **Xiao-Long Yang**
BC 1 **Ya-Ru Ren**
BC 1 **Xin-Yu Zhuang**
ADE 2 **Lei Zhang**
ADE 1 **Xiao-Feng Zhang**

1 Department of Ophthalmology, First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, P.R. China
2 Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Suzhou, Jiangsu, P.R. China

* Yue Xu and Xiao-Lin Yang contributed equally

Corresponding Authors: Xiao-Feng Zhang, e-mail: zhangxiaofeng@suda.edu.cn, Lei Zhang, e-mail: lzhang6@suda.edu.cn
Source of support: Departmental sources

Background: Keratitis is a complex condition in humans and is the second most common cause of legal blindness worldwide.
Material/Methods: To reveal the genomic *loci* underlying keratitis, we performed functional annotations of SNP-based and gene-based genome-wide association studies of keratitis in the UK Biobank (UKB) cohort with 337 199 subjects of European ancestry.
Results: The publicly available SNP-based association results showed a total of 34 SNPs, from 14 distinct *loci*, associated with keratitis in the UKB. Gene-based association analysis identified 2 significant genes: *IQCF3* ($p=2.0\times10^{-6}$) and *SOD3* ($p=2.0\times10^{-6}$). Thirty-two candidate genes were then prioritized using information from multiple sources. The overlap of *IQCF3* in these 2 analyses resulted in a total of 33 hub genes. Functional annotation of hub genes was performed and transcriptional factors of *IQCF3* and *SOD3* were predicted.
Conclusions: A total of 34 SNPs from 14 distinct *loci* were identified as being associated with keratitis, and 32 candidate genes were then prioritized. In addition, *IQCF3* and *SOD3* were identified by their p values through gene-based tests on the basis of individual SNP-based tests. The functional relationship between these suspect genes and keratitis suggest that *IQCF3* and *SOD3* are candidate genes underlying keratitis.

MeSH Keywords: **Genes, vif • Genome-Wide Association Study • Keratitis • Polymorphism, Single Nucleotide • Superoxide Dismutase**

Abbreviations: **SNP** – single-nucleotide polymorphism; **GWAS** – genome-wide association study; **VEGAS** – versatile gene-based association study; **QC** – quality control; **MAF** – minor allele frequency; **UTR** – untranslated regions; **CRV** – credible risk variant; **LD** – linkage disequilibrium; **eQTL** – expression quantitative trait *loci*; **GO** – Gene Ontology; **PPI** – protein–protein interaction

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/922710

🗐 3303   ▦ 4   📊 5   📖 50

Background

Keratitis is a condition in which the cornea becomes inflamed. Keratitis is the second most common cause of legal blindness worldwide, after cataracts [1]. Approximately 12.2% of all corneal transplantations are performed as the result of active infectious keratitis [2]. The prevalence of corneal diseases is 0.8%, whereas the prevalence of infectious keratitis is 0.15% [3]. The American National Ambulatory Care and Emergency Department databases show that episodes of keratitis and contact lens-related disorders cost about $175 million annually in direct healthcare expenditures [4], and that office and outpatient clinic visits require over 250 000 hours of clinician time annually [4].

Keratitis can arise from noninfectious causes (e.g., autoimmune and aseptic inflammation) or infectious agents (e.g., microbes such as bacteria, fungi, amebae, and viruses) [4,5]. Multiple genes have been identified as the basis for noninfectious keratitis. Autosomal dominant keratitis (ADK) is associated with the *PAX6* gene [6]. Mutations in *PAX6* were shown to underlie ADK when significant linkage was detected in both the *PAX6* region and ADK in 15 affected members in a single family [6]. Keratitis is one of the characteristics of tyrosinemia type II, which is an autosomal recessive disease [7]. Keratitis-ichthyosis-deafness (KID) syndrome is also characterized by vascularizing keratitis, and is caused by heterozygous missense mutations in the connexin-26 gene, *GJB2* [8]. Induction of permanent connexin-26 mutation expression can result in progressive keratitis in KID syndrome [8].

Genetics plays a role in infectious keratitis, from various types of infections and epithelial defects, as reported previously [9]. Even though commonly exposed to these pathogens in daily life, not everyone who contacts them is infected, perhaps because of differences in genetic susceptibility to infections. Multiple miRNAs, expressed on the ocular surface, function to regulate immunity on the ocular surface [10]. Resistance to *Pseudomonas aeruginosa* infection was higher in miR-183/96/182 cluster knockdown mice compared to wild-type mice [11,12] and the upregulation of miR-155 was reported to increase corneal susceptibility to *P. aeruginosa* [13]. Furthermore, a relationship between recurrent herpes simplex keratitis and host gene polymorphisms has been proposed [14].

Numerous alleles, isolated in many *loci*, lead to complex disease susceptibility through the contributions of common and rare alleles [15]. There has not yet been a description of the genetic risk factors for keratitis on a genome-wide scale, so there are still many keratitis candidate genes to be identified.

Because of the existence of allelic heterogeneity, genetic association analysis and the statistical power of gene-based association tests are improved through integrating diverse single-nucleotide polymorphisms (SNPs) into a single statistic [16,17]. Gene-based analysis using the versatile gene-based association study (VEGAS) approach, is widely used in modern genome-wide association studies. In the present study, a genome-wide association study of keratitis was performed using the large-scale UK Biobank (UKB) cohort. Both SNP-based and gene-based analyses were performed, and functional investigations of the identified genes were conducted.

Material and Methods

Ethics statement

This study was conducted in accordance with the Helsinki Declaration II and was approved by the Ethics Committee of First Affiliated Hospital of Soochow University.

Subjects

The sample consisted of 337 199 white participants of European ancestry from the UKB, including 239 cases and 336 960 controls. The diagnosis of keratitis was coded according to the World Health Organization's International Classification of Diseases (ICD) and Related Health Problems.

Quality control

This study had strict quality control (QC). Duplicate SNPs or SNPs deviating from Hardy-Weinberg equilibrium ($p < 10^{-6}$) or SNPs including a minor allele frequency (MAF) <0.01 or imputation r^2 <0.4 were excluded.

SNP-based analysis and enrichment analysis

The SNP-based association analysis was performed by Dr. Benjamin Neale's lab, and the summary results were generously made publicly available. The SNP association results were downloaded from the web link (*http://www.nealelab.is/uk-biobank*).

Fifty-two genomic features are available from a study by Finucane [18], including 3'- and 5'-untranslated regions (UTR). Every SNP was annotated into at least 1 genomic feature. A set of credible risk variants (CRVs) in each novel identified locus was identified. We selected the variants which were located in 500 kb around the lead SNPs and with p values within 2 orders of magnitude of the lead SNPs [19]. The enrichment of CRV in each genomic feature was then examined using the method described in Pei et al. [20]. Specifically, 100 1-Mb control regions were randomly sampled across the genome so that the distance between any identified region was at least 1 Mb.

A logistic regression model was utilized to examine the association of each feature, with the dependent variable of the status of being CRVs and the independent variable of the status with the feature. The model was adjusted for the dependence due to linkage disequilibrium (LD) within each region using robust variance estimation, clustering on region, using the R package "multiwayvcov".

Gene-based analyses using VEGAS

With the SNP-based genome-wide association statistics, a gene-based association analysis was performed with VEGAS [21]. P values of diverse SNPs in a gene region were combined into a gene-based score. Then, permutation-based testing, which simulated genotype data from a multivariate normal distribution, was used to account for LD. VEGAS has efficient computing power, considering the number of permutation simulations is adaptive. Individual SNP p values and information of LD structure were required as input by VEGAS [21].

Candidate gene prioritization

To prioritize potential candidate genes from SNP-based association results, 5 data sources were used to find the causality of a gene: i) genes nearest to the lead CRV (N); ii) genes including a mis-sense coding CRV (M); iii) genes related to 1 or more CRVs on the level of mRNA (cis-eQTL, Q); iv) genes having biological function with keratitis (B); and v) genes prioritized by summary data-based Mendelian randomization method (SMR) analysis (S).

Cis-eQTL was calculated by the GTEx (v7) project and obtained from the GTEx web portal (*www.gtexportal.org/*). SMR analysis is a kind of prioritization analysis integrating summary data from genome-wide association studies (GWAS). SMR analysis was performed to identify the correlation between the expression levels of genes and the trait due to pleiotropy effects or causality [22].

Module analysis

We used Cytoscape (version 3.6.1) bioinformatics software to analyze and visualize molecular interaction networks of identified hub genes [23]. Molecular Complex Detection (MCODE) (version 1.5.1), a plug-in of Cytoscape, was utilized to find densely connected regions with clustering in a given network according to topology [24]. MCODE was used to detect the most significant and densest module in the protein-protein interaction (PPI) network based on the following criteria: K-score=2, degree cut-off=4, max depth=100 and node score cut-off=0.20.

PPI network construction

The online database Search Tool for the Retrieval of Interacting Genes (STRING; version 10.0) was used to find the PPI network of the hub genes and to further analyze the functional interactions among proteins (*http://string-db.org*) [25], which elucidated the initiation or development mechanisms in keratitis. Confidence scores of interactions greater than 0.15 were regarded as statistically significant.

Functional enrichment analysis

Recognizing the roles of the hub genes in keratitis, Gene Ontology (GO) enrichment analysis, including biological processes (BP), cellular components (CC), and molecular functions (MF), were analyzed [26]. The online Database for Annotation, Visualization, and Integrated Discovery (DAVID; Version 6.8) was used to detect the function of development-related signaling pathways in keratitis (*http://david.ncifcrf.gov*) [27]. P values smaller than 0.05 were regarded as statistically significant. A bubble chart was used to illustrate GO enrichment.

Transcription factor network construction

A transcription factor network was constructed in the 2 genes (*IQCF3* and *SOD3*) using R software (Version 3.3.2). In line with hub genes, significant nodes were marked in red.

Results

This study consisted of 3 stages. In the first stage, a total of 34 SNPs from 14 distinct *loci* were identified from UKB, the enrichment of the variants at these newly identified *loci* was assessed in terms of diverse genomic features, and candidate genes were then prioritized. In the second stage, *IQCF3* and *SOD3* were identified by their p values through gene-based tests on the basis of individual SNP-based tests. In the third stage, 33 hub genes were selected, functional annotation of these hub genes was performed, and transcriptional factors of *IQCF3* and *SOD3* were predicted.

SNP-based analyses

After filtering, a total of 7 601 143 SNPs remained for subsequent analyses. A logarithmic quantile-quantile (QQ) plot of SNP-based association results is shown in Figure 1A. The obvious deviation in the tail of the distribution indicated true association signals. Additionally, a Manhattan plot of SNP-based associations is shown in Figure 1B. At the genome-wide significance level ($\alpha=5.0\times10^{-8}$), a total of 34 SNPs from 14 distinct *loci* were identified. Table 1 lists the major results of the novel identified *loci*.
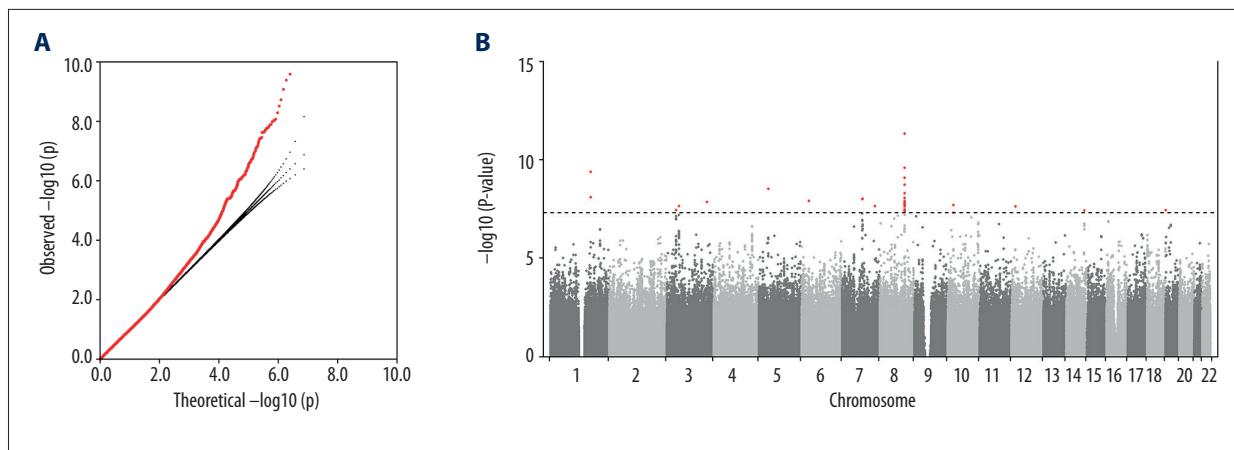
Figure 1. Logarithmic quantile-quantile (QQ) plot and Manhattan plot of individual SNP-based association tests. (**A**) QQ plot of individual SNP-based association tests. The x axis shows the theoretical $-\log_{10}$ (p value). The y axis shows the observed $-\log_{10}$ (p value). (**B**) Manhattan plot of individual SNP-based association tests. The x axis shows the chromosome numbers. The y axis shows the $-\log_{10}$ (p value). The dotted line is the gene-wide significance p value of 5×10^{-8}.

Table 1. Main results of the identified novel *loci* in UKB.

Lead SNP	CHR	POS	Locus	Beta	SE	P	N
rs3766652	1	171758370	1q24.3	9.24E-04	1.48E-04	**4.10E-10**	337199
rs6796010	3	169613751	3q26.2	2.63E-03	4.64E-04	**1.43E-08**	337199
rs9861402	3	39293844	3p22.2	6.37E-04	1.16E-04	**3.61E-08**	337199
rs139240291	3	51883928	3p21.2	1.21E-03	2.16E-04	**2.26E-08**	337199
rs2324638	4	25008851	4p15.2	5.14E-03	6.29E-04	**2.84E-16**	337199
rs150160025	5	40059390	5p13.1	1.17E-03	1.97E-04	**3.01E-09**	337199
rs6910636	6	30985793	6p21.33	3.39E-03	5.96E-04	**1.23E-08**	337199
rs144144964	7	138460918	7q34	1.04E-03	1.86E-04	**2.30E-08**	337199
rs76229182	7	86504797	7q21.12	1.03E-03	1.79E-04	**9.47E-09**	337199
rs149409568	8	105903422	8q22.3	1.18E-03	1.70E-04	**4.54E-12**	337199
rs138303599	10	23672519	10p12.2	7.77E-04	1.38E-04	**2.02E-08**	337199
rs117460695	12	15958922	12p12.3	1.19E-03	2.12E-04	**2.35E-08**	337199
rs748230	14	96240891	14q32.13	1.12E-03	2.03E-04	**3.91E-08**	337199
rs62129468	19	1944342	19p13.3	7.75E-04	1.41E-04	**3.67E-08**	337199

CHR – chromosome; POS – genomic position based on GRCH37 genome assembly; Beta – regression coefficient; SE – standard error of beta; N – sample size.

The enrichment of the variants at the novel *loci* was assessed for diverse genomic features. A set of CRVs was identified as SNPs with p values within 2 orders of magnitude of the lead SNP at each locus. Ninety-five CRVs were identified in this study. Then, the enrichment of these CRVs was evaluated for 52 genomic features, including transcription factor binding sites (TFBSs), transcription start sites (TSSs), 5'-untranslated regions (5'-UTRs) and 3'-UTRs, promoters, and histone mark active regions. At the level of Bonferroni-corrected significance ($\alpha=9.62 \times 10^{-4}$), CRVs were significantly enriched in 8 features: FANTOM5 enhancers ($p=5.00 \times 10^{-117}$), 5'-UTRs ($p=2.13 \times 10^{-109}$), conserved regions ($p=1.17 \times 10^{-85}$), TSSs ($p=3.10 \times 10^{-73}$), repressed regions ($p=4.41 \times 10^{-6}$), TFBSs ($p=5.26 \times 10^{-4}$), enhancers ($p=6.55 \times 10^{-4}$), and introns ($p=9.37 \times 10^{-4}$).
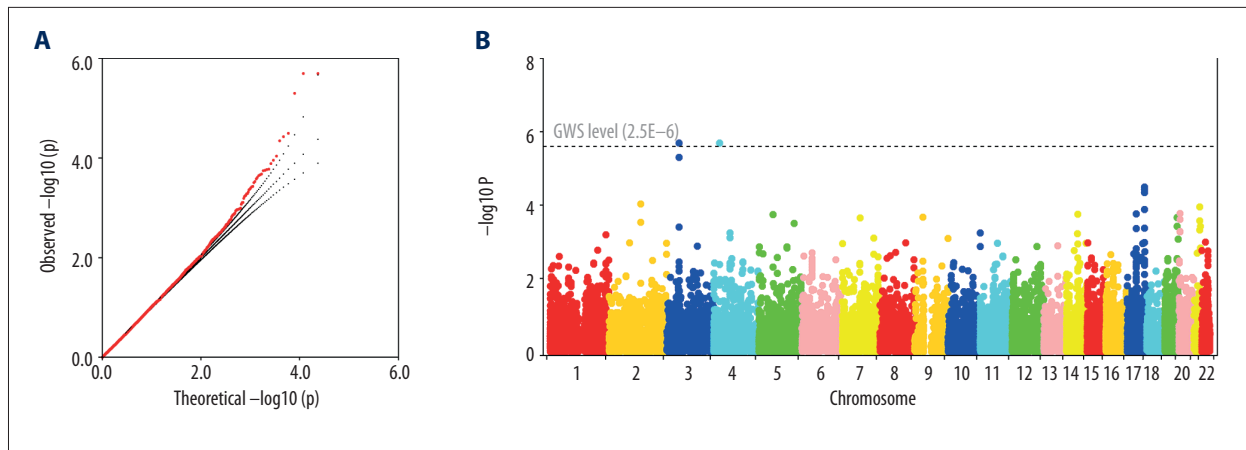
Figure 2. QQ plot and Manhattan plot of gene-based association tests. (**A**) QQ plot of gene-based association tests. The x axis shows the theoretical $-\log_{10}$ (p value). The y axis shows the observed $-\log_{10}$ (p value). (**B**) Manhattan plot of gene-based association tests. The x axis shows the chromosome numbers. The y axis shows the $-\log_{10}$ (p value). The dotted line is the gene-wide significance p value of 2.5×10^{-6}.

Table 2. Main results of the identified novel genes.

Locus	Gene	Start	Stop	No. SNPs	p Value	Top SNP	Top SNP – p value
3p21.2	*IQCF3*	51850898	51874874	42	2.00×10^{-6}	rs144319860	4.38×10^{-7}
4p15.2	*SOD3*	24787084	24812467	71	2.00×10^{-6}	rs800445	2.40×10^{-6}

Gene-based analyses

Gene-based tests were performed using VEGAS software on the basis of individual SNP-based tests. The gene-based association tests involved 23 779 genes. Bonferroni correction was utilized to determine the genome-wide significance level (GWAS, $0.05/23,779 = 2.10 \times 10^{-6}$). Figure 2A and 2B illustrate the QQ plot and Manhattan plot of the gene-based association tests, respectively. The marked deviation in the tail of the distribution in Figure 2A demonstrates a genuine connection between the genes and keratitis.

Table 2 shows that *IQCF3* and *SOD3* were identified according to their p values from the VEGAS analysis (p<2.50×10^{-6}). This study investigated whether allelic heterogeneity existed in the identified novel genes by studying SNP-level LD structure within each gene. The regional LD structure and association signals of the 2 novel genes are shown in Figure 3. The 42 SNPs in the *IQCF3* gene clustered into a single haplotype block (Figure 3A). The 71 SNPs in *SOD3* clustered into 2 haplotype blocks (Figure 3B), but the significant SNP association signals were only present in the larger block. Thus, the association signals of both novel genes might arise from a single source.

To prioritize candidate genes from the SNP-based association results, 5 data sources for CRV annotation were incorporated

to explore causality: i) the genes nearest to the lead CRV (N); ii) the genes including a mis-sense coding CRV (M); iii) the genes related to 1 or more CRVs at the level of mRNA (cis-eQTL, Q); iv) the genes having biological function with keratitis (B); and v) the genes prioritized by SMR analysis (S). With these strategies, 32 candidate genes were prioritized within these 14 new *loci* (Table 3). More than one kind of supporting evidence was found for some of these genes. For example, *CX3CR1* at 3p22.2 contained a mis-sense coding CRV of the lead CRV rs9861402 within this locus. *CX3CR1* was also prioritized due to biological function with keratitis and being the nearest to the lead CRV.

Hub-gene selection and analyses

As described above, *IQCF3* and *SOD3* were identified based on their p values found in gene-based analysis. In addition, 32 genes were seriously considered as candidate genes related to keratitis through candidate gene prioritization. Remarkably, *IQCF3* was identified by both analyses. In conclusion, a total of 33 hub genes were selected through these 2 analyses.

This study assessed the PPI networks among 33 hub genes. A network among the hub genes and their co-expressed genes was constructed visually (Figure 4A). After statistical selection, *SOD3*, *MGP*, *VARS2*, *RPL14*, *HIST4H4*, and *CSNK1G2* were considered to be densely connected nodes in the PPI network.
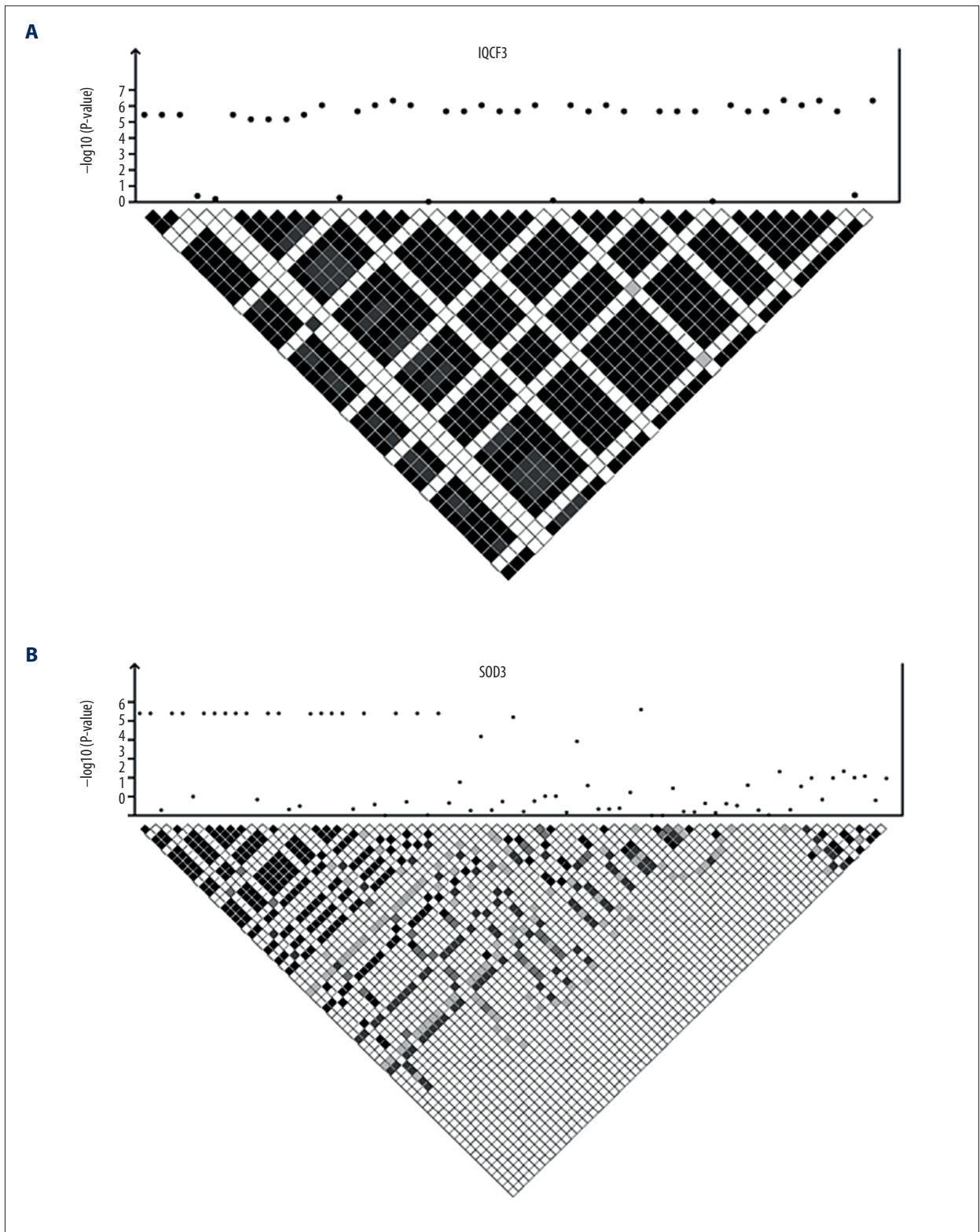
Figure 3. Regional plot of SNP associations of (**A**) *IQCF3* and (**B**) *SOD3*.

Table 3. Prioritized candidate genes at the identified novel *loci*.

Locus	Lead SNP	Candidate gene
1q24.3	rs3766652	*METTL13*(N), *PIGC*(S)
3q26.2	rs6796010	*RP11-379K17.4*(Q), *SEC62*(Q, B), *SAMD7*(N)
3p22.2	rs9861402	*CX3CR1*(M, N, B), *RPL14*(S)
3p21.2	rs139240291	*ABHD14A*(Q), *IQCF3*(Q), *RP11-314A5.3*(Q), *PPM1M*(Q), *IQCF2*(N)
4p15.2	rs2324638	*LGI2*(N), *ANAPC4*(S), *CCDC149*(S)
5p13.1	rs150160025	*BC026261*(N)
6p21.33	rs6910636	*HLA-B*(Q, B), *MUC21*(Q, B), *XXbac-BPG248L24.12*(Q), *VARS2*(Q), *MUC22*(N)
7q34	rs144144964	*ATP6V0A4*(N)
7q21.12	rs76229182	*KIAA1324L*(N)
8q22.3	rs149409568	*Mir_584*(N)
10p12.2	rs138303599	*C10orf67*(N)
12p12.3	rs117460695	*EPS8*(N), *HIST4H4*(S), *C12orf60*(S), *MGP*(S)
14q32.13	rs748230	*BX247990*(N)
19p13.3	rs62129468	*CSNK1G2*(N), *MKNK2*(S)

Genes are prioritized as follow: gene nearest to the lead credible risk variant (CRV) (N); gene containing a mis-sense CRV (M); gene with mRNA levels in association with one or more CRVs (cis-eQTL, Q); gene with biology function with keratitis (B); and gene prioritized by SMR analysis (S).
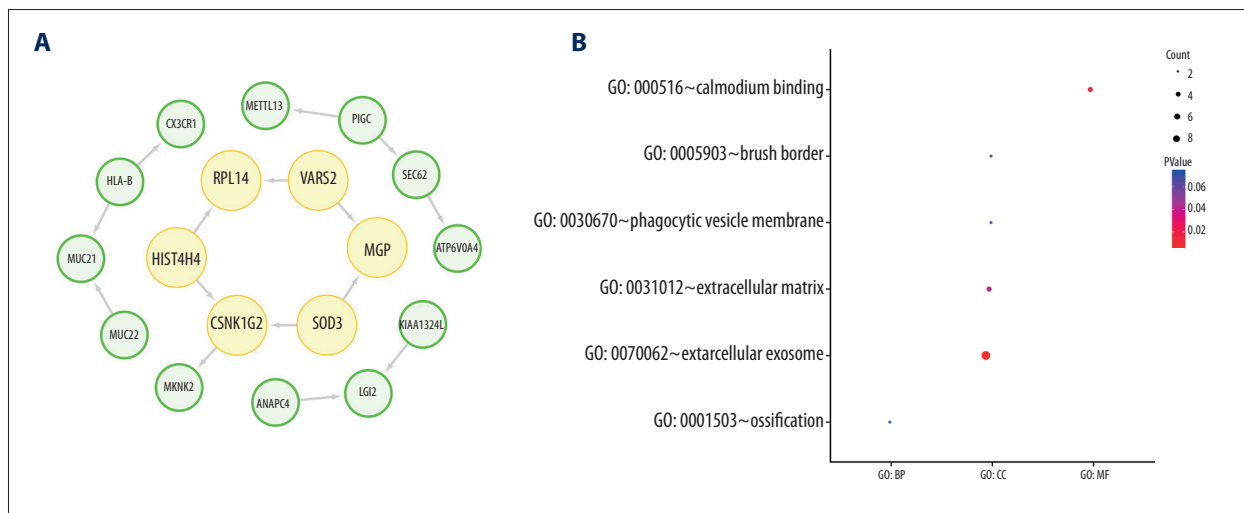

Figure 4. Functional annotations and predicted signaling pathways. (**A**) The PPI network of the 33 hub genes and their coexpression genes was constructed. Six genes (*SOD3, MGP, VARS2, RPL14, HIST4H4,* and *CSNK1G2*) in yellow were considered densely connected nodes in the PPI network. (**B**) Functional enrichment analyses of hub genes were performed by DAVID in a bubble chart.

Functional enrichment analyses of hub genes were performed in bubble charts by DAVID (Figure 4B). Through functional analyses of the 33 hub genes, enrichment profiles indicated that the hub genes were mainly enriched in the extracellular exosome, calmodulin binding, and extracellular matrix (Supplementary Table 1). GO analysis suggested that changes in the cellular components of the hub genes were significantly enriched in the extracellular exosome and extracellular matrix. Changes in

molecular function of the hub genes were mostly enriched in calmodulin binding.

Transcription factor network construction

The transcription factor network was predicted for the 2 genes (*IQCF3* and *SOD3*), as shown in Figure 5. In connection with these 2 hub genes, significant nodes are marked in red.
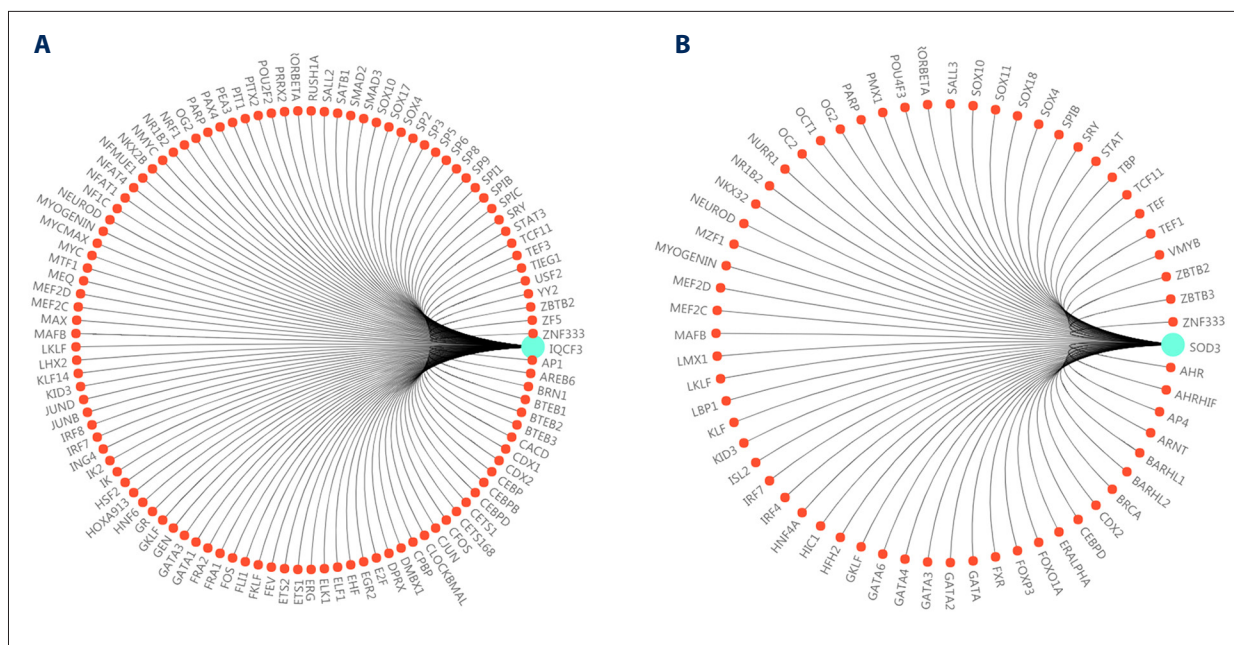
Figure 5. Transcription factor network of *IQCF3* and *SOD3*. Transcription factor network was predicted in *IQCF3* (**A**) and *SOD3* (**B**). Significant nodes were marked in red in line with these 2 genes.

Discussion

A total of 34 SNPs from 14 distinct *loci* were identified as associated with keratitis in samples from the UKB. In addition, *IQCF3* and *SOD3* were identified by their p values through gene-based tests on the basis of individual SNP-based tests. Thirty-two candidate genes were then prioritized using multiple sources of information. Four of these candidate genes (*CX3CR1*, *SEC62*, *HLA-B*, and *MUC21*) were found to be closely associated with keratitis using more than 1 source of information. The overlapping of *IQCF3* in both the gene-based association study and candidate gene prioritization resulted in a total of 33 hub genes. Functional annotation of the hub genes was performed and transcriptional factors for *IQCF3* and *SOD3* were predicted.

IQ motif containing F3 (*IQCF3*) is located at 3p21.2. The molecular function of IQCF3 is calmodulin binding. *IQCF3* was found in both the gene-based association study and in candidate gene prioritization analysis. Few previous studies have illustrated the relationship between keratitis and *IQCF3*. *IQCF3* appears to be a novel gene underlying keratitis and may provide a potential target for the treatment of keratitis.

Superoxide dismutase 3 (*SOD3*) is located at 4p15.2. *SOD3* encodes extracellular superoxide dismutase. SOD3 is an antioxidant enzyme that catalyzes the dismutation of 2 superoxide radicals into hydrogen peroxide and oxygen, which are thought to play major roles in changes in vascular structure and function in pathophysiology [28]. *SOD3* exists as a copper- and zinc-containing tetramer, and is synthesized containing a signal peptide that directs this enzyme exclusively to extracellular spaces [29]. Reactive oxygen species (ROS)-induced oxidative injury is involved in the pathogenesis of fungal keratitis via the p38 MAPK pathway, and *SOD3* is one of the important molecules in biological oxidation in the cornea [30]. *SOD3* protects the cornea by converting ROS into less-reactive products [31,32]. It appears that *SOD3* could be a therapeutic anti-inflammatory agent due to its strong antioxidant activity and immune-regulatory function [33]. Some researchers have synthesized a SOD derivative to reduce tissue injury in keratitis caused by ROS [34].

C-X3-C motif chemokine receptor 1 (*CX3CR1*) is located at 3p22.2. Protein CX3CR1, encoded by *CX3CR1*, is the sole receptor for fractalkine, which is a transmembrane protein and chemokine involved in the adhesion and migration of leukocytes [35]. It is well-recognized that the peripheral corneal and limbal epithelia contain a resident population of major histocompatibility complex (MHC) class II+ intraepithelial dendritic cells (DCs), sometimes called Langerhans cells (LCs). *CX3CR1* was shown to play a role in the normal recruitment of MHC class II+ putative DCs into the corneal epithelium [36]. The association between nerves and resident immune cells in normal eyes was reported to be partly mediated by *CX3CR1* signaling in response to central corneal epithelial injury [37]. Other research indicated that accumulation of CX3CR1-positive macrophages can dampen alkali-induced corneal neovascularization by producing antiangiogenic factors such as TSP-1 and ADAMTS-1 [38].

SEC62 Homolog, preprotein translocation factor (*SEC62*), located at 3q26.2, encodes translocation protein SEC62 [39,40]. It participates in post-translational protein translocation into the endoplasmic reticulum (ER) and may also be involved in the backward transport of ER proteins [40]. Dilated ER is considered a general characteristic of keratitis in acute and chronic inflammation [41]. The change becomes more pronounced as the endothelial cells become detached from the Descemet's membrane [41].

Major histocompatibility complex, Class I, B (*HLA-B*), located at 6p21.33, belongs to a gene family called the human leukocyte antigen (HLA) complex class I. HLA is the human version of major histocompatibility complex (MHC). The HLA complex assists the immune system in distinguishing proteins made by foreign invaders, such as viruses and bacteria, from the body's own proteins. The *HLA-B* genes show various polymorphic forms, one of which is *HLA-B5*. *HLA-B5* was found to be related to recurrent herpetic keratitis and recurrent stromal keratitis [42].

Mucin 21 (*MUC21*) is also located at 6p21.33. This gene encodes a large membrane-bound glycoprotein that is a member of the mucin family. Mucins play an important role not only in intracellular signaling, but also in forming protective mucous barriers on epithelial surfaces. The transmembrane mucin MUC21 has been shown to be one of the first lines of defense by mucosal surfaces to limit the access of viruses into the host [43,44]. It has been suggested that transmembrane mucins act as space-filling molecules in glycocalyx and are capable of forming dense network structures that selectively limit viral infection [45]. In the corneal epithelium, the carbohydrate-dependent interactions of galectin-3 with transmembrane mucins help maintain the epithelial barrier [46].

The relationships between keratitis and the other genes identified here have not been determined previously. However, the results of the present study suggest that *SAMD7* is essential for rod photoreceptor cells and is associated with retinitis pigmentosa [47,48]. The downregulation of *PPM1M* significantly inhibits HSV-1 infection, which is associated with herpetic stromal keratitis [49]. A study investigating the molecular mechanisms underlying activation of cell death pathways in human limbal epithelial cell cultures suggested that *HIST4H4* is a significantly upregulated gene during hypothermic storage and that it is involved in a functional network highly associated with cell death, necrosis, and transcription of RNA [50]. These genes were also identified as candidate genes in our study and may be novel susceptible genes underlying keratitis, but this needs further study.

Keratitis is a complex clinical trait that can arise from infectious agents or noninfectious causes. Even though confronted by these pathogens everywhere in daily life, not everyone who is contacted becomes infected, perhaps due to differences in genetic susceptibility to infections. Based on the PPI network and functional annotations of these hub genes, targeted therapies or drug development for keratitis should focus on these genes and pathways.

Although they are potentially related to keratitis, it is unlikely that all of these genes are causally related, as they were identified by statistical association rather than biological validation. In addition, this kind of research is influenced by the structure of the LD, and is generally considered to be indirect gene mapping. The identification of causal genes associated with keratitis still awaits further functional studies.

Conclusions

In conclusion, a total of 34 SNPs from 14 distinct *loci* were identified as associated with keratitis from UKB. Additionally, *IQCF3* and *SOD3* were identified by their p values through gene-based tests on the basis of individual SNP-based tests. Thirty-two candidate genes were then prioritized. The overlapping of *IQCF3* in the gene-based association study and candidate gene prioritization resulted in a total of 33 hub genes. Functional annotation of hub genes was performed and transcriptional factors of *IQCF3* and *SOD3* were predicted. These results, together with the known functional relationship between these genes and keratitis, suggest that *IQCF3* and *SOD3* are candidate genes underlying keratitis. These genes, with good predictive power, may provide a novel approach for selecting high-risk keratitis patients, as well as acting as biomarkers for monitoring, treatments, and drug development.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of First Affiliated Hospital of Soochow University.

Availability of data and material

The datasets analyzed in our study are available from the corresponding author on reasonable request.

Conflict of interests

None.

CLINICAL RESEARCH

Supplementary Data

Supplementary Table 1. GO enrichment analysis of hub genes.

Term	Description	Count in gene set	P value
GO: 0070062	Extracellular exosome	9	7.00E-03
GO: 0005516	Calmodulin binding	3	1.70E-02
GO: 0031012	Extracellular matrix	3	4.12E-02

GO – Gene Ontology.

References:

1. Al-Mujaini A, Al-Kharusi N, Thakral A, Wali UK: Bacterial keratitis: Perspective on epidemiology, Clinico-Pathogenesis, diagnosis and treatment. Sultan Qaboos Univ Med J, 2009; 9(2): 184–95

2. Sharma S: Keratitis. Biosci Rep, 2001; 21(4): 419–44

3. Cao J, Yang Y, Yang W et al: Prevalence of infectious keratitis in Central China. BMC Ophthalmol, 2014; 14: 43

4. Collier SA, Gronostaj MP, MacGurn AK et al: Estimated burden of keratitis – United States, 2010. Morb Mortal Wkly Rep, 2014; 63(45): 1027–30

5. Srinivasan M, Mascarenhas J, Prashanth CN: Distinguishing infective versus noninfective keratitis. Indian J Ophthalmol, 2008; 56(3): 203–7

6. Mirzayans F, Pearce WG, MacDonald IM, Walter MA: Mutation of the PAX6 gene in patients with autosomal dominant keratitis. Am J Hum Genet, 1995; 57(3): 539–48

7. Natt E, Kida K, Odievre M et al: Point mutations in the tyrosine aminotransferase gene in tyrosinemia type II. Proc Natl Acad Sci USA, 1992; 89(19): 9297–301

8. Richard G, Rouan F, Willoughby CE et al: Missense mutations in GJB2 encoding Connexin-26 cause the ectodermal dysplasia keratitis-ichthyosis-deafness syndrome. Am J Hum Genet, 2002; 70(5): 1341–48

9. Lakhundi S, Siddiqui R, Khan NA: Pathogenesis of microbial keratitis. Microb Pathog, 2017; 104: 97–109

10. Hazlett L, Suvas S, McClellan S, Ekanayaka S: Challenges of corneal infections. Expert Rev Ophthalmol, 2016; 11(4): 285–97

11. Lumayag S, Haldin CE, Corbett NJ et al: Inactivation of the microRNA-183/96/182 cluster results in syndromic retinal degeneration. Proc Natl Acad Sci USA, 2013; 110(6): E507–16

12. Muraleedharan CK, McClellan SA, Barrett RP et al: Inactivation of the miR-183/96/182 cluster decreases the severity of Pseudomonas aeruginosa-induced keratitis. Investig Ophthalmol Vis Sci, 2016; 57(4): 1506–17

13. Yang K, Wu M, Li M et al: MiR-155 suppresses bacterial clearance in *Pseudomonas aeruginosa*-induced keratitis by targeting Rheb. J Infect Dis, 2014; 210(1): 89–98

14. Borivoje S, Svetlana S, Milan HM et al: IL28B genetic variations in patients with recurrent herpes simplex keratitis. Medicina (Kaunas), 2019; 55(10): pii: E642

15. Yang J, Manolio TA, Pasquale LR et al: Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet, 2011; 43(6): 519–25

16. Huang H, Chanda P, Alonso A et al: Gene-based tests of association. PLoS Genet, 2011; 7(7): e1002177

17. Li MX, Gui HS, Kwan JSH, Sham PC: GATES: A rapid and powerful gene-based association test using extended Simes procedure. Am J Hum Genet, 2011; 88(3): 283–93

18. Finucane HK, Bulik-Sullivan B, Gusev A et al: Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet, 2015; 47(11): 1228–35

19. Michailidou K, Lindstrom S, Dennis J et al: Association analysis identifies 65 new breast cancer risk *loci*. Nature, 2017; 551(7678): 92–94

20. Pei YF, Liu L, Liu T Le et al: Joint association analysis identified 18 new *loci* for bone mineral density. J Bone Miner Res, 2019; 34(6): 1086–94

21. Liu JZ, McRae AF, Nyholt DR et al: A versatile gene-based test for genome-wide association studies. Am J Hum Genet, 2010; 87(1): 139–45

22. Zhu Z, Zhang F, Hu H et al: Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet, 2016; 48(5): 481–87

23. Smoot ME, Ono K, Ruscheinski J et al: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics, 2011; 27(3): 431–32

24. Bandettini WP, Wilson JR, Leung SW et al: MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: A clinical validation study. J Cardiovasc Magn Reson, 2012; 14(1): 83

25. Franceschini A, Szklarczyk D, Frankild S et al: STRING v9.1: Protein-protein interaction networks,with increased coverage and integration. Nucleic Acids Res, 2013; 41: D808–15

26. Ashburner M, Ball CA, Blake JA et al: Gene Ontology: Tool for the unification of biology. Nat Genet, 2000; 25(1): 25–29

27. Huang DW, Sherman BT, Tan Q et al: The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol, 2007; 8(9): R183

28. Faraci FM, Didion SP: Vascular protection: Superoxide dismutase isoforms in the vessel wall. Arterioscler Thromb Vasc Biol, 2004; 24(8): 1367–73

29. Zelko IN, Mariani TJ, Folz RJ: Superoxide dismutase multigene family: A comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. Free Radic Biol Med, 2002; 33(3): 337–49

30. Hua X, Chi W, Su L et al: ROS-induced oxidative injury involved in pathogenesis of fungal keratitis via p38 MAPK activation. Sci Rep, 2017; 7(1): 10421

31. Leal SM, Vareechon C, Cowden S et al: Fungal antioxidant pathways promote survival against neutrophils during infection. J Clin Invest, 2012; 122(7): 2482–98

32. Chen Y, Mehta G, Vasiliou V: Antioxidant defenses in the ocular surface. Ocul Surf, 2009; 7(4): 176–85

33. Gao B, Flores SC, Leff JA et al: Synthesis and anti-inflammatory activity of a chimeric recombinant superoxide dismutase: SOD2/3. Am J Physiol Lung Cell Mol Physiol, 2003; 284(6): L917–25

34. Ando E, Ando Y, Inoue M et al: Inhibition of corneal inflammation by an acylated superoxide dismutase derivative. Investig Ophthalmol Vis Sci, 1990; 31(10): 1963–67

35. Dorgham K, Ghadiri A, Hermand P et al: An engineered CX3CR1 antagonist endowed with anti-inflammatory activity. J Leukoc Biol, 2009; 86(4): 903–11

36. Chinnery HR, Ruitenberg MJ, Plant GW et al: The chemokine receptor CX3CR1 mediates homing of MHC class II – positive cells to the normal mouse corneal epithelium. Invest Ophthalmol Vis Sci, 2007; 48(4): 1568–74

37. Seyed-Razavi Y, Chinnery HR, McMenamin PG: A novel association between resident tissue macrophages and nerves in the peripheral stroma of the murine cornea. Investig Opthalmology Vis Sci, 2014; 55(3): 1313–20

38. Lu P, Li L, Kuno K et al: Protective roles of the fractalkine-CX3CL1-CX3CR1 interactions in alkali-induced corneal neovascularization through enhanced antiangiogenic factor expression. J Immunol, 2008; 180(6): 4283–91

39. Daimon M, Susa S, Suzuki K et al: Identification of a human cDNA homologue to the Drosophila translocation protein 1 (Dtrp1). Biochem Biophys Res Commun, 1997; 230(1): 100–4

40. Meyer HA, Grau H, Kraft R et al: Mammalian Sec61 is associated with Sec62 and Sec63. J Biol Chem, 2000; 275(19): 14550–57

41. Elgebaly SA, Forouhar F, Gillies C et al: Leukocyte-mediated injury to corneal endothelial cells. A model of tissue injury. Am J Pathol, 1984; 116(3): 407–16

42. Jensen KB, Nissen SH, Svejgaard A et al: Recurrent herpetic keratitis and HlA antigens. Acta Ophthalmol, 1984; 62(1): 61–68

43. Woodward AM, Mauris J, Argueso P: Binding of transmembrane mucins to Galectin-3 limits herpesvirus 1 infection of human corneal keratinocytes. J Virol, 2013; 87(10): 5841–47

44. Pickles RJ: Physical and biological barriers to viral vector-mediated delivery of genes to the airway epithelium. Proc Am Thorac Soc, 2004; 1(4): 302–8

45. Kesimer M, Ehre C, Burns KA et al: Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. Mucosal Immunol, 2013; 6(2): 379–92

46. Argüeso P, Guzman-Aranguez A, Mantelli F et al: Association of cell surface mucins with galectin-3 contributes to the ocular surface epithelial barrier. J Biol Chem, 2009; 284(34): 23037–45

47. Omori Y, Kubo S, Kon T et al: Samd7 is a cell type-specific PRC1 component essential for establishing retinal rod photoreceptor identity. Proc Natl Acad Sci, 2017; 114(39): E8264–73

48. Hlawatsch J, Karlstetter M, Aslanidis A et al: Sterile alpha motif containing 7 (Samd7) is a novel Crx-regulated transcriptional repressor in the retina. PLoS One, 2013; 8(4): e60633

49. Yue L, Guo S, Zhang Y et al: The modulation of phosphatase expression impacts the proliferation efficiency of HSV-1 in infected astrocytes. PLoS One, 2013; 8(11): e79648

50. Paaske Utheim T, Salvanos P, Aass Utheim Ø et al: Transcriptome analysis of cultured limbal epithelial cells on an intact amniotic membrane following hypothermic storage in optisol-GS. J Funct Biomater, 2016; 7(1): 4