Screening of single nucleotide polymorphisms among fuchs' endothelial corneal dystrophy subjects in Malaysia

Ker Hsin Ng[1*], Visvaraja Subrayan[2], Vasudevan Ramachandran[3,4*] and Fazliana Ismail[2*]

Abstract

Background: The pathophysiology underlying Fuchs' Endothelial Corneal Dystrophy (FECD), especially in older individuals, remains unclear, with a genetic predisposition being reported as the single best predictor of the disease. Genetic studies have shown that several genes in various loci such as *COL8A2, SLC4A11, TCF8/ZEB1* and *TCF4* are associated with FECD in different populations and ethnicities. A case–control study was conducted to determine the association between genetic variants and FECD in a tertiary care setting in Malaysia. A total number of 12 patients with clinically diagnosed FECD and 12 age, gender and race matched control subjects were recruited. Extracted genomic DNA were genotyped using Infinium Global Screening Array (GSA)-24 version 1.0 BeadChip with iScan high-throughput system. Illumina GenomeStudio 2.0 Data Analysis and PLINK version 1.9 software were used to perform association tests and determine the distribution of obtained variants among the cases and controls.

Results: A significant novel genetic variant, *rs11626651*, a variant of the *LOC105370676* gene or known as the *LINC02320* gene, located at chromosome 14, has been identified as a suggestive association with FECD ($p < 5 \times 10^{-6}$). Further analysis in this study suggested that candidate genes such as *COL8A2, ZEB1/TCF8, TCF4* and *SLC4A11* had no significant associations with FECD.

Conclusions: The discovery of a novel variant may influence the underlying pathogenic basis of FECD in Malaysia. The current study is the first genetic study on FECD to use Infinium GSA. It is the first comprehensive report in Malaysia to provide genetic information of potential relevance to FECD, which may pave the way for new therapeutic strategies in the future. A detailed analysis with a larger sample size is recommended for further evaluation.

Keywords: Genetic polymorphism, Illumina, Fuchs, Endothelial corneal dystrophy

Background

Fuchs' Endothelial Corneal Dystrophy (FECD) is a bilateral, asymmetric, slowly progressive disorder affecting the corneal endothelium. It is characterized by the

*Correspondence: alene.khng@gmail.com; vasuphd@gmail.com; ifazliana@gmail.com
[1] Department of Ophthalmology, University Malaya Medical Centre, Kuala Lumpur, Malaysia
[2] Department of Ophthalmology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[3] Malaysian Research Institute of Ageing, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
Full list of author information is available at the end of the article

formation of guttata, resulting in thickening of Descemet's membrane and loss of corneal endothelial cells. This dysfunction leads to corneal decompensation with subsequent visual impairment [1]. FECD affects approximately 4% of Americans [2], 4.5 to 9% of Europeans [3], and 3.8 to 8% of Asians [4].

Surgery is a widely known treatment option in FECD. By understanding the pathogenesis of this multifactorial disorder, new therapies and preventive interventions may be discovered. Environmental factors such as oxidative stress and accumulation of advanced glycation end-products play key roles in FECD pathogenesis [5–8]. Smoking has been reported to increase the risk

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 2 of 11

of advanced disease, possibly due to oxidative stress [9]. Chronic exposure to ultraviolet (UV) radiation like

UV-A activates antioxidant defense, inducing apoptosis in corneal endothelial cells [6, 10]. The pathophysiology of FECD remains unclear, especially in older individuals, with a genetic predisposition being reported as the single best predictor of the disease. The susceptibility of genes to mutations can vary in different ethnicities. The discovery of single nucleotide polymorphisms (SNPs) and genetic variants in FECD is helpful as it may influence future therapies of this disease. Genetic studies have replicated numerous quantitative trait loci throughout the genome and provide considerable evidences that genes play essential roles in the genetic architecture of ocular diseases [11]. Although the genetics remain unclear, several studies have supported the existence of genetic variants that increase the risk of FECD [12–14]. The involvement of collagen, type VIII, alpha 2 (*COL8A2*) gene in early-onset FECD has been found [15]. In the late-onset form of FECD, candidate gene studies and genome-wide linkage analyses showed that multiple genetic loci had been mapped on chromosomes 13, 18, 5 and 9, respectively [16–19]. Genetic variants found in zinc finger E-box binding homeobox 1 (*ZEB1*) or transcription factor 8 (*TCF8*), the solute carrier family 4 member 11 (*SLC4A11*), and transcription factor 4 (*TCF4*) genes have been established to play important roles in late-onset forms of FECD [12, 19, 20].

In the past two decades, many efforts have been made to discover previously unknown biological insights in various ocular diseases other than FECD, such as age-related macular degeneration, glaucoma, keratoconus, diabetic retinopathy and retinitis pigmentosa [21]. The advent of next-generation sequencing (NGS) can identify genetic variations in ocular diseases and lay the groundwork for the era of personalized medicine [11, 22, 23]. The genetics of FECD can be determined well by using bioinformatics tools to identify the candidate genes and merge multiple genetic data sets from individuals with this disease.

To date, there is a lack of information available on the genetic susceptibility of FECD among Malaysians. Hence, we conducted this study to analyze the association between genetic variants with FECD in a tertiary care setting. Infinium Global Screening Array (GSA)-24 version 1.0 BeadChip with iScan system [24, 25] was used for detection of the variants in FECD. The Infinium GSA is an advanced genotyping array that enables high content flexibility, high-throughput capacity, and genotyping accuracy [26].

This genotypic approach will provide a better understanding in identifying the underlying genetic defects.

The possibility of new medical interventions or gene therapies may be enhanced in the future.

Methods

Participants and data collection

The subjects were recruited from the Ophthalmology clinic of a tertiary care hospital in Malaysia from April 2016 until March 2017. A total of 12 patients with FECD and 12 age, gender and race matched individuals without FECD were recruited based on the inclusion and exclusion criteria.

Inclusion criteria

Patients with FECD who were above 40 years old from the three main Malaysian ethnicities (Malay, Chinese and Indian) were included. The diagnosis of FECD was principally based on clinical signs and symptoms. The presence of guttae was identified using slit-lamp biomicroscope via specular reflection. Specular microscopy was used to aid in demonstrating polymegathism and pleomorphism of the corneal endothelial cells, as well as measuring endothelial cell density.

Exclusion criteria

Patients who were diagnosed with pseudophakic bullous keratopathy with no evidence of FECD in the other eye, previous ocular trauma, history of intraocular inflammation and other causes of corneal decompensation like corneal inflammation, posterior polymorphous dystrophy, interstitial keratitis, iridocorneal endothelial syndrome, and herpetic keratitis were excluded.

Individuals who were free of any ocular disorders except cataracts at the time of ascertainment were recruited as control subjects. All patients who fulfilled the selection criteria were explained the nature of the study and written consents were obtained. They underwent a complete ophthalmic evaluation including a detailed medical, ocular and family history.

Sampling methods

5 ml (mL) of venous blood was drawn from the peripheral veins of all the subjects. Genomic DNA was extracted from the whole blood using a Qiagen kit (Qiagen Operation, Hilden, Germany). DNA quantification and qualification were done. Genomic DNA samples were genotyped using Infinium GSA on a bead array, following the manufacturer's instructions. The array has a SNP panel that includes 642,824 markers and uses a 24-sample Infinium high-throughput screening (HTS) format. All the samples were screened using the GSA Kit, which includes convenient packaging of BeadChips and various reagents. The standard protocol was followed, using a streamlined Infinium workflow for a minimum of

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 3 of 11

3 days. The protocol includes amplification, fragmentation, precipitation, resuspension, hybridization, staining, extension, and labelling to detect the genetic variants. The Illumina iScan® system was used to scan the BeadChip and record high-resolution images for analysis.

Data and association analysis of genetic variants

The genotyping data were extracted with the GenomeStudio Genotyping Module and displayed in the Illumina Genome Viewer. Illumina Chromosome Browser was used to visualize the data of all samples. The data were normalized and recorded in different file formats.

A range of standard case–control association analyses were performed with PLINK software version 1.9. The frequencies of the alleles were calculated and Chi-square test was used to compare their distributions between cases and controls. The results were used to generate Multidimensional Scaling (MDS) Plot, Quantile–quantile (Q-Q) Plot and Manhattan Plot.

The PLINK files obtained from Genome Studio were first subjected to quality control. A quality measure, known as a "GenCall score", was determined for each genotype. Samples with low genotyping efficiency or call rate were eliminated from further analysis. A recommended threshold is 98–99% efficiency, after first removing markers with a low genotype call rate across samples. Separate quality filtering was performed for variants in autosomal and sex chromosomes. Parameters that were used to filter variants in autosomal chromosomes were: including genetic variants with a threshold of 95% genotyping rate (geno) > 0.95, minor allele frequency (MAF) > 0.01, and excluding genetic variants that failed Hardy–Weinberg equilibrium (HWE) with corresponding $p < 0.000001$.

Markers in the X chromosome required different quality filtering steps. The steps performed were: removing heterozygous genotypes (likely due to genotyping errors), performing sex check (X chromosome homozygosity estimate), selecting only female control samples that passed sex check to identify variants that had HWE with $p < 0.000001$, and lastly, excluding all individuals that did not pass sex check, genetic variants with MAF < 0.01 and high missing call rates.

The biological interpretation and function of the proteins were analyzed through the Clusters of Orthologous Groups (COG) database. Eukaryotic Orthologous Groups (KOG), a eukaryote-specific version of the COG tool, was used to identify the ortholog and paralog proteins.

Statistical analysis

IBM Statistical Package for Social Sciences (SPSS) Version 24.0 (SPSS Inc., Chicago, IL, USA) software was used to analyze the socio-demographic data. Continuous variables were expressed as mean ± standard deviation and categorical variables. Chi-square test was used to compare genders, races, smoking status and systemic illnesses (such as hypertension and diabetes mellitus) between the cases and controls. The student's t-test was used to calculate the mean ages between and within the case–control groups. A p-value of < 0.05 was considered to be statistically significant.

Results

Demographic data

A total of 24 participants (12 cases and 12 controls) were recruited in this study (Table 1). They were all aged between 53 and 86 years with a mean of 68.25 years. The mean ages of the cases (68.67 ± 7.67) were similar to the controls (67.83 ± 8.57), with no statistical difference. Among the cases, there were 7 (58.3%) females and 5 (41.7%) males who were matched with the controls. Genders and ethnicities were equally distributed among the cases and controls. There were no significant differences between the cases and controls who had diabetes, hypertension, and history of smoking ($p > 0.05$). Family histories of FECD were denied in both groups.

Variant analysis

Out of 616,835 genetic variants in autosomal chromosomes, 16,232 variants were filtered out due to high missing call rates. 231,362 variants were filtered out with MAF less than 1% from the data screened for the genetic variants of all subjects. However, none of these variants failed HWE. The remaining 369,241 variants passed quality control steps. A total of 16,919 variants were found

Table 1 Demographic characteristics of the study subjects

Parameters	Case (n = 12)	Control (n = 12)	*p*-value[a]
Age (years)	68.67 ± 7.67	67.83 ± 8.57	0.804[b]
Race			
Malay	5 (41.7)	5 (41.7)	1.000
Chinese	4 (33.3)	4 (33.3)	
Indians	3 (25.0)	3 (25.0)	
Gender			
Male	5 (41.7)	5 (41.7)	1.000
Female	7 (58.3)	7 (58.3)	
Smoking	4 (33.3)	3 (25.0)	0.653
Diabetes	4 (33.3)	3 (25.0)	0.653
Hypertension	9 (75.0)	8 (66.7)	0.653
Family history of FECD	0	0	-

The age was expressed in Means ± SD. Data was presented as percentages in parentheses

[a] Chi-square test; [b] Student's t-test; $p > 0.05$

FECD: Fuchs' Endothelial Corneal Dystrophy

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 4 of 11

in the X chromosome. Among these variants, 7645 variants were marked as heterozygous haploid genotypes in males (likely due to genotyping errors) and subsequently removed. A total of 4199 variants did not pass the MAF threshold. 1022 variants were removed due to high missing call rates. However, none of the variants were removed by the HWE test, leaving a total of 4053 variants in the X chromosome for the association test.

Filtering culled the total number of variants from 633,754 to 373,294. Out of these 373,294 variants, 373,192 SNPs were found, and the remaining 102 variants were insertions/deletions (INDELS). A total of 17,401 markers were found in the exonic regions. The remaining 355,893 markers that were found in the intronic regions had no information on the genetic structures. The majority of the variant effects were silent (163,400), followed by missense (7273), synonymous (3088) and lastly, nonsense (75) markers. Among all the chromosome locations, the most significant number of markers were found in chromosome 2 (29,962), with the least in chromosome 21 (5108) and X chromosome (4053).
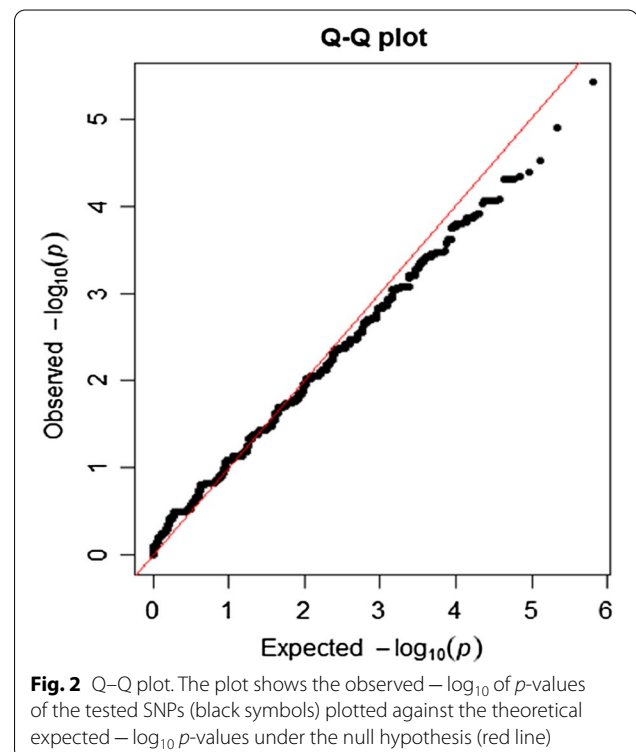
Association analysis

The results depicted in the MDS plot (Fig. 1) represent that most of the cases and controls shared similarities in genetic patterns by forming a cluster of variant alleles at the center of the MDS plot. The matrix distances between most of these sample coordinates in the plot were very close, indicating that the cases and controls had more similarities in their genetic patterns.
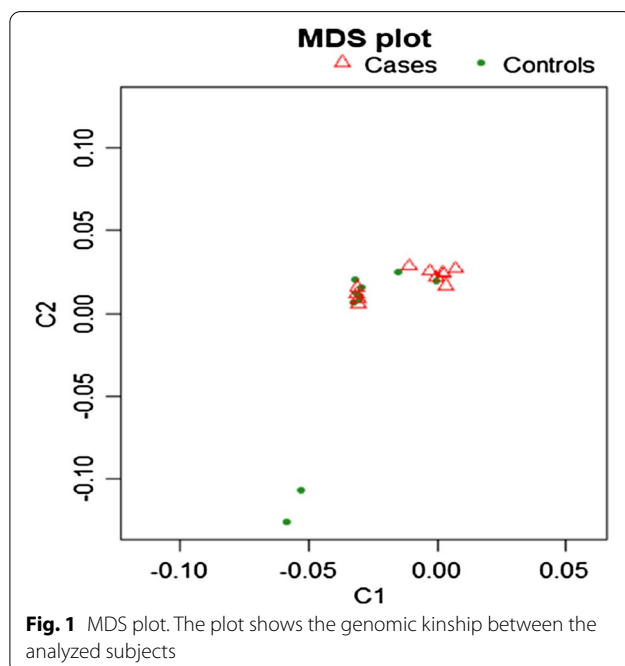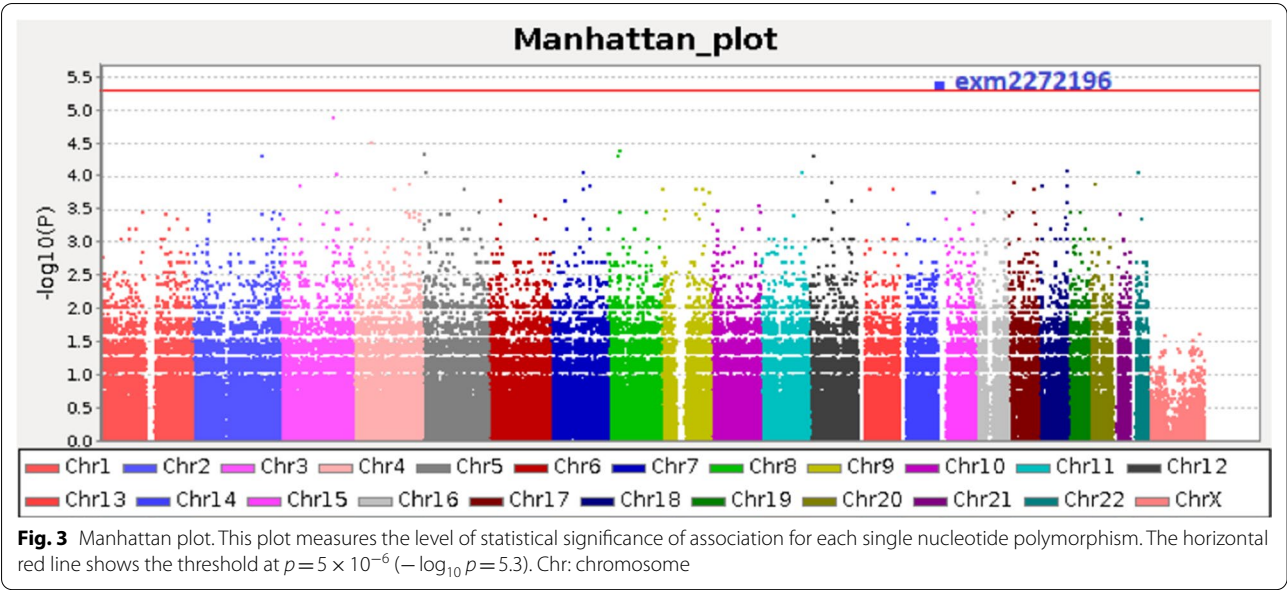
The Q–Q plot (Fig. 2) compares observed $-\log_{10}$ p values of the tested SNPs on the vertical *y*-axis to expected $-\log_{10}$ *p* values under the null hypothesis on the horizontal *x*-axis. The null hypothesis is indicated by the red, standard line plotted on the Q–Q plot. If the observed values correspond to the expected values, all points will be on or near the standard line between the x-axis and y-axis. The black symbols indicate the tested variants. The results depicted on the plot reveal that the overall data distribution is linear. The deviation occurs towards the extreme tail end, where the black symbols do not fall on the standard line. This indicates that there were suggestive true associations of the genetic variants in FECD found in this study.

Manhattan plot (Fig. 3) confirms the findings on the Q–Q plot. It represents the *P* values of the entire GWAS on a genomic scale. The genomic coordinates are displayed along the *x*-axis, with the negative logarithm of the association *P*-value for each variant displayed on the *y*-axis. Each colored dot corresponds to a SNP located in different chromosomes. This plot illustrates the level of statistical significance of association by the *y*-axis, as measured by $-\log_{10}$ *p* values for each variant. The *x*-axis represents the various genomic locations by chromosomes. The strongest associations have the smallest *p*-values which correspond to the large values of their negative logarithms ($-\log_{10}$ *p*). The solid, red

Fig. 1 MDS plot. The plot shows the genomic kinship between the analyzed subjects

Fig. 2 Q–Q plot. The plot shows the observed $-\log_{10}$ of *p*-values of the tested SNPs (black symbols) plotted against the theoretical expected $-\log_{10}$ *p*-values under the null hypothesis (red line)

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 5 of 11

Fig. 3 Manhattan plot. This plot measures the level of statistical significance of association for each single nucleotide polymorphism. The horizontal red line shows the threshold at $p = 5 \times 10^{-6}$ ($-\log_{10} p = 5.3$). Chr: chromosome

Table 2 The most significant variant retrieved from Manhattan Plot

Single nucleotide polymorphism	rs11626651 (exm2272196)
Gene symbol	*LOC105370676*
Chromosome	14
Position (base pair)	101,717,022
Minor allele	A
Number of cases with minor allele ($n = 12$)	11 (91.67%)
Number of controls with minor allele ($n = 12$)	3 (25%)
Frequency of allele (cases)	0.8333
Frequency of allele (controls)	0.1667
Chi-square	21.33
p value	3.86×10^{-6}
$-\log_{10} p$ value	5.41
Odds ratio (confidence interval)	25 (5.478–114.1)

p-value $< 5 \times 10^{-6}$

horizontal line represents an indicative threshold of suggestive evidence of association with the disease, which is $p = 5 \times 10^{-6}$ ($-\log_{10} p = 5.3$). This analysis identified a variant (*rs11626651*) that was strongly associated as it reached the suggestive threshold of $p < 5 \times 10^{-6}$ ($> -\log_{10} p = 5.3$). However, it did not reach the Bonferonni-corrected genome-wide significance threshold of $- < 5 \times 10^{-8}$ ($> -\log_{10} p = 7.3$).

A genetic variant, *rs11626651* (*exm2272196*), located in the *LOC105370676* gene within chromosome 14 (Table 2), was identified from this analysis with $p = 3.86 \times 10^{-6}$ which was significant. The minor allele

(A) distributions of *rs11626651* were higher among FECD cases (11) in comparison to the controls (3). The frequencies of this minor allele were 0.8333 (cases) and 0.1667 (controls) respectively. This finding corresponds to the results of Chi-square, which was 21.33 and Odd ratio (OR) of 25, with a confidence interval of 5.478 to 114.1. Hence, this variant has been deemed as a potential risk factor for FECD.

The candidate genes such as *COL8A2, ZEB1/TCF8, TCF4* and *SLC4A11* were also analyzed to find any associations with FECD in this study. A total of 11 SNPs were found for these genes but they were all not significantly associated with FECD ($p > 5 \times 10^{-6}$) (Table 3).

KOG functional classification analysis

Gene names for markers that passed quality control were obtained from the Infinium Global Screening Array v1.0 Gene Annotation File. There were 17,982 unique genes obtained. Among these genes, 6474 genes were unable to be mapped to the UniProtKB database (https://www.uniprot.org/uploadlists). These unmapped genes include non-coding RNAs, pseudogenes, RNA antisense genes and other small RNAs. In this study, we utilized only the reviewed UniProtKB ID. These UniProtKB IDs were subsequently used to retrieve proteins sequences from Batch Entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez). A custom script was used to compare the sequences to the KOG database to classify the sequences into KOG categories and predict the functions of the proteins. KOG annotation provides four functional groups, each of which is divided into KOG classifications identified by letters of the alphabet (25 in total). A total of 11,508 genes

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 6 of 11

Table 3 Association of candidate genes with FECD

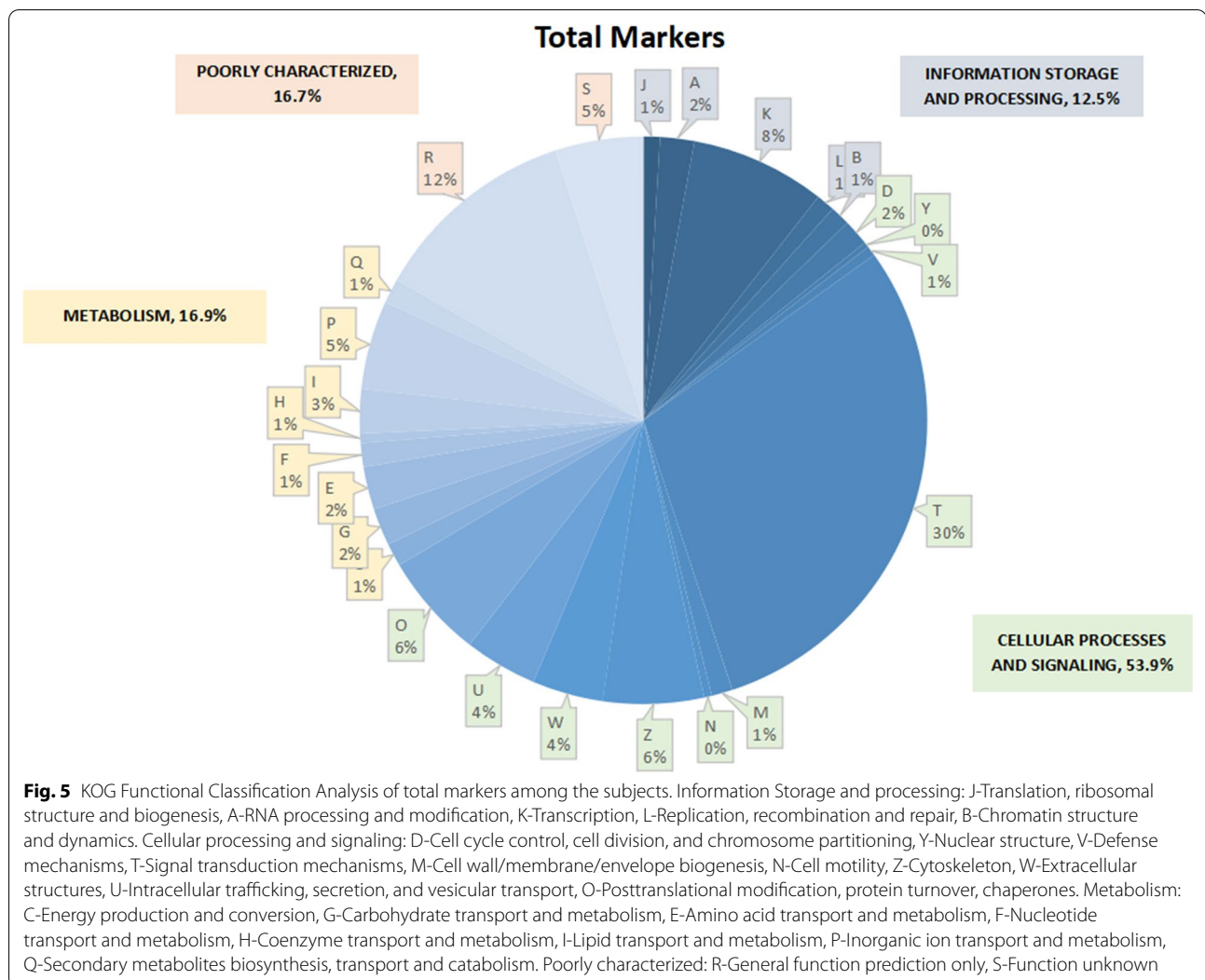| Chromosome number	Gene	SNPs	Position (base pair)	Minor allele	Major allele	Chi square	*p*-value*	Odds ratio (OR)
1	*COL8A2*	rs96067	36106319	G	A	0.7513	0.3861	0.6044
1	*COL8A2*	rs142956038	36110260	G	A	0.3556	0.551	0.4783
1	*COL8A2*	rs274776	36121728	A	G	0.1355	0.7128	1.25
10	*ZEB1/TCF8*	rs431073	31388277	C	A	1.917	0.1662	0
10	*ZEB1/TCF8*	rs74905658	31399430	A	G	0.3556	0.551	0.4783
10	*ZEB1/TCF8*	rs187774661	31405487	C	A	1.021	0.3122	0
10	*ZEB1/TCF8*	rs112139244	31415097	A	G	3.501	0.06133	0
18	*TCF4*	rs77452859	55320989	A	G	0.3556	0.551	2.091
18	*TCF4*	rs9966430	55458133	C	A	1.422	0.233	2.055
18	*TCF4*	rs9636107	55532886	A	G	0.5053	0.4772	0.6
20	*SLC4A11*	rs797045107	3234173	C	A	0.08466	0.7711	0.8442

p-value $> 5 \times 10^{-6}$ ($< -\log_{10} p = 5.3$); SNP: single nucleotide polymorphism

that were retrieved from UniProtKB were annotated and classified. Figures 4 and 5 depict the KOG analysis for the protein functions of the genes and markers.

Among the four major pathways, "Cellular process and signaling" ranked the highest category (53.9% of total KOG classifications). Subsequent categories include the "Metabolism" pathway (16.9%), "Information storage and processing" (12.5%) and the rest were poorly characterized (16.7%). In the largest group of "Cellular processes and signaling", the most frequently observed classes were "Signal transduction" (2275 genes, 20%; 40,532 markers, 30%), "Post-translational

Fig. 4 KOG functional classification analysis of total genes

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 7 of 11

Fig. 5 KOG Functional Classification Analysis of total markers among the subjects. Information Storage and processing: J-Translation, ribosomal structure and biogenesis, A-RNA processing and modification, K-Transcription, L-Replication, recombination and repair, B-Chromatin structure and dynamics. Cellular processing and signaling: D-Cell cycle control, cell division, and chromosome partitioning, Y-Nuclear structure, V-Defense mechanisms, T-Signal transduction mechanisms, M-Cell wall/membrane/envelope biogenesis, N-Cell motility, Z-Cytoskeleton, W-Extracellular structures, U-Intracellular trafficking, secretion, and vesicular transport, O-Posttranslational modification, protein turnover, chaperones. Metabolism: C-Energy production and conversion, G-Carbohydrate transport and metabolism, E-Amino acid transport and metabolism, F-Nucleotide transport and metabolism, H-Coenzyme transport and metabolism, I-Lipid transport and metabolism, P-Inorganic ion transport and metabolism, Q-Secondary metabolites biosynthesis, transport and catabolism. Poorly characterized: R-General function prediction only, S-Function unknown

modification, protein turnover, chaperones" (849 genes, 7%; 8142 markers, 6%) and "Cytoskeleton" (568 genes, 5%; 7784 markers, 6%). The second major pathway was the "Metabolism pathway" of which the most prominent class was "Inorganic ion transport and metabolism" (369 genes, 3%; 6889 markers, 5%), followed by "Lipid transport and metabolism" (359 genes, 3%; 3410 markers, 3%) and "Amino acid transport and metabolism" (327 genes, 3%; 3317 markers, 2%). The smallest group predicted by KOG was "Information storage and processing" of which "Transcription" was categorized as the highest (1217 genes, 11%; 10,449 markers, 8%). Of the "Poorly characterized" category, "General functional prediction" represented the highest of the transcripts (1374 genes, 12%; 15,980 markers, 12%).

Discussion

In this study, the Infinium GSA-24 version 1.0 Bead-Chip was used to determine the genetic variants associated with FECD subjects. This genotyping microarray allows rapid variant screening and high-throughput sample processing, using HTS format [27]. It also provides an economical tool for population-scale genomics and screening. The Illumina iScan is a laser-based, high-resolution optical imaging system that integrates the high-throughput capability of genotyping. It rapidly scans to collect large volumes of data from the BeadChips [28].

The genetic basis of FECD is complex and heterogeneous, demonstrating variable expressivity and incomplete penetrance. Early-onset FECD is rare. It is generally regarded as an autosomal dominant disorder with a Mendelian inheritance pattern. Late-onset FECD is more

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 8 of 11

common, but the genetic cause of this form remains unclear, with some studies describing it as autosomal dominant with variable penetrance [16, 18]. These conditions are seen in the candidate genes like *COL8A2* [15] and *SLC4A11* [12].

TCF8 mutations that have been seen in posterior polymorphous dystrophy can cause late-onset FECD [19]. Five loss-of-function mutations in the *TCF8* gene were detected. This rare variation in the gene encodes the *ZEB1* on chromosome 10p11.2. A recent genome-wide association study (GWAS) revealed an association between FECD and alleles of *TCF4* [20]. It concluded that a SNP (*rs613872*) on chromosome 18q21 had a significant association with late-onset FECD.

We discovered a novel genetic variant (*rs11626651*), a variant of the *LOC105370676* gene, that was associated with FECD. This variant is a C/T single nucleotide variation in human chromosome 14, which is located in the intron region of *XR_001750895.1* (locus position). *LOC105370676* is a gene with uncertain functions. When using the gene ID in GeneCards Human Gene Database (https://www.genecards.org), *LOC105370676* is matched with Long Intergenic Non-Protein Coding RNA 2320 (*LINC02320*). It is an RNA Gene that is affiliated with the non-coding RNA class. *LINC02320* is a Long Intergenic Non-Protein Coding RNA (lincRNA) with unrecognized biological interplay and no protein-coding capacity. The LincRNAs are one of the subgroups of a long non-coding RNA (lncRNAs) which are transcripts longer than 200 nucleotides that are not translated into protein. Mechanisms and functions of lncRNAs are very poorly understood. The LincRNAs are a heterogeneous class of endogenous RNA molecules that are transcribed from intergenic regions of the genome. They are found in between coding genes that do not overlap with protein-coding genes [29].

Although the functions of most lincRNAs are unknown, a few lincRNAs have been discovered to regulate gene expression at the level of chromatin modification, transcription and post-transcriptional processing [30, 31]. Some lincRNAs are involved in regulating the expression of coding genes. An alteration in the expression is associated with dysregulation of the specific proteins leading to various human diseases such as cancer, ageing and ocular disorders [32–34]. Studies have shown that this non-coding RNA can be misexpressed in solid tumors and leukemias [35]. Although the lincRNAs may have impacts on several human diseases, only less than 1% has been characterized out of over 3000 human lincRNAs [36] due to an unclear basis of molecular mechanisms.

In relation to the human eye, the presence of the lincRNAs may contribute to mediating vision which is a process dependent on strict transcriptional control of genes specific to terminally differentiated photoreceptor neurons [37]. A few other lincRNAs play important roles in the development of ocular conditions such as choroidal neovascularization [38], diabetic retinopathy [39], glaucoma [40] and corneal neovascularization [41]. However, the mechanisms of how they affect these diseases remain questionable. There is no established relationship between the *LINC02320* gene and FECD. Thus, further evaluation using microarray analysis or high throughput RNA sequencing may provide a better understanding of lincRNA regulation and the disease etiology especially in uncommon polygenic diseases like FECD [33].

Globally, several studies have been conducted using different platforms of genotyping and study designs to detect the genetic variations in FECD [13, 42–44]. These studies have shown that candidate genes such as *COL8A2, SLC4A11, TCF4,* and *TCF8/ZEB1* are significantly associated with FECD subjects in various populations. However, Table 4 represents the association of candidate genes in several populations of different ethnicities with conflicting results. Some population studies involving Indians [13, 45], Chinese [43], Singaporeans [44], and Japanese [46] have failed to show genetic variations of these genes. The susceptibility of genes to mutations which can vary in different ethnicities and racial backgrounds may have caused the failure. In this study, these candidate genes were not significantly associated with our FECD subjects. This is also possibly due to the variation of gene susceptibility in various ethnicities of Malaysians (Malay, Chinese and Indian).

This study has to be interpreted within the context of various limitations. One of the most important limitations of this study is the small sample size (12 cases and 12 controls). Larger sample size is needed to confirm minimal effects or subtle differences with statistical confidence. Type I (false positives) or II errors (false negatives) may occur in small sample- sized studies. A significant genetic variant that was found in this study could be caused by a false positive error. The underpowered statistically significant study can result in false-negative errors. This could hinder a reliable assessment of significant associations between FECD and candidate genes like *COL8A2, SLC4A11, TCF8/ZEB1* and *TCF4* in this study. The differences in ethnicities in this study may cause the possible variation in genetic risks, which could falsely identify the subgroup of ethnically associated genes as being FECD related.

Conclusions
To our knowledge, this is the first report from Malaysia that involves screening of the genetic variants among FECD subjects in a tertiary care setting using

Ng *et al. Egypt J Med Hum Genet*　　(2021) 22:73

Page 9 of 11

Table 4 Associations of candidate genes in various populations

First Author (year)	Country	Ethnicity	Study design	Genotyping method	Case/Control	SNP	Gene	Association
Baratz et al.,2010 [20]	USA	Caucasian	GWAS	Taqman assay	130/260	rs613872	*TCF4*	*
Thalamuthu et al.,2011 [44]	Singapore	Asian	Case–control	MassArray	57/121	rs613872	*TCF4*	**
Wieben et al., 2014 [47]	USA	Caucasian	Case–control	Sequencing	68/16	rs613872	*TCF4*	*
Nanda et al., 2014 [45]	India	Asian	Case–control	Sequencing	44/108	rs613872	*TCF4*	**
Hemadevi et al., 2010 [13]	India	Asian	Case–control	Sequencing	80/100	Several SNPs	*COL8A2, SLC4A11*	**
Kobayashi et al., 2004 [46]	Japan	Asian	Case–control	Sequencing	15/36	R155Q and T502M	*COL8A2*	**
Mehta et al., 2008 [43]	Chinese	Asian	Case–control	Sequencing	74/93	Asn696Ser	*TCF8*	**
Current study	Malaysia	Malaysian	GWAS	Global screening array	12/12	rs11626651	*LINC02320*	*

* $p < 0.05$, ** $p > 0.05$

Infinium GSA. A single novel variant (*rs11626651*) of the *LOC105370676* gene, also known as the *LINC02320* gene, located at chromosome 14 was found to be associated with FECD. Although this particular gene has no functional role and protein-coding capacity, a few other types of lincRNAs have been found to contribute to the development of some ocular conditions. This variant could play a plausible role in the development of FECD, at least in Malaysia. Further studies with a larger sample size, RNA sequencing, and lincRNA regulation are much recommended. This would be beneficial to confirm the association found and discover other genetic variations. The results of this study could be the primary research platform for the researchers to conduct future studies related to FECD. The identification of additional genetic risk factors adds to our understanding of FECD, which may pave the way for therapeutic strategies in the future.

Abbreviations
COG: Clusters of orthologous groups; DNA: Deoxyribonucleic acid; FECD: Fuchs' endothelial corneal dystrophy; GSA: Global screening array; GWAS: Genome-wide association study; HTS: High throughput screening; HWE: Hardy–Weinberg equilibrium; KOG: Eukaryotic orthologous groups; LincRNA: Long intergenic non-protein coding RNA; lncRNA: Long non-coding RNA; MAF: Minor allele frequency; MDS: Multidimensional scaling; NGS: Next-generation sequencing; Q–Q: Quantile–quantile; SNPs: Single nucleotide polymorphisms.

Authors' contributions
KHN: Conceptualization, data acquisition, analysis validation, writing and editing original draft. VS: Conceptualization, visualization, supervision and final manuscript review and editing. VR: Conceptualization, formal analysis, methodology, writing, reviewing and editing original draft. FI: Conceptualization, data acquisition, visualization, fund acquisition, supervision, final manuscript review and editing. All authors have read and approved the manuscript. The authors agree to be personally accountable for the integrity and accuracy of the study.

Availability of data and material
The data are available upon the request to the corresponding author.

Declarations

Ethics approval and consent to participate
Ethical approval (MECID NO. 201512–1947) was obtained from Medical Ethics Committee of University Malaya Medical Centre prior to the study. All the participants were given written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare that there are no competing interests for this publication.

Author details
[1]Department of Ophthalmology, University Malaya Medical Centre, Kuala Lumpur, Malaysia. [2]Department of Ophthalmology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [3]Malaysian Research Institute of Ageing, Universiti Putra Malaysia, Serdang, Selangor, Malaysia. [4]Centre for Materials Engineering and Regenerative Medicine, Bharath Institute of Higher Education and Research, Selaiyur, Chennai 600 073, Tamil Nadu, India.

References

1. Adamis AP, Filatov V, Tripathi BJ (1993) Fuchs' endothelial dystrophy of the cornea. Surv Ophthalmol 38(2):149–168
2. Lorenzetti DW, Uotila MH, Parikh N, Kaufman HE (1967) Central cornea guttata: incidence in the general population. Am J Ophthalmol 64(6):1155–1158
3. Zoega GM, Fujisawa A, Sasaki H, Kubota A, Sasaki K, Kitagawa K et al (2006) Prevalence and risk factors for cornea guttata in the Reykjavik Eye Study. Ophthalmology 113(4):565–569
4. Kitagawa K, Kojima M, Sasaki H, Shui YB, Chew SJ, Cheng HM et al (2002) Prevalence of primary cornea guttata and morphology of corneal endothelium in aging Japanese and Singaporean subjects. Ophthalmic Res 34(3):135–138
5. Jurkunas UV, Rawe I, Bitar MS, Zhu C, Harris DL, Colby K et al (2008) Decreased expression of peroxiredoxins in Fuchs' endothelial dystrophy. Investig Ophthalmol Vis Sci 49(7):2956–2963
6. Jurkunas UV, Bitar MS, Funaki T, Azizi B (2010) Evidence of oxidative stress in the pathogenesis of fuchs endothelial corneal dystrophy. Am J Clin Pathol 177(5):2278–2289
7. Kaji Y, Amano S, Usui T, Oshika T, Yamashiro K, Ishida S et al (2003) Expression and function of receptors for advanced glycation end products in bovine corneal endothelial cells. Investig Ophthalmol Vis Sci 44(2):521–528
8. Wang Z, Handa JT, Green WR, Stark WJ, Weinberg RS, Jun AS (2007) Advanced glycation end products and receptors in Fuchs' dystrophy corneas undergoing Descemet's stripping with endothelial keratoplasty. Ophthalmology 114(8):1453–1460
9. Zhang X, Igo RP, Fondran J, Mootha VV, Oliva M, Hammersmith K et al (2013) Association of smoking and other risk factors with Fuchs' endothelial corneal dystrophy severity and corneal thickness. Investig Ophthalmol Vis Sci 54(8):5829–5835
10. Liu C, Vojnovic D, Kochevar IE, Jurkunas UV (2016) UV-A irradiation activates Nrf2-regulated antioxidant defense and induces p53/caspase3-dependent apoptosis in corneal endothelial cells. Investig Ophthalmol Vis Sci 57(4):2319–2327
11. Tiwari A, Bahr A, Bähr L, Fleischhauer J, Zinkernagel MS, Winkler N et al (2016) Next generation sequencing based identification of disease-associated mutations in Swiss patients with retinal dystrophies. Sci Rep 6(1):1–1
12. Vithana EN, Morgan PE, Ramprasad V, Tan DT, Yong VH, Venkataraman D et al (2008) SLC4A11 mutations in Fuchs endothelial corneal dystrophy. Hum Mol Genet 17(5):656–666
13. Hemadevi B, Srinivasan M, Arunkumar J, Prajna NV, Sundaresan P (2010) Genetic analysis of patients with Fuchs endothelial corneal dystrophy in India. BMC Ophthalmol 10(1):1–6
14. Iliff BW, Riazuddin SA, Gottsch JD (2012) The genetics of Fuchs' corneal dystrophy. Exp Rev Ophthalmol 7(4):363–375
15. Gottsch JD, Sundin OH, Liu SH, Jun AS, Broman KW, Stark WJ et al (2005) Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of fuchs corneal dystrophy. Investig Ophthalmol Vis Sci 46(6):1934–1939
16. Sundin OH, Jun AS, Broman KW, Liu SH, Sheehan SE, Vito EC et al (2006) Linkage of late-onset Fuchs corneal dystrophy to a novel locus at 13pTel-13q12. 13. Investig Ophthalmol Vis Sci 47(1):140–145
17. Sundin OH, Broman KW, Chang HH, Vito EC, Stark WJ, Gottsch JD (2006) A common locus for late-onset Fuchs corneal dystrophy maps to 18q21. 2–q21. 32. Investig Ophthalmol Vis Sci 47(9):3919–3926
18. Riazuddin SA, Eghrari AO, Al-Saif A, Davey L, Meadows DN, Katsanis N et al (2009) Linkage of a mild late-onset phenotype of Fuchs corneal dystrophy to a novel locus at 5q33. 1–q35. 2. Investig Ophthalmol Vis Sci 50(12):5667–5671
19. Riazuddin SA, Zaghloul NA, Al-Saif A, Davey L, Diplas BH, Meadows DN et al (2010) Missense mutations in TCF8 cause late-onset Fuchs corneal dystrophy and interact with FCD4 on chromosome 9p. Am J Hum Genet 86(1):45–53
20. Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W et al (2010) E2–2 protein and Fuchs's corneal dystrophy. N Engl J Med 363(11):1016–1024
21. Chandra A, Mitry D, Wright A, Campbell H, Charteris DG (2014) Genome-wide association studies: applications and insights gained in ophthalmology. Eye 28(9):1066–1079

22. Glöckle N, Kohl S, Mohr J, Scheurenbrand T, Sprecher A, Weisschuh N et al (2014) Panel-based next generation sequencing as a reliable and efficient technique to detect mutations i- unselected patients with retinal dystrophies. Eur J Hum Genet 22(1):99–104
23. Zhao L, Wang F, Wang H, Li Y, Alexander S, Wang K et al (2015) Next-generation sequencing-based molecular diagnosis of 82 retinitis pigmentosa probands from Northern Ireland. Hum Genet 134(2):217–230
24. Gelernter J, Zhou H, Nuñez YZ, Mutirangura A, Malison RT, Kalayasiri R (2018) Genome-wide association study of alcohol dependence and related traits in a Thai population. Alcohol Clin Exp Res 42(5):861–868
25. Mueller SH, Färber A, Prüss H, Melzer N, Golombeck KS, Kümpfel T et al (2018) Genetic predisposition in anti-LGI1 and anti-NMDA receptor encephalitis. Ann Neurol 83(4):863–869
26. Huang J, Lin Y. Genotyping technologies and applications in the era of precision medicine. N Am J Med Sci. 2017;10(4).
27. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol cell 58(4):586–597
28. Wu MC, Kuan PF (2018) A guide to Illumina BeadChip data analysis. In DNA methylation protocols. Humana Press, New York, p. 303–330.
29. Kung JT, Colognori D, Lee JT (2013) Long noncoding RNAs: past, present, and future. Genetics 193(3):651–669
30. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10(3):155–159
31. Ransohoff JD, Wei Y, Khavari PA (2018) The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 19(3):143
32. Tsai MC, Spitale RC, Chang HY (2011) Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res 71(1):3–7
33. Li F, Wen X, Zhang H, Fan X (2016) Novel insights into the role of long noncoding RNA in ocular diseases. Int J Mol Sci 17(4):478
34. Sahu A, Singhal U, Chinnaiyan AM (2015) Long noncoding RNAs in cancer: from function to translation. Trends Cancer 1(2):93–109
35. Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C et al (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell 12(3):215–229
36. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Morales DR et al (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106(28):11667–11672
37. Mustafi D, Kevany BM, Bai X, Maeda T, Sears JE, Khalil AM et al (2013) Evolutionarily conserved long intergenic non-coding RNAs in the eye. Hum Mol Genet 22(15):2992–3002
38. Xu XD, Li KR, Li XM, Yao J, Qin J, Yan B (2014) Long non-coding RNAs: new players in ocular neovascularization. Mol Biol Rep 41(7):4493–4505
39. Yan B, Tao ZF, Li XM, Zhang H, Yao J, Jiang Q (2014) Aberrant expression of long noncoding RNAs in early diabetic retinopathy. Investig Ophthalmol Vis Sci 55(2):941–951
40. Wiggs JL, Yaspan BL, Hauser MA, Kang JH, Allingham RR, Olson LM et al (2012) Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. PLoS Genet 8(4):e1002654
41. Huang J, Li YJ, Liu JY, Zhang YY, Li XM, Wang LN et al (2015) Identification of corneal neovascularization-related long noncoding RNAs through microarray analysis. Cornea 34(5):580–587
42. Li YJ, Minear MA, Rimmler J, Zhao B, Balajonda E, Hauser MA et al (2011) Replication of TCF4 through association and linkage studies in late-onset Fuchs endothelial corneal dystrophy. PLoS ONE 6(4):e18044
43. Mehta JS, Vithana EN, Tan DT, Yong VH, Yam GH, Law RW et al (2008) Analysis of the posterior polymorphous corneal dystrophy 3 gene, TCF8, in late-onset Fuchs endothelial corneal dystrophy. Investig Ophthalmol Vis Sci 49(1):184–188
44. Thalamuthu A, Khor CC, Venkataraman D, Koh LW, Tan DT, Aung T et al (2011) Association of TCF4 gene polymorphisms with Fuchs' corneal dystrophy in the Chinese. Investig Ophthalmol Vis Sci 52(8):5573–5578
45. Nanda GG, Padhy B, Samal S, Das S, Alone DP (2014) Genetic association of TCF4 intronic polymorphisms, CTG18. 1 and rs17089887, with Fuchs' endothelial corneal dystrophy in an Indian population. Investig Ophthalmol Vis Sci 55(11):7674–7680
46. Kobayashi A, Fujiki K, Murakami A, Kato T, Chen LZ, Onoe H et al (2004) Analysis of COL8A2 gene mutation in Japanese patients with Fuchs' endothelial dystrophy and posterior polymorphous dystrophy. Jpn J Ophthalmol 48(3):195–198

Ng *et al. Egypt J Med Hum Genet*   (2021) 22:73

Page 11 of 11

47. Wieben ED, Aleff RA, Eckloff BW, Atkinson EJ, Baheti S, Middha S et al (2014) Comprehensive assessment of genetic variants within TCF4 in Fuchs' endothelial corneal dystrophy. Investig Ophthalmol Vis Sci 55(9):6101–6107