Molecular Vision 2025; 31:486-500 <http://www.molvis.org/molvis/v31/486>
Received 24 May 2025 | Accepted 28 November 2025 | Published 1 December 2025

Targeted sequencing with single-molecule molecular inversion probes highlights a gap in understanding the cause of Fuchs endothelial corneal dystrophy

Bushra Alayed,[1,2] Danah Albuainain,[1,3] Salina Siddiqui,[1,4] Weijia Li,[1] Ummey Hany,[1] Seema Anand,[4] Chris F. Inglehearn,[1] Christopher M. Watson,[1,5] Manir Ali[1]

[1]Division of Molecular Medicine, Leeds Institute of Medical Research, St. James's University Hospital, University of Leeds, Leeds, UK; [2]Department of Medical Laboratories, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia; [3]Department of Anatomy, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia; [4]The Eye Department, St. James's University Hospital, Leeds, UK; [5]North East and Yorkshire Genomic Laboratory Hub, Central Lab, St James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

Purpose: A trinucleotide repeat expansion in *TCF4* is thought to cause Fuchs endothelial corneal dystrophy (FECD) in ~70% of European patients. In addition, strong evidence exists for the involvement of rare variants in *COL8A2* and *SLC4A11* in a small number of FECD cases, and more controversially, it has been suggested that variants in *ZEB1*, *AGBL1*, and *LOXHD1* may also be involved. We screened patients without a *TCF4* repeat expansion for causative variants in the other candidate FECD genes.

Methods: Genomic DNA from blood was genotyped for expansion of the CTG18.1 repeat in intron 2 of *TCF4* using short-tandem repeat PCR, followed by triplet-repeat primed PCR (STR/TP-PCR). Single-molecule molecular inversion probes (smMIPs) were designed to amplify the coding exons and splice recognition sites of FECD candidate genes *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1* using MIPGEN, and the libraries generated were sequenced on a NextSeq 2000. FASTQ-formatted sequence reads were aligned to the human reference genome with MIPVAR, and identified variants were annotated using Annovar. Rare potentially pathogenic variants (minor allele frequency ≤0.01 for assumed dominant inheritance, combined annotation-dependent depletion ≥15) were confirmed by Sanger sequencing and interpreted according to American College of Medical Genetics and Genomics criteria using the Franklin by Genoox interface.

Results: Analysis of 114 FECD cases by STR/TP-PCR stratified the patients into FECD expansion-negative cases that had <50 trinucleotide repeats on both alleles at the CTG18.1 locus (n = 33 probands and three additional family members) and FECD expansion-positive (n = 78) cases with at least one allele harboring ≥50 repeats in size. All 36 expansion negative cases were then analyzed by smMIP targeted capture and short-read sequencing of the five other genes implicated in FECD causation. For comparison, two control groups were similarly analyzed: a subset of 29 of the expansion-positive cases whose FECD was assumed to be caused by the repeat expansion and 29 expansion-negative unaffected individuals. Across all groups, 13 variants passed filtration criteria: 1 in *COL8A2*, 2 in *SLC4A11*, 4 in *AGBL1*, and 6 in *LOXHD1*. No variants were identified in *ZEB1.* Eight of the variants identified were found in seven FECD expansion-negative cases, five in four FECD expansion-positive cases, and three in two non-FECD controls, with three variants appearing in both expansion-positive and expansion-negative patient groups. Statistical analysis indicated no significant enrichment of variants in expansion-negative cases compared to the other two groups (p = 0.7612 and 0.3275). Only one variant (*LOXHD1* NM_144612.7, c.5545G>A, p.(Gly1849Arg)), found in an expansion-negative case, was classified as likely pathogenic, while others were classed as variant of unknown significance (n = 4), likely benign (n = 4), or benign (n = 4).

Conclusions: smMIPs were used for targeted screening of candidate genes implicated in FECD and proved a versatile, economic approach for prescreening before whole-exome or genome sequencing. This study confirmed the well-documented enrichment of the *TCF4* repeat expansion in cases over controls but found a paucity of evidence for the involvement of variants in *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, or *LOXHD1* in this set of expansion-negative cases, implying knowledge of the causes of FECD remains incomplete.

Fuchs endothelial corneal dystrophy (FECD) is a progressive, bilateral eye disease that primarily affects the corneal endothelium, the monolayer of cells in the innermost part of the cornea, and occurs in 1 in 25 individuals over 40 years of age [1,2]. The condition is marked by the presence of "guttae," which are drop-like excrescences that protrude posteriorly from the Descemet membrane, and by gradual loss of endothelial cells that affects the barrier function of the endothelium, causing hydration of the cornea due to fluid entering from the anterior chamber [3]. This leads to edema, corneal clouding, blurred vision, glare, and, if left untreated,

Correspondence to: Manir Ali, Division of Molecular Medicine, Leeds Institute of Medical Research, St. James's University Hospital, University of Leeds, Leeds, UK; email: m.ali@leeds.ac.uk

blindness. Endothelial transplantation is the preferred method of treatment, with FECD the main cause of corneal transplantation worldwide [4].

A trinucleotide repeat expansion of ≥50 copies in intron 2 of the gene encoding transcription factor 4 (*TCF4*, OMIM *602272, #613267) accounts for over 70% of Caucasian cases with FECD [5–7]. The repeat expansion causes disease in corneal endothelial cells by forming RNA nuclear foci that sequester RNA-binding proteins, leading to global missplicing [8–10]. Although much less common than the *TCF4* expanded repeat sequence, heterozygous variants in collagen type VIII alpha 2 chain (*COL8A2*, OMIM *120252, #136800) cause earlier-onset FECD in the 20s [11,12]. COL8A2 is expressed by corneal endothelial cells and is one subunit of a heterotrimer that makes up short-chain collagen fibrils and forms part of the hexagonal lattice structure in the Descemet membrane [13]. A knock-in mutant mouse model that has similar features to the human disease showed that the predominant effect of the mutation on corneal endothelium was activation of the unfolded protein response, leading to endoplasmic reticulum stress and apoptosis [14]. Classical FECD with onset in the 40s can also be caused by heterozygous mutations in solute carrier family 4 member 11 (*SLC4A11*, OMIM *610206, #613268) [15,16]. SLC4A11, which is expressed in corneal endothelial cells, is a transmembrane-bound sodium-coupled borate transporter that is required for intracellular boron homeostasis, cell growth, and proliferation [17]. Biallelic variants in *SLC4A11* cause the corneal condition, congenital hereditary endothelial dystrophy type 2 (CHED2) [18]. Coexpression studies have shown that SLC4A11 is dimeric, and variants that cause disease are retained intracellularly and can be partially rescued by wild-type SLC4A11 for CHED2-causing recessive variants, but not for FECD-causing dominant variants [19].

There have also been studies suggesting a causal role for rare heterozygous variants in zinc finger E-box binding homeobox 1 (*ZEB1*, OMIM *189909, #613270) [20,21], ATP/GTP binding protein like 1 (*AGBL1*, OMIM *615496, #615523) [22,23], and lipoxygenase homology polycystin/lipoxygenase/alpha-toxin domain 1 (*LOXHD1*, OMIM *613072) [24] in FECD. ZEB1, also called transcription factor 8 (TCF8), is a transcriptional repressor of the epithelial cell marker E-cadherin, thereby contributing to the epithelial-mesenchymal cell transition [25]. ZEB1 is expressed in the cornea, and dominant variants have previously been shown to cause the related eye condition posterior polymorphous corneal dystrophy, characterized by abnormal changes in the corneal endothelium and Descemet's membrane, which result in cell differentiation and the presence of multiple layers of epithelial-like cells [26,27]. AGBL1, also called cytosolic carboxypeptidase 4, encodes a glutamate decarboxylase that catalyzes the deglutamylation of polyglutamylated proteins [28]. Functional analysis suggested that AGBL1 interacts with TCF4, and that disease-causing variants diminish that interaction [22]. LOXHD1 is predominantly expressed in the hair cells of the inner ear, and biallelic variants cause progressive hearing loss in humans [29]. Immunohistochemical analysis of mouse sections suggested that there is expression, particularly in the corneal epithelium [22]. However, transcriptome analyses of normal [30–32] and FECD-derived corneal endothelium [33] noted the absence of *AGBL1* and *LOXHD1* transcripts. To date, there remains significant doubt as to the involvement of variants in *ZEB1*, *AGBL1*, and *LOXHD1* in causing FECD, and further independent familial studies, which are currently lacking, and functional analyses are required [34].

Single-molecule molecular inversion probes (smMIPs), also sometimes called padlock probes, are linear oligonucleotides that contain a common DNA backbone with extension and ligation arms at each end (see Figure 1A). The two arms provide dual recognition of specific sequences to target the region of interest. The gap between the two arms is filled by the action of DNA polymerase, and the hybridized sequences are then circularized by DNA ligase. This circular DNA is PCR amplified using the universal primers attached to the smMIP backbone. Thousands of probes can be pooled together without interfering with each other due to their combined intrinsic specificity. Molecular inversion probes were first used for high-throughput genotyping with microarrays [35], while the smMIP adaptation was developed for the detection of low-frequency variants in mixed cell populations [36]. smMIPs have unique barcoded indices added by PCR to permit pooling and then subsequent deconvolution of multiple patient samples in a single sequencing experiment [37]. smMIPs are an efficient and cost-effective technique for variant scanning; numerous genes can be sequenced concurrently, they require a low starting mass of DNA (200 ng), and generated libraries can be multiplexed [37–39].

Here, we report results from a custom smMIP panel designed to capture the coding sequence and splice recognition sites of five candidate genes, *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1*, with the last three being large multiexon genes, in which variants have been suggested to be associated with FECD. The reagent was used to screen cases that did not have a repeat expansion in intron 2 of *TCF4*, in an attempt to identify the genetic causes of FECD in these cases.

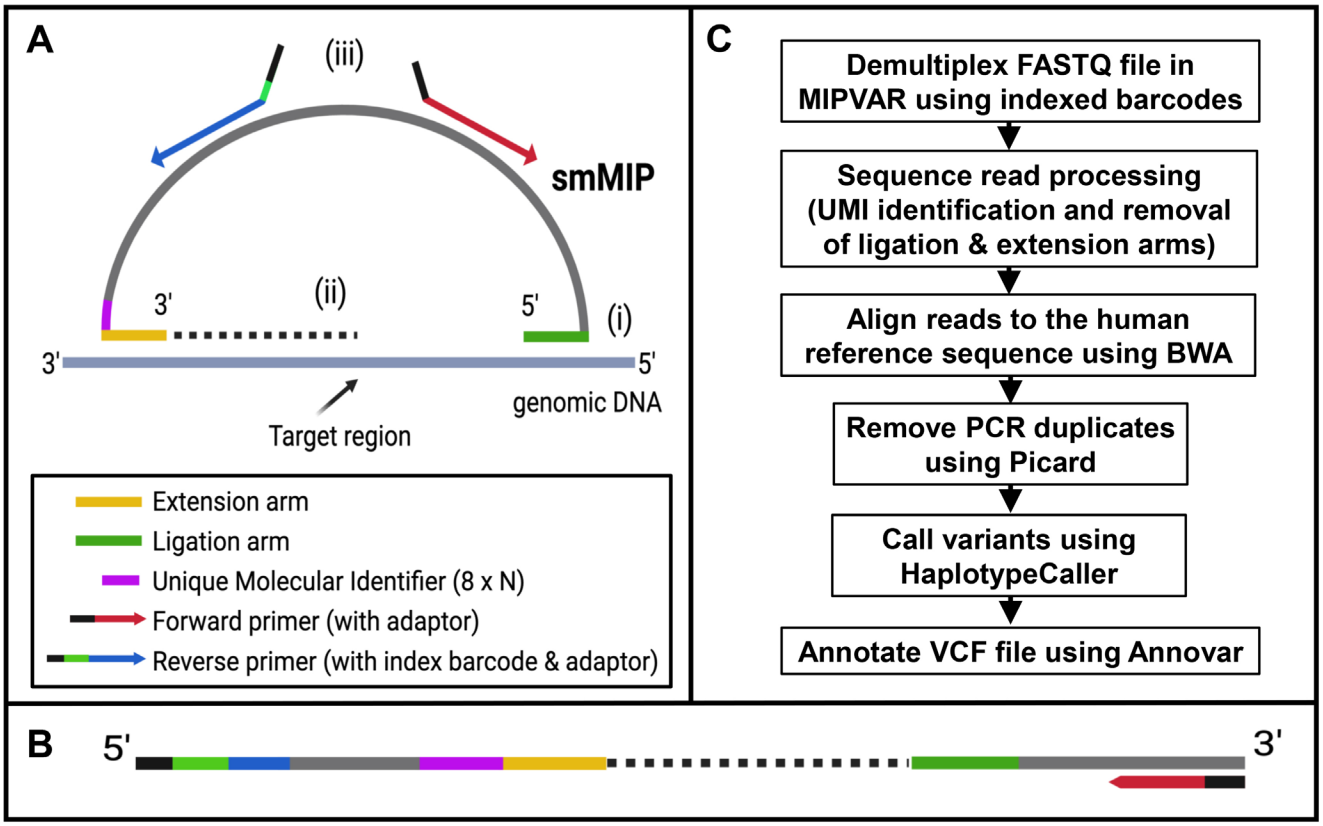


Figure 1. Summary of smMIP-based targeted sequencing workflow. **A**. Schematic of one smMIP spanning target region of interest in genomic DNA. The 5′ and 3′ orientations are shown. Each smMIP of approximately 83 nucleotides in length consists of an oligonucleotide backbone (light gray, 30 bp) with specific sequences for primer binding (red and blue), extension (yellow, 16–22 bp), and ligation (in green, 22–28 bp) arms and the unique molecular identifier (UMI; purple, 8 bp). The smMIP hybridizes to the target region of interest (i), then undergoes extension and ligation to complete a DNA circle (ii). This is linearized following PCR with universal primers containing an 8-bp index barcode unique to that sample and adaptor sequences specific to the sequencing platform being used (iii), to give the linearized product shown in (**B**). For each patient sample, a megapool of smMIPs can be used simultaneously to create a library with its own unique index, which can be pooled with other libraries and sequenced. **C**. Pipeline of data analysis following sequencing. Panels A and B were created in BioRender. Alayed, B. (2025).

METHODS

Patient recruitment and sampling: Patients with FECD, along with their affected family members and age-, sex-, and ethnically matched non-FECD controls, were recruited in the Eye Department at St. James's University Hospital in Leeds, UK. All patients with FECD were ascertained from outpatient or operating theater lists and identified as having endothelial dystrophies in the electronic records. They had all undergone detailed slit-lamp examination and specular microscopy at the time of listing and satisfied the criteria of FECD grade 1 or above, according to the modified Krachmer grading scale. Informed consent was obtained using a process approved by the Leeds East Research Ethics Committee (Reference Number: 17/YH/0032) and following the principles of the Declaration of Helsinki. The cohort consisted of Caucasians aged 50 or older at recruitment and was not indicative of disease onset. Non-FECD controls were recruited from cataract clinics and, before surgery, had their corneal endothelium inspected using a specular microscope to exclude FECD. A total of 3 ml of peripheral blood was sampled by venipuncture into standard EDTA blood collection tubes (BD Biosciences, Wokingham, UK) for genomic DNA extraction using established methods.

Short tandem repeat and triplet-repeat primed PCR genotyping assays: Genotyping assays for measuring the size of the CTG18.1 repeat sequence have been described elsewhere [6]. Briefly, short tandem repeat (STR) genotyping was initially performed on genomic DNA, followed by triplet-repeat PCR (TP-PCR) on those patients in whom both alleles could not be identified using the STR assay, to confirm the

presence or absence of an allele expansion. The oligonucleotide primer sequences used were P1 (FAM-AAT CCA AAC CGC CTT CCA AGT) and P2 (CAA AAC TTC CGA AAG CCA TTT CT) for the STR assay and primers P1, P3 (TAC GCA TCC CAG TTT GAG ACG), and P4 (TAC GCA TCC CAG TTT GAG ACG CAG CAG CAG CAG CAG), in a ratio of 1:1:0.3, for the TP-PCR assay. For the STR assay, 95 °C for 2 min was followed by 30 PCR cycles of 94 °C denaturation for 30 s, 61 °C annealing for 30 s, and 72 °C extension for 30 s, then a final extension step of 72 °C for 5 min. For TP-PCR, conditions were 40 cycles of 94 °C for 30 s, 60 °C for 45 s, and 72 °C for 2 min. PCR products were resolved on an ABI 3130xl Genetic Analyzer and analyzed with GeneMapper version 4.0 (Thermo Fisher Scientific, Waltham, MA).

Gene selection and designing the smMIP oligonucleotides: To identify candidate genes in which protein-coding nucleotide variants have been implicated in causing FECD, a keyword search for "FECD" was performed in the Online Mendelian Inheritance in Man (OMIM) database (accessed July 22, 2021). The "Phenotypic Series" tab was selected to identify six candidate genes: *COL8A2*, *ZEB1*, *AGBL1*, *LOXHD1*, *TCF4*, and *SLC4A11*. Of these, the *TCF4* variant is an intronic, trinucleotide repeat expansion that does not alter the protein-coding sequence. Hence, for smMIP design, only the protein-coding exons and splice recognition sites of the other five genes were selected by downloading the hg19 version of human reference sequences as a bed (browser extensible data) file from the UCSC genome browser.

Oligonucleotide probes were designed from genomic coordinates described in the bed file to amplify a specific target sequence between 95 and 115 nucleotides in length, using the MIPGEN program [40]. All the probes were visualized on the Integrative Genomics Viewer (IGV; v.2.16.2) [41], and a single probe was selected to target each sequence on either DNA strand. Each probe overlapped with the adjacent probe with a minimum of 10 nucleotides to ensure that no gaps were present in the target sequence. Probe selection was based on high logistic scores from MIPGEN, with probes scoring≥0.7 preferred for the panel design. However, where that was not possible, probes with lower logistic scores were chosen.

Preparation of the smMIP oligonucleotide megapool: Probes, which were around 80 nucleotides in length, were synthesized at a 100-nmol scale in 96-well formatted plates (Integrated DNA Technologies, Leuven, Belgium). Probes were initially pooled at an equimolar concentration by combining 5 µl of each probe to generate a smMIP megapool. Then, 25 µl of the megapool was phosphorylated with 3 µl 10x T4 ligase buffer containing 10 mM ATP, 1 µl T4 polynucleotide kinase (New England Biolabs, Hitchin, UK), and 1 µl nuclease-free water. The smMIP megapool was then incubated at 37 °C for 45 min, followed by 65 °C for 20 min, in a thermocycler before subsequent use.

The smMIP-based targeted sequencing workflow is shown in Figure 1. Following each optimization run using control genomic DNA, the volume of each probe in the smMIP megapool was adjusted according to average read depth, so that probes with a lower read depth between 10 and 99 were increased in volume, and higher read depth probes over 600 were decreased for subsequent runs. Any probes with a read depth below 10 were increased in volume if they had a favorable logistic score of ≥0.7 or redesigned if they had a lower logistic score.

Preparation of targeted capture libraries and next-generation sequencing: For each subject, 200 ng genomic DNA was hybridized against the phosphorylated smMIP megapool so that the final reaction contained a ratio of 800 smMIP copies for each DNA molecule. Briefly, 25 µl capturing reaction was set up containing genomic DNA in 10 µl volume, 4 µl diluted smMIP megapool (10^{-4} dilution), 2.5 µl Ampligase 10x reaction buffer (Cambio Limited, Cambridge, UK), 0.32 µL dNTP (0.25 mM stock; New England Biolabs), 0.32 µl Hemo KlenTaq (10 U/µl; New England Biolabs), 0.2 µl Ampligase (5 U/µl; Cambio Limited), and 7.66 µl nuclease-free water. This mixture was incubated at 95 °C for 3 min and 65 °C for 22 h. This was followed by chilling on ice and adding 2 µl exonuclease master mix consisting of 0.5 µl Exonuclease I (20,000 U/ml; New England Biolabs), 0.5 µl Exonuclease III (100,000 U/ml; New England Biolabs), 0.2 µlAmpligase 10x reaction buffer, and 0.8 µl nuclease-free water. The reaction was mixed and incubated at 37 °C for 45 min and 95 °C for 2 min. Following exonuclease treatment, 10 µl of the exonuclease-treated capturing reaction was mixed with 12.5 µl Q5 Hot Start High Fidelity 2x master mix (New England Biolabs) and 1.25 µl of each of the forward (dAAT GAT ACG GCG ACC ACC GAG ATC TAC ACA TAC GAG ATC CGT AAT CGG GAA GCT GAA G) and barcoded reverse primers (dCAA GCA GAA GAC GGC ATA CGA GAT XXXXXXXX ACA CGC ACG ATC CGA CGG TAG TGT), where XXXXXXXX are 96 unique index sequences corresponding to 96 reverse primers (Illumina, San Diego, CA; 10 mM stocks). Underlined are the sequences complementary to the DNA backbone in the smMIP oligonucleotides to enable incorporating the indexed barcode in the PCR. The PCR was performed for 1 cycle at 98 °C for 30 s, followed by 22 cycles of 98 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s, with a final extension step of 72 °C for 2 min, and then the reactions were stored at 4 °C. The reaction

products were purified using 20 µl AppMag PCR clean-up beads (Appleton Scientific, Birmingham, UK) according to the supplier's instructions, and the distribution of fragment sizes in the library was visualized using a Tapestation with the High Sensitivity D1000 assay kit (Agilent Technologies Limited, Stockport, UK). The concentration of each individually barcoded library was determined using a Qubit 2.0 fluorometer with High-Sensitivity DNA reagents (Thermo Fisher Scientific).

To generate a single, multiplexed smMIP captured library, 5 ng of each individual library was combined before sequencing. Depending on the number of samples in the pooled library and the depth of coverage required, the library was sequenced using either a MiSeq (Illumina) or NextSeq 2000 (Illumina) instrument, generating paired-end 151-bp reads according to the manufacturer's instructions.

Data analysis and variant interpretation: Raw data were demultiplexed and converted to FASTQ format using BCL Convert (v.3.9.3), then processed using MIPVAR (v.0.1.0); MIP arms were removed and sequence reads aligned to human reference genome build hg19 using the Burrow-Wheeler Aligner (v.0.7.12). Picard (v.1.119) was used to identify and remove PCR duplicates (read pairs with the same mapping coordinates). The Genome Analysis Toolkit HaplotypeCaller (v.3.7.0) was used to identify and record nonreference bases in a variant call format file [42]. Each patient's variant call format was annotated with population allele frequency data and information relevant to pathogenicity using Annovar [43]. To enrich for rare, potentially pathogenic variants, only those with a minor allele frequency below 1% in the Exome Aggregation Consortium (v.0.3) and a combined annotation-dependent depletion (v.1.3) score greater than 15 were retained.

Variants were interpreted according to the American College of Medical Genetics and Genomics (ACMG) criteria using the Franklin by Genoox website [44]. The Genome Aggregation Database (gnomAD; v.4.1) was used to identify variant allele frequencies. Splicing predictions were obtained using SpliceAI. Variant pathogenicity data were obtained from ClinVar. Evolutionary conservation across multiple vertebrate species was assessed using the multiple-protein-sequence alignment resource, the constraint-based alignment tool (COBALT), available on the NCBI website. Copy number variant (CNV) analysis was performed using ExomeDepth [45]. Briefly, samples to be compared were sequenced at the same time on the same machine to prevent confounding batch effects. For each sample, within the FECD expansion-negative, FECD expansion-positive, and non-FECD control categories, the total number of sequencing reads in the binary alignment map file was recorded and aligned in ascending order. Ten samples from each category, with similar total sequencing reads, were compared in batches of 30, using the ExomeDepth commands. The program compared the observed reads for each exon in the test sample, one at a time, against the expected reads from the remaining samples in that batch. The output results were documented in an Excel (Microsoft, Redmond, WA) file. The Bayes factor score indicated the likelihood of the CNV being present in the tested sample, and any with a score greater than 5.0 were prioritized. Read alignments supporting identified variants were visualized using IGV.

Statistical analysis: Fisher's exact test was performed in GraphPad (GraphPad Software, La Jolla, CA). The test was used to compare categorical data in a 2 × 2 contingency table at the 95% significance level.

Variant validation by Sanger sequencing: Primers were designed to generate an amplification product spanning the variant of interest, using Primer3 (v.4.1.0). Each PCR was performed in a 10-µl reaction volume using 40 ng genomic DNA with the primer pair and HotShot Diamond master mix (Clent Life Science, Stourbridge, UK) according to the supplier's instructions. An aliquot of the PCR was treated with ExoSAP-IT (Thermo Fisher Scientific), followed by incubation at 37 °C for 30 min and 80 °C for 15 min. The treated PCR was Sanger sequenced by an external service provider (Azenta Life Sciences, Manchester, UK). Chromatograms were visualized using Sequence Analysis software (v.5.2; Thermo Fisher Scientific).

RESULTS

Optimization of the smMIP targeted capture reagent: The final pool of the target capture reagent consisted of 380 smMIPs that covered the coding sequences and splice recognition signals of five genes, *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1* (Table 1), in which variants had previously been implicated in causing FECD. This rebalanced pool was derived following three optimization runs on the MiSeq, using five non-FECD, genomic DNA samples to evaluate the average read depth of each probe (Figure 2). Following the third optimization run, an average read depth of over 50 reads was achieved by 95.7% (360/376) of probes and over 10 reads by 99.5% (374/376) of probes when covering the targeted regions for the five control samples. Regions with an average read depth below 50 are shown in Appendix 1.

Prescreening samples for repeat expansion in TCF4 using STR/TP-PCR genotyping: Genomic DNA from 114 FECD cases and 29 non-FECD controls were genotyped for the CTG18.1 repeat expansion in intron 2 of *TCF4* using the

STR/TP-PCR assay. Alleles with fewer than 50 repeats at the CTG18.1 locus were measured using the STR assay. Any patients in whom both alleles could not be identified this way were then tested using the TP-PCR assay to confirm the presence or absence of expansions greater than 50 repeats. FECD cases with at least one allele greater than 50 repeats were considered expansion positive, and those without an expanded allele were classed as expansion negative. This analysis stratified the patients into FECD expansion-negative (n = 33 probands and three additional affected family members) and FECD expansion-positive (n = 78) cases and confirmed the absence of *TCF4* repeat expansions in all controls. The cohort demographics following STR/TP-PCR genotyping are summarized in Table 2, and the individual patients' genotyping results are presented in Appendix 2, Appendix 3, and Appendix 4. Briefly, the age range and gender ratios between the FECD cases and non-FECD controls were similar and well matched.

Screening for variants in the known candidate genes using the smMIP targeted capture reagent: All 36 expansion-negative cases were then analyzed by smMIP targeted capture and short-read sequencing of the five other genes implicated in FECD causation. For comparison, two control groups were also screened, one composed of a subset of 29 of the expansion-positive cases and the other composed of the 29 expansion-negative unaffected individuals sampled via cataract clinics. These 94 samples were screened for variants in the FECD candidate genes using the smMIP capture reagent. Following preparation of libraries and pooling of each sample, the mixture was sequenced on a NextSeq 2000. After demultiplexing and bioinformatics processing, the data were analyzed and filtered to identify single-nucleotide variants with a minor allele frequency <1% in the Exome Aggregation Consortium and a combined annotation-dependent depletion score >15.

Eight variants, confirmed by Sanger sequencing (Appendix 5), were identified in seven FECD expansion-negative cases (Table 3A). Variant NM_005202.4: c.7G>A, p.(Gly3Arg) in *COL8A2* was classed as benign by ACMG criteria. Two variants were identified in *AGBL1*, NM_152336.4: c.806C>T, p.(Pro269Leu) and c.3323+2dup, predicted to be benign and a variant of unknown significance (VUS), respectively. The c.3323+2dup variant, which was present in patient 589, was homozygous in six individuals in the European (non-Finnish) general population in gnomAD. There were also five variants identified in *LOXHD1*, NM_01145472.3: c.3340G>A, p.(Gly1114Arg); NM_144612.7: c.5545G>A, p.(Gly1849Arg); c.4504C>T, p.(Arg1502Trp); c.3269G>A, p.(Arg1090Gln); and c.1570C>T, p.(Arg524Cys). Of these, the c.4504C>T variant found in patient 817, classed as a VUS by ACMG criteria, affects an arginine residue that is not evolutionarily conserved among vertebrates (data not shown). The c.5545G>A variant in patient 596, which is classed as likely pathogenic according to ACMG criteria, is predicted to create a new splice acceptor site (Splice AI = 0.79). This patient also harbored a second *LOXHD1* variant, c.1570C>T, p.(Arg524Cys), considered likely benign, although the phase could not be established. As recessive variants in *LOXHD1* cause nonsyndromic, progressive hearing loss [29], patient notes were examined, revealing that the patient developed a hearing deficit in their 70s and required hearing aids. However, this is not uncommon at that age and may be unrelated. The remaining two *LOXHD1* variants, c.3340G>A (NM_001145472.3) and c.3269G>A (NM_144612.7), were

Table 1. Gene candidates in which variants have been associated with FECD.

HGNC gene symbol	Gene name	OMIM number	Cytogenetic location	Transcript	Genomic coordinates	Strand
COL8A2	Collagen Type VIII Alpha 2 Chain	120252	1p34.3	NM_005202.4	chr1:36,560,840–36,590,823	-
SLC4A11	Solute Carrier Family 4 Member 11	610206	20p13	NM_001174089.2	chr20:3,208,068–3,219,836	-
ZEB1	Zinc Finger E-box Binding Homeobox 1	189909	10p11.22	NM_001174096.2	chr10:31,608,145–31,818,732	+
AGBL1	AGBL Carboxypeptidase 1	615496	15q25.3	NM_001386094.1	chr15:86,623,102–87,459,387	+
LOXHD1	Lipoxygenase Homology PLAT Domains 1	613072	18q21.1	NM_001384474.1	chr18:44,057,236–44,237,183	-

Genomic coordinates reported for the human reference genome are provided for build hg19. HGNC: Human Genome Organisation (HUGO) Gene Nomenclature Committee. OMIM: Online Mendelian Inheritance in Man database (https://www.omim.org/). The strand used for transcription, whether it is the sense (+) or the anti-sense (-) strand, is shown for each gene entry.

classed as benign. No rare variants in *ZEB1* or *SLC4A11* were identified in the FECD expansion-negative cohort.

Five rare variants were identified in four FECD expansion-positive cases (Table 3B). The two variants identified in *AGBL1*, NM_152336.4: c.806C>T, p.(Pro269Leu) and c.3323+2dup, which were classified as benign and VUS, respectively, according to ACMG criteria, were identified in expansion-positive patients 558 and 948, as well as in expansion-negative cases 758 and 589. Two variants were also identified in *LOXHD1* in expansion-positive cases. One of these, a benign variant, NM_144612.7: c.3269G>A, p.(Arg1090Gln), was identified in expansion-positive patient 778 and also in expansion-negative case 765. The other variant, c.2998C>T, p.(Arg1000Trp), which was VUS by ACMG criteria, affects an arginine residue that is not evolutionarily conserved among vertebrates (data not shown). One variant in *SLC4A11* was identified in expansion-positive case 958, NM_032034.4: c.2041G>A, p.(Ala681Thr), which

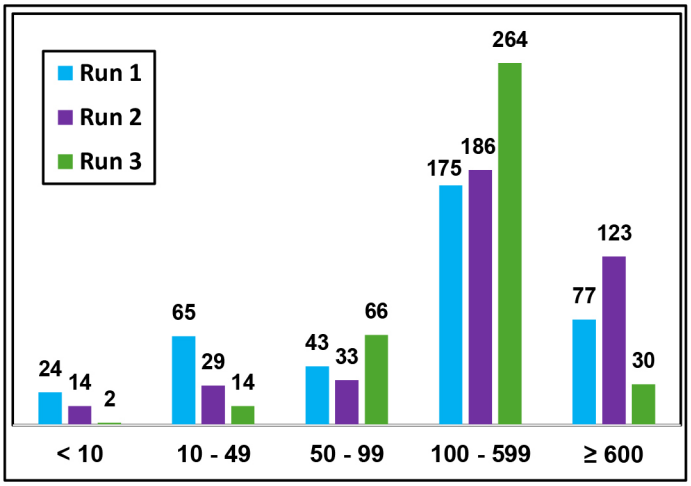
A

Mean Read Depth	Optimisation Run 1	Optimisation Run 2	Optimisation Run 3
≥ 600	77	123	30
100 – 599	175	186	264
50 – 99	43	33	66
10 – 49	65	29	14
< 10	24	14	2
Total	384	385	376

B

Excluded	11	9	N/A
Re-designed	12	None	N/A

C



D

Sample ID	Total reads sequenced	Usable reads	Duplicate reads (%)
1	5,659,798	225,547	96.01
2	4,827,222	172,060	96.44
3	6,049,174	194,342	96.79
4	6,546,854	144,493	97.79
5	4,607,876	140,574	96.95
Total	27,690,924		

Figure 2. Summary of smMIP optimization runs. Three optimization runs were performed on genomic DNA from five non-FECD control samples. **A**. Number of probes with mean read depth following each optimization run. **B**. Number of probes that were excluded and redesigned after each run. N/A, not applicable. **C**. Graphical representation of the number of probes (on the y-axis) and average read depth (on the x-axis) following each run. **D**. MiSeq data for the third optimization run for each of the five control samples, highlighting total reads sequenced, number of usable reads, and percentage of PCR duplicates. Note following each optimization run, the number of probes with an average read depth greater than 10 increased gradually from 94.8%, then 96.4% to 99.5%.

Table 2. Summary of cohort demographics following STR/TP-PCR genotyping.

Various	FECD affected		non-FECD
Number in cohort	78	36	29
CTG18.1 repeat expansion	YES	NO	NO
Mean age at recruitment (range)	71.5 (53 - 88)	68.8 (51 - 95)	76.1 (65 - 92)
Gender; F:M (%)	53: 25 (67.9%: 32.1%)	28: 8 (77.8%: 22.2%)	23: 6 (79.3%: 20.7%)

Mean age at recruitment in years and gender of the FECD cases and non-FECD controls, after classifying the cohort into those that harbor the expanded repeat of ≥50 copies, and those that do not. All the subjects are Caucasian. Note the individuals studied were all over 50 years of age at the time of recruitment. The mean age, and the age range, were similar between each sub-category and there was a consistent predominance of females in each group. A Fischer exact statistical test to compare the gender ratios between FECD cases (F: M, 81: 33) and non-FECD controls (23: 6), showed that the groups were well matched (p=0.4857).

was VUS by ACMG criteria. This variant, which affects an alanine residue in the protein, is not evolutionarily conserved among vertebrates (data not shown). Three rare variants were identified in 2 non-FECD controls (Table 3C) and were likely benign according to ACMG criteria. Two variants, both identified in *AGBL1*, NM_152336.4: c.1352C>A, p.(Ser451Tyr) and c.2457G>T, p.(Glu819Asp), were found in the same individual (patient 1149). The other variant in *SLC4A11*, NM_032034.4: c.1039C>T, p.(Arg347Trp) was identified in patient 1159.

A statistical comparison between the number of variants identified after filtration in each cohort highlighted the absence of variant enrichment in the FECD expansion-negative cases compared to the FECD expansion-positive cases ($p = 0.7612$) or when compared to the non-FECD controls ($p = 0.3275$; Table 4).

To look for copy number changes, the smMIP-generated targeted sequences were compared using ExomeDepth between FECD expansion-negative and FECD expansion-positive cases and non-FECD controls. The putative CNVs that were identified are presented in Appendix 6. All Bayesian factor scores were below 10, and none of the putative copy number changes could be distinguished from samples without the CNV following manual scrutiny in IGV, suggesting the putative CNV changes are likely to represent false-positive variant calls.

Familial cases: The 36 FECD expansion-negative cases included two families with multiple affected members. One consisted of three affected siblings (patients 572, 573, and 580), while the other was a sibling pair of affected cases (patients 807 and 818). All affected members of both families were tested and confirmed to be expansion negative. No rare variants passing filtration criteria or any shared haplotypes were found within these families in the genes screened.

DISCUSSION

This study describes the use of a smMIP reagent targeting the exons of five candidate genes, *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1*, potentially implicated in causing FECD, to screen a cohort of patients who lack a repeat expansion in *TCF4*, together with suitable control groups. Targeting the coding exons of *TCF4* in the smMIP reagent was not considered, as dominantly inherited Mendelian variants in *TCF4* result in haploinsufficiency of the protein and cause Pitt-Hopkins syndrome, a severe, congenital neurologic condition with characteristic facial features [46,47] that is distinct from late-onset FECD. Targeting the intronic CTG18.1 trinucleotide repeat sequence in *TCF4* was not attempted by smMIPs, since the oligonucleotide probes used in the current study were designed to amplify up to 115 nucleotides of the target sequence. In theory, if an appropriate probe is designed to span the repeat sequence, it would only allow for the detection of up to 38 copies of the trinucleotide repeat, assuming suitable software is used to analyze the output. However, this detection would only be similar in size to the STR assay results presented in the current study. This limitation in size detection makes the smMIP method unsuitable for detecting repeat expansions.

The smMIP reagent described here provided a comparatively low-cost, rapid screening strategy, with cost per sample approximately a fifth of that for whole-exome sequencing (WES). The method required only 200 ng of starting genomic DNA for library preparation, as it is based on PCR enrichment and scalable through the addition of a unique indexed barcode to allow pooling of multiple samples in a single sequencing run. The data generated required less data storage, were simpler to analyze than the exome sequence, and detected variants that were verified by Sanger sequencing

Table 3. Single nucleotide variants identified in FECD cases that did not have an expanded repeat (A), had an expanded repeat (B) and in non-FECD controls (C).

HGNC gene symbol	Sample ID	Genomic nomenclature	Transcript	c.Nomen	p.Nomen	dbSNP ID	ExAC frequency	CADD score	gnomAD frequency	Splice AI score	Franklin
(A)											
COL8A2	1068	chr1:g.36565837C>T	NM_005202.4	c.7G>A	p.(Gly3Arg)	rs115156902	0.007	15.17	0.005322	0.07	B
AGBL1	758	chr15:g.86800154C>T	NM_152336.4	c.806C>T	p.(Pro269Leu)	rs185896474	0.0083	31	0.01173	0.01	B
AGBL1	589	chr15:g.87531321dup	NM_152336.4	c.3323+2dup	NK	rs547400346	0.0015	16.4	0.001999	0.34	VUS
LOXHD1	831	chr18:g.44057151C>T	NM_001145472.3	c.3340G>A	p.(Gly1114Arg)	rs142931455	0.0018	17.26	0.001364	0.05	B
LOXHD1	596	chr18:g.44085948C>T	NM_144612.7	c.5545G>A	p.(Gly1849Arg)	rs780560784	4.36E-05	32	1.03E-05	0.79	LP
LOXHD1	817	chr18:g.44109166G>A	NM_144612.7	c.4504C>T	p.(Arg1502Trp)	rs200819355	0.0002	28.2	0.000405	0.05	VUS
LOXHD1	765	chr18:g.44137400C>T	NM_144612.7	c.3269G>A	p.(Arg1090Gln)	rs118174674	0.0096	20.2	0.01747	0.01	B
LOXHD1	596	chr18:g.44171980G>A	NM_144612.7	c.1570C>T	p.(Arg524Cys)	rs192376005	0.0019	33	0.003051	0	LB
(B)											
AGBL1	558	chr15:g.86800154C>T	NM_152336.4	c.806C>T	p.(Pro269Leu)	rs185896474	0.0083	31	0.01173	0.01	B
AGBL1	948	chr15:g.87531321dup	NM_152336.4	c.3323+2dup	NK	rs547400346	0.0015	16.4	0.001999	0.34	VUS
LOXHD1	558	chr18:g.44140109G>A	NM_144612.7	c.2998C>T	p.(Arg1000Trp)	rs199847981	0.0005	21.1	0.000841	0	VUS
LOXHD1	778	chr18:g.44137400C>T	NM_144612.7	c.3269G>A	p.(Arg1090Gln)	rs118174674	0.0096	20.2	0.01747	0.01	B
SLC4A11	958	chr20:g.3209766C>T	NM_032034.4	c.2041G>A	p.(Ala681Thr)	rs770312190	4.58E-05	24	5.96E-05	0.01	VUS
(C)											
AGBL1	1149	chr15:g.86807754C>A	NM_152336.4	c.1352C>A	p.(Ser451Tyr)	rs79186009	0.0009	25	0.000577	0	LB
AGBL1	1149	chr15:g.86940679G>T	NM_152336.4	c.2457G>T	p.(Glu819Asp)	rs200431794	0.0002	25.7	0.000147	0	LB
SLC4A11	1159	chr20:g.3211846G>A	NM_032034.4	c.1039C>T	p.(Arg347Trp)	rs138137682	0.001	21.6	0.001843	0	LB

Genomic coordinates reported for the human reference genome are according to build hg19. ACMG classification determined using Franklin. This classification system considers up to 28 specific criteria with varying degrees of strength for pathogenic or benign prediction. Combining the scores for these criteria, assigns each variant to one of five tiers. The variant can either be pathogenic (P), likely pathogenic (LP), variant of unknown significance (VUS), likely benign (LB) or benign (B). Nomen=Nomenclature. ExAc=Exome Aggregation Consortium, version 0.3, included all populations. CADD=Combined Annotation Dependent Depletion, version 1.3. NK=the consequence on the protein is not known.

Table 4. Statistical comparison to analyze variant enrichment in each cohort.

Various	FECD affected		non-FECD
CTG18.1 repeat expansion	YES	NO	NO
Number in cohort	29	33	29
Number of rare variants identified after filtration	5	8	3
P-value	0.7612 (NS)		0.3275 (NS)

A summary of the number of rare variants identified after filtration in each of the FECD and non-FECD cohorts is shown with the number of cases studied in each group. A Fischer exact test highlighted the absence of variant enrichment in the FECD expansion negative cohort when comparing to the other two groups (p=0.7612 and 0.3275). NS=not significant.

using an aliquot of the original genomic DNA sample. Other studies have used smMIPs to screen candidate genes, including testing 10 genes in focal epilepsies [48], looking for *ABCA4* variants in Stargardt's disease [37,49], screening 18 genes in patients with congenital glycosylation disorders [50], testing 113 genes in patients with retinitis pigmentosa and Leber congenital amaurosis [51], and testing 19 genes in the molecular diagnosis of patients with amelogenesis imperfecta [52].

This approach aimed to assess the relative contributions of rare, predicted pathogenic, dominant variants in *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1* to FECD causation in cases that do not appear to result from a *TCF4* repeat expansion, as well as compare variant frequencies in the expansion-negative FECD group with those in suitable control groups. Previous studies have supported the involvement of *COL8A2* variants in early-onset FECD [11,12] and *SLC4A11* variants in classical FECD [15,16]. In addition, a role has been suggested for dominant variants in *ZEB1* [20,21], *AGBL1* [22,23], and *LOXHD1* [24] in causing FECD. However, the lack of multiple, independent genetic studies confirming findings in *ZEB1*, *AGBL1*, and *LOXHD1* means the link between FECD and these genes remains unproven [34].

Screening *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1* in 36 expansion-negative FECD cases revealed eight rare variants that passed filtering criteria in seven of these patients. These consisted of one variant in *COL8A2*, two in *AGBL1*, and five in *LOXHD1*. Screening of the two control groups revealed five variants in four cases in the FECD expansion-positive group, as well as three variants in two cases in cataract patients, with three variants appearing in both the expansion-positive and expansion-negative FECD patient groups. Comparing the three groups by Fisher's exact test indicated no significant enrichment of variants that passed the filtration criteria in expansion-negative cases over either of the other groups (p = 0.7612 and 0.3275). Furthermore, only one variant (*LOXHD1* NM_144612.7, c.5545G>A, p.(Gly1849Arg)), found in an expansion-negative case, was classified as likely pathogenic. According to ACMG criteria, this variant is predicted to create a new splice acceptor site (Splice AI = 0.79) and is classified in the ClinVar database as being of uncertain significance. Other variants were classed as VUS (n = 4), likely benign (n = 4), or benign (n = 4), with no obvious clustering of VUS within any group.

The relatively small size of the cohort used in this study means that these findings should be interpreted with caution and cannot be taken to disprove the involvement of variants in any one gene. Nevertheless, these data are consistent with previous studies (reviewed by [34]) suggesting that pathogenic variants in *ZEB1*, *AGBL1*, and *LOXHD1* are rarely a cause of nonfamilial FECD. These data also show that repeat expansions in *TCF4*, together with rare variants in *COL8A2*, *SLC4A11*, *ZEB1*, *AGBL1*, and *LOXHD1*, do not account for FECD in all cases, implying a gap in our knowledge of the genetic basis of FECD. Of note, the FECD expansion-negative cases included two small families with multiple affected members. The fact that expansion-negative FECD is consistent within families provides further support for the hypothesis that another genetic cause or causes exist for FECD, although environmental effects cannot be ruled out. Similar findings were reported in a recent study [53] that used WES in 128 expansion-negative cases to show that only approximately 10% (13/128) carried rare, potentially deleterious variants in known FECD candidate genes.

Further contributions to FECD susceptibility could be genetic or environmental. Functional validation of the VUSs identified in this study may be an option to clarify their significance, although additional Mendelian variants in coding or noncoding regions may yet be discovered, particularly by studying familial cases, which are more likely to highlight highly penetrant alleles in novel genes. However, a polygenic form of FECD is also possible. Variants with an

allele frequency greater than 1% in the general population were excluded during filtering in this study, but genome-wide association studies (GWASs) of FECD have proved revealing. Since the original discovery through GWASs of the single-nucleotide polymorphism (SNP) rs613872 on chromosome 18 in an intron of *TCF4* and its connection with FECD [54], two further GWASs [55,56] have highlighted SNPs at 11 novel loci: *KANK4*, *LAMC1*, *LINC00970/ATP1B1*, *SSBP2*, *THSD7A*, *LAMB1*, *PIDD1*, *RORA*, *HS3ST3B1*, *LAMA5*, and *COL18A1*. WES or whole-genome sequencing of expansion-negative cases could identify rare Mendelian variants in these genes, while a GWAS using only expansion-negative FECD may help to clarify the relative contributions of these loci more fully and also offer the possibility of polygenic risk assessment. Alternatively, the versatility of the smMIP reagent allows the addition of specifically designed oligonucleotide probes spanning the GWAS SNPs as targets in the mega-pool. Likewise, as novel gene findings are reported, tiled oligonucleotides corresponding to the new targets could be included in the smMIP reagent.

A recent study identified several environmental risk factors contributing to FECD severity or age at onset of symptoms, including obesity, diabetes, and smoking [57]. Also, studies using mouse models have shown that increased corneal endothelial cell loss in females in response to ultraviolet light was driven in part by increased production of estrogen metabolites that caused DNA damage, which may explain the increased incidence of FECD in females [58,59]. Further studies are required to investigate whether any of these modulating factors could contribute to FECD onset in expansion-negative cases or protect against disease in individuals who have the expansion but do not manifest disease symptoms.

In summary, we have designed a smMIP-targeted capture reagent for variant screening of known gene candidates implicated in FECD causation. Although some DNA sequence variants were identified by this reagent, they were not significantly enriched in expansion-negative FECD cases compared with controls, suggesting that FECD in this group may have another cause or causes. Further work is therefore required to identify the cause of disease in the expansion-negative cases to account for the missing heritability in FECD.

APPENDIX 1. REGIONS WITH AN AVERAGE READ DEPTH BELOW 50 FOLLOWING OPTIMISATION RUN 3.

To access the data, click or select the words "Appendix 1." Genomic coordinates reported for the human reference genome are provided for build hg19. HGNC: Human Genome Organisation (HUGO) Gene Nomenclature Committee. Size is in nucleotides. The gene feature refers to the position of the region relative to the gene architecture.

APPENDIX 2. STR/TP-PCR GENOTYPING OF THE CTG18.1 LOCUS IN 33 FECD PROBANDS AND 3 ADDITIONAL FAMILY MEMBERS THAT DO NOT HAVE THE EXPANDED REPEAT.

To access the data, click or select the words "Appendix 2." Sample ID is shown with the number of repeats following STR/TP-PCR genotyping. A repeat size of more than, or equal to, 50 was considered an expanded repeat.

APPENDIX 3. STR/TP-PCR GENOTYPING OF THE CTG18.1 LOCUS IN 78 FECD CASES THAT HAVE AN EXPANDED REPEAT.

To access the data, click or select the words "Appendix 3." Sample ID is shown with the number of repeats following STR/TP-PCR genotyping. -: size unknown. Most cases are monoallelic for the expansion, apart from subjects with the genotype −/−, who are biallelic for the expansion. Twenty-nine FECD cases that were included in the smMIPs screen are indicated.

APPENDIX 4. STR/TP-PCR GENOTYPING OF THE CTG18.1 LOCUS IN 29 NON-FECD CONTROLS.

To access the data, click or select the words "Appendix 4." Sample ID is shown with the number of repeats following STR/TP-PCR genotyping. ND: No data. Greater than, or equal to, 50 repeats were considered expanded repeats. None of the controls that were tested had an expanded repeat.

APPENDIX 5. SANGER SEQUENCE VERIFICATION OF VARIANT CANDIDATES.

To access the data, click or select the words "Appendix 5." Chromatograms from expansion negative patients and normal control sequences for variant candidates identified in smMIPs. These included one variant in *COL8A2*, two in *AGBL1* and five variants in *LOXHD1*.

APPENDIX 6. PUTATIVE COPY NUMBER VARIANTS IDENTIFIED IN FECD CASES THAT DID NOT HAVE AN EXPANDED REPEAT (A), HAD AN EXPANDED REPEAT (B) AND IN NON-FECD CONTROLS (C).

To access the data, click or select the words "Appendix 6." HGNC gene symbol, sample ID, genomic coordinates of variant in hg19 version of the genome, the type of copy

number variant (CNV), transcript accession number, number of exons affected, the expected and observed reads and their ratio relative to each other, and the Bayesian factor (BF) score are shown. There was no evidence to support the CNVs when the samples were reviewed using the IGV software.

ACKNOWLEDGMENTS

The authors wish to thank the patients who were involved in this study. Author contributions: BA contributed to study design, acquired, analyzed and interpreted data and wrote the first draft of the manuscript. DA contributed to study design, interpreted data and commented on the manuscript draft. SS acquired and analyzed data and commented on the manuscript draft. WL analyzed data and commented on the manuscript draft. UH contributed to study design and commented on the manuscript draft. SA acquired data and commented on the manuscript draft. CFI contributed to study design, interpreted data and critically revised the manuscript draft. CW contributed to study design, interpreted data and critically revised the manuscript draft. MA contributed to study conception and design, analyzed and interpreted data and critically revised the initial manuscript draft for important intellectual content. All authors read and approved the final manuscript. Funding: This study was funded by a Saudi Arabian Government Scholarship to BA and an MRC Clinical Research Training Fellowship to SS (grant number G1002002/1). DA was funded by a Saudi Arabian Government Scholarship and UH was funded by a Leeds Doctoral Scholarship during this work. Conflict of interest statement: All other authors declare no competing financial interests of relevance to the contents of this manuscript.

REFERENCES

1. Stocker FW. The endothelium of the cornea and its clinical implications. Trans Am Ophthalmol Soc 1953; 51:669-786. [PMID: 13216798].

2. Lorenzetti DW, Uotila MH, Parikh N, Kaufman HE. Central cornea guttata. Incidence in the general population. Am J Ophthalmol 1967; 64:1155-8. [PMID: 6072991].

3. Bonanno JA. Identity and regulation of ion transport mechanisms in the corneal endothelium. Prog Retin Eye Res 2003; 22:69-94. [PMID: 12597924].

4. Gain P, Jullienne R, He Z, Aldossary M, Acquart S, Cognasse F, Thuret G. Global Survey of Corneal Transplantation and Eye Banking. JAMA Ophthalmol 2016; 134:167-73. [PMID: 26633035].

5. Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, Baratz KH. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. PLoS One 2012; 7:e49083[PMID: 23185296].

6. Mootha VV, Gong X, Ku HC, Xing C. Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. Invest Ophthalmol Vis Sci 2014; 55:33-42. [PMID: 24255041].

7. Zarouchlioti C, Sanchez-Pintado B, Hafford Tear NJ, Klein P, Liskova P, Dulla K, Semo M, Vugler AA, Muthusamy K, Dudakova L, Levis HJ, Skalicka P, Hysi P, Cheetham ME, Tuft SJ, Adamson P, Hardcastle AJ, Davidson AE. Antisense Therapy for a Common Corneal Dystrophy Ameliorates TCF4 Repeat Expansion-Mediated Toxicity. Am J Hum Genet 2018; 102:528-39. [PMID: 29526280].

8. Du J, Aleff RA, Soragni E, Kalari K, Nie J, Tang X, Davila J, Kocher JP, Patel SV, Gottesfeld JM, Baratz KH, Wieben ED. RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy. J Biol Chem 2015; 290:5979-90. [PMID: 25593321].

9. Mootha VV, Hussain I, Cunnusamy K, Graham E, Gong X, Neelam S, Xing C, Kittler R, Petroll WM. TCF4 Triplet Repeat Expansion and Nuclear RNA Foci in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci 2015; 56:2003-11. [PMID: 25722209].

10. Wieben ED, Aleff RA, Tang X, Butz ML, Kalari KR, Highsmith EW, Jen J, Vasmatzis G, Patel SV, Maguire LJ, Baratz KH, Fautsch MP. Trinucleotide Repeat Expansion in the Transcription Factor 4 (TCF4) Gene Leads to Widespread mRNA Splicing Changes in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci 2017; 58:343-52. [PMID: 28118661].

11. Biswas S, Munier FL, Yardley J, Hart-Holden N, Perveen R, Cousin P, Sutphin JE, Noble B, Batterbury M, Kielty C, Hackett A, Bonshek R, Ridgway A, McLeod D, Sheffield VC, Stone EM, Schorderet DF, Black GC. Missense mutations in COL8A2, the gene encoding the alpha2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. Hum Mol Genet 2001; 10:2415-23. [PMID: 11689488].

12. Gottsch JD, Sundin OH, Liu SH, Jun AS, Broman KW, Stark WJ, Vito ECL, Narang AK, Thompson JM, Magovern M. Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of fuchs corneal dystrophy. Invest Ophthalmol Vis Sci 2005; 46:1934-9. [PMID: 15914606].

13. Muragaki Y, Mattei M-G, Yamaguchi N, Olsen BR, Ninomiya Y. The complete primary structure of the human alpha 1 (VIII) chain and assignment of its gene (COL8A1) to chromosome 3. Eur J Biochem 1991; 197:615-22. [PMID: 2029894].

14. Jun AS, Meng H, Ramanan N, Matthaei M, Chakravarti S, Bonshek R, Black GCM, Grebe R, Kimos M. An alpha 2 collagen VIII transgenic knock-in mouse model of Fuchs endothelial corneal dystrophy shows early endothelial cell unfolded protein response and apoptosis. Hum Mol Genet 2012; 21:384-93. [PMID: 22002996].

15. Vithana EN, Morgan PE, Ramprasad V, Tan DT, Yong VH, Venkataraman D, Venkatraman A, Yam GH, Nagasamy S, Law RW, Rajagopal R, Pang CP, Kumaramanickevel G,

Casey JR, Aung T. SLC4A11 mutations in Fuchs endothelial corneal dystrophy. Hum Mol Genet 2008; 17:656-66. doi:10.1093/hmg/ddm337[PMID: 18024964].

16. Riazuddin SA, Vithana EN, Seet L-F, Liu Y, Al-Saif A, Koh LW, Heng YM, Aung T, Meadows DN, Eghrari AO, Gottsch JD, Katsanis N. Missense mutations in the sodium borate cotransporter SLC4A11 cause late-onset Fuchs corneal dystrophy. Hum Mutat 2010; 31:1261-8. [PMID: 20848555].

17. Park M, Li Q, Shcheynikov N, Zeng W, Muallem S. NaBC1 is a ubiquitous electrogenic Na+ -coupled borate transporter essential for cellular boron homeostasis and cell growth and proliferation. Mol Cell 2004; 16:331-41. [PMID: 15525507].

18. Vithana EN, Morgan P, Sundaresan P, Ebenezer ND, Tan DTH, Mohamed MD, Anand S, Khine KO, Venkataraman D, Yong VHK, Salto-Tellez M, Venkatraman A, Guo K, Hemadevi B, Srinivasan M, Prajna V, Khine M, Casey JR, Inglehearn CF, Aung T. Mutations in sodium-borate cotransporter SLC4A11 cause recessive congenital hereditary endothelial dystrophy (CHED2). Nat Genet 2006; 38:755-7. [PMID: 16767101].

19. Vilas GL, Loganathan SK, Quon A, Sundaresan P, Vithana EN, Casey J. Oligomerization of SLC4A11 protein and the severity of FECD and CHED2 corneal dystrophies caused by SLC4A11 mutations. Hum Mutat 2012; 33:419-28. [PMID: 22072594].

20. Riazuddin SA, Zaghloul NA, Al-Saif A, Davey L, Diplas BH, Meadows DN, Eghrari AO, Minear MA, Li YJ, Klintworth GK, Afshari N, Gregory SG, Gottsch JD, Katsanis N. Missense mutations in TCF8 cause late-onset Fuchs corneal dystrophy and interact with FCD4 on chromosome 9p. Am J Hum Genet 2010; 86:45-53. [PMID: 20036349].

21. Gupta R, Kumawat BL, Paliwal P, Tandon R, Sharma N, Sen S, Kashyap S, Nag TC, Vajpayee RB, Sharma A. Association of ZEB1 and TCF4 rs613872 changes with late onset Fuchs endothelial corneal dystrophy in patients from northern India. Mol Vis 2015; 21:1252-60. [PMID: 26622166].

22. Riazuddin SA, Vasanth S, Katsanis N, Gottsch JD. Mutations in AGBL1 cause dominant late-onset Fuchs corneal dystrophy and alter protein-protein interaction with TCF4. Am J Hum Genet 2013; 93:758-64. doi:10.1016/j.ajhg.2013.08.010[PMID: 24094747].

23. Zhang J, Wu D, Li Y, Fan Y, Chen H, Hong J, Xu J. Novel mutations associated with various types of corneal dystrophies in a Han Chinese population. Front Genet 2019; 10:881-[PMID: 31555324].

24. Riazuddin SA, Parker DS, McGlumphy EJ, Oh EC, Iliff BW, Schmedt T, Jurkunas U, Schleif R, Katsanis N, Gottsch JD. Mutations in LOXHD1, a recessive-deafness locus, cause dominant late-onset Fuchs corneal dystrophy. Am J Hum Genet 2012; 90:533-9. doi:10.1016/j.ajhg.2012.01.013[PMID: 22341973].

25. Eger A, Aigner K, Sonderegger S, Dampier B, Oehler S, Schreiber M, Berx G, Cano A, Beug H, Foisner R. DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene 2005; 24:2375-85. [PMID: 15674322].

26. Krafchak CM, Pawar H, Moroi SE, Sugar A, Lichter PR, Mackey DA, Mian S, Nairus T, Elner V, Schteingart MT, Downs CA, Kijek TG, Johnson JM, Trager EH, Rozsa FW, Mandal MN, Epstein MP, Vollrath D, Ayyagari R, Boehnke M, Richards JE. Mutations in TCF8 cause posterior polymorphous corneal dystrophy and ectopic expression of COL4A3 by corneal endothelial cells. Am J Hum Genet 2005; 77:694-708. [PMID: 16252232].

27. Lechner J, Dash DP, Muszynska D, Hosseini M, Segev F, George S, Frazer DG, Moore JE, Kaye SB, Young T, Simpson DA, Churchill AJ, Héon E, Willoughby CE. Mutational spectrum of the ZEB1 gene in corneal dystrophies supports a genotype-phenotype correlation. Invest Ophthalmol Vis Sci 2013; 54:3215-23. [PMID: 23599324].

28. Rogowski K, van Dijk J, Magiera MM, Bosc C, Deloulme J-C, Bosson A, Peris L, Gold ND, Lacroix B, Bosch Grau M, Bec N, Larroque C, Desagher S, Holzer M, Andrieux A, Moutin M-J, Janke C. A family of protein-deglutamylating enzymes associated with neurodegeneration. Cell 2010; 143:564-78. [PMID: 21074048].

29. Grillet N, Schwander M, Hildebrand MS, Sczaniecka A, Kolatkar A, Velasco J, Webster JA, Kahrizi K, Najmabadi H, Kimberling WJ, Stephan D, Bahlo M, Wiltshire T, Tarantino LM, Kuhn P, Smith RJ, Müller U. Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. Am J Hum Genet 2009; 85:328-37. [PMID: 19732867].

30. Frausto RF, Wang C, Aldave AJ. Transcriptome analysis of the human corneal endothelium. Invest Ophthalmol Vis Sci 2014; 55:7821-30. [PMID: 25377225].

31. Frausto RF, Le DJ, Aldave AJ. Transcriptomic Analysis of Cultured Corneal Endothelial Cells as a Validation for Their Use in Cell Replacement Therapy. Cell Transplant 2016; 25:1159-76. [PMID: 26337789].

32. Ali M, Khan SY, Kabir F, Gottsch JD, Riazuddin SA. Comparative transcriptome analysis of hESC- and iPSC-derived corneal endothelial cells. Exp Eye Res 2018; 176:252-7. Epub2018Sep6[PMID: 30196069].

33. Wieben ED, Aleff RA, Tang X, Kalari KR, Maguire LJ, Patel SV, Baratz KH, Fautsch MP. Gene expression in the corneal endothelium of Fuchs endothelial corneal dystrophy patients with and without expansion of a trinucleotide repeat in TCF4. PLoS One 2018; 13:e0200005[PMID: 29966009].

34. Tsedilina TR, Sharova E, Iakovets V, Skorodumova LO. Systematic review of SLC4A11, ZEB1, LOXHD1, and AGBL1 variants in the development of Fuchs' endothelial corneal dystrophy. Front Med (Lausanne) 2023; 10:1153122[PMID: 37441688].

35. Hardenbol P, Banér J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 2003; 21:673-8. [PMID: 12730666].

36. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 2013; 23:843-54. [PMID: 23382536].

37. Khan M, Cornelis SS, Khan MI, Elmelik D, Manders E, Bakker S, Derks R, Neveling K, van de Vorst M, Gilissen C, Meunier I, Defoort S, Puech B, Devos A, Schulz HL, Stöhr H, Grassmann F, Weber BHF, Dhaenens CM, Cremers FPM. Cost-effective molecular inversion probe-based ABCA4 sequencing reveals deep-intronic variants in Stargardt disease. Hum Mutat 2019; 40:1749-59. [PMID: 31212395].

38. Hitti-Malin RJ, Dhaenens CM, Panneman DM, Corradi Z, Khan M, den Hollander AI, Farrar GJ, Gilissen C, Hoischen A, van de Vorst M, Bults F, Boonen EGM, Saunders P, Roosing S, Cremers FPM. MD Study Group. Using single molecule Molecular Inversion Probes as a cost-effective, high-throughput sequencing approach to target all genes and loci associated with macular diseases. Hum Mutat 2022; 43:2234-50. [PMID: 36259723].

39. Panneman DM, Hitti-Malin RJ, Holtes LK, de Bruijn SE, Reurink J, Boonen EGM, Khan MI, Ali M, Andréasson S, De Baere E, Banfi S, Bauwens M, Ben-Yosef T, Bocquet B, De Bruyne M, de la Cerda B, Coppieters F, Farinelli P, Guignard T, Inglehearn CF, Karali M, Kjellström U, Koenekoop R, de Koning B, Leroy BP, McKibbin M, Meunier I, Nikopoulos K, Nishiguchi KM, Poulter JA, Rivolta C, Rodríguez de la Rúa E, Saunders P, Simonelli F, Tatour Y, Testa F, Thiadens AAHJ, Toomes C, Tracewska AM, Tran HV, Ushida H, Vaclavik V, Verhoeven VJM, van de Vorst M, Gilissen C, Hoischen A, Cremers FPM, Roosing S. Cost-effective sequence analysis of 113 genes in 1,192 probands with retinitis pigmentosa and Leber congenital amaurosis. Front Cell Dev Biol 2023; 11:1112270[PMID: 36819107].

40. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics 2014; 30:2670-2. Epub2014May26[PMID: 24867941].

41. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol 2011; 29:24-6. [PMID: 21221095].

42. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011; 43:491-8. [PMID: 21478889].

43. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38:e164[PMID: 20601685].

44. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17:405-24. [PMID: 25741868].

45. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, Wood NW, Hambleton S, Burns SO, Thrasher AJ, Kumararatne D, Doffinger R, Nejentsev S. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics 2012; 28:2747-54. [PMID: 22942019].

46. Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, Plouin P, Carter NP, Lyonnet S, Munnich A, Colleaux L. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet 2007; 80:988-93. [PMID: 17436254].

47. Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, Reardon W, Saraiva J, Cabral A, Gohring I, Devriendt K, de Ravel T, Bijlsma EK, Hennekam RCM, Orrico A, Cohen M, Dreweke A, Reis A, Nurnberg P, Rauch A. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet 2007; 80:994-1001. [PMID: 17436255].

48. Pippucci T, Licchetta L, Baldassari S, Marconi C, De Luise M, Myers C, Nardi E, Provini F, Cameli C, Minardi R, Bacchelli E, Giordano L, Crichiutti G, d'Orsi G, Seri M, Gasparre G, Mefford HC, Tinuper P, Bisulli F. Collaborative Group of Italian League Against Epilepsy (LICE) Genetic Commission. Contribution of ultrarare variants in mTOR pathway genes to sporadic focal epilepsies. Ann Clin Transl Neurol 2019; 6:475-85. [PMID: 30911571].

49. Khan M, Cornelis SS, Pozo-Valero MD, Whelan L, Runhart EH, Mishra K, Bults F, AlSwaiti Y, AlTalbishi A, De Baere E, Banfi S, Banin E, Bauwens M, Ben-Yosef T, Boon CJF, van den Born LI, Defoort S, Devos A, Dockery A, Dudakova L, Fakin A, Farrar GJ, Sallum JMF, Fujinami K, Gilissen C, Glavač D, Gorin MB, Greenberg J, Hayashi T, Hettinga YM, Hoischen A, Hoyng CB, Hufendiek K, Jägle H, Kamakari S, Karali M, Kellner U, Klaver CCW, Kousal B, Lamey TM, MacDonald IM, Matynia A, McLaren TL, Mena MD, Meunier I, Miller R, Newman H, Ntozini B, Oldak M, Pieterse M, Podhajcer OL, Puech B, Ramesar R, Rüther K, Salameh M, Salles MV, Sharon D, Simonelli F, Spital G, Steehouwer M, Szaflik JP, Thompson JA, Thuillier C, Tracewska AM, van Zweeden M, Vincent AL, Zanlonghi X, Liskova P, Stöhr H, Roach JN, Ayuso C, Roberts L, Weber BHF, Dhaenens CM, Cremers FPM. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. Genet Med 2020; 22:1235-46. [PMID: 32307445].

50. Abu Bakar N, Ashikov A, Brum JM, Smeets R, Kersten M, Huijben K, Keng WT, Speck-Martins CE, de Carvalho DR, de Rizzo IMPO, de Mello WD, Heiner-Fokkema R, Gorman K, Grunewald S, Michelakakis H, Moraitou M, Martinelli D, van Scherpenzeel M, Janssen M, de Boer L, van den Heuvel LP, Thiel C, Lefeber DJ. Synergistic use of glycomics and single-molecule molecular inversion probes for identification

of congenital disorders of glycosylation type-1. J Inherit Metab Dis 2022; 45:769-81. [PMID: 35279850].

51. Panneman DM, Hitti-Malin RJ, Holtes LK, de Bruijn SE, Reurink J, Boonen EGM, Khan MI, Ali M, Andréasson S, De Baere E, Banfi S, Bauwens M, Ben-Yosef T, Bocquet B, De Bruyne M, de la Cerda B, Coppieters F, Farinelli P, Guignard T, Inglehearn CF, Karali M, Kjellström U, Koenekoop R, de Koning B, Leroy BP, McKibbin M, Meunier I, Nikopoulos K, Nishiguchi KM, Poulter JA, Rivolta C, Rodríguez de la Rúa E, Saunders P, Simonelli F, Tatour Y, Testa F, Thiadens AAHJ, Toomes C, Tracewska AM, Tran HV, Ushida H, Vaclavik V, Verhoeven VJM, van de Vorst M, Gilissen C, Hoischen A, Cremers FPM, Roosing S. Cost-effective sequence analysis of 113 genes in 1,192 probands with retinitis pigmentosa and Leber congenital amaurosis. Front Cell Dev Biol 2023; 11:1112270[PMID: 36819107].

52. Hany U, Watson CM, Liu L, Nikolopoulos G, Smith CEL, Poulter JA, Antanaviciute A, Rigby A, Balmer R, Brown CJ, Patel A, de Camargo MGA, Rodd HD, Moffat M, Murillo G, Mudawi A, Jafri H, Mighell AJ, Inglehearn CF. Genetic Screening of a Nonsyndromic Amelogenesis Imperfecta Patient Cohort Using a Custom smMIP Reagent for Selective Enrichment of Target Loci. Hum Mutat 2025; 2025:8942542[PMID: 40741335].

53. Liu S, Sadan AN, Bhattacharyya N, Zarouchlioti C, Szabo A, Abreu Costa M, Hafford-Tear NJ, Kladny AS, Dudakova L, Ciosi M, Moghul I, Wilkins MR, Allan B, Skalicka P, Hardcastle AJ, Pontikos N, Bunce C, Monckton DG, Muthusamy K, Liskova P, Tuft SJ, Davidson AE. Genetic and Demographic Determinants of Fuchs Endothelial Corneal Dystrophy Risk and Severity. JAMA Ophthalmol 2025; 143:338-47. [PMID: 40079965].

54. Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W, Tran KD, Schmid-Kubista KE, Heckenlively JR, Swaroop A, Abecasis G, Bailey KR, Edwards AO. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med 2010; 363:1016-24. [PMID: 20825314].

55. Afshari NA, Igo RP Jr, Morris NJ, Stambolian D, Sharma S, Pulagam VL, Dunn S, Stamler JF, Truitt BJ, Rimmler J, Kuot A, Croasdale CR, Qin X, Burdon KP, Riazuddin SA, Mills R, Klebe S, Minear MA, Zhao J, Balajonda E, Rosenwasser GO, Baratz KH, Mootha VV, Patel SV, Gregory SG, Bailey-Wilson JE, Price MO, Price FW Jr, Craig JE, Fingert JH, Gottsch JD, Aldave AJ, Klintworth GK, Lass JH, Li YJ, Iyengar SK. Genome-wide association study identifies three novel loci in Fuchs endothelial corneal dystrophy. Nat Commun 2017; 8:14898-[PMID: 28358029].

56. Gorman BR, Francis M, Nealon CL, Halladay CW, Duro N, Markianos K, Genovese G, Hysi PG, Choquet H, Afshari NA, Li YJ, Gaziano JM, Hung AM, Wu WC, Greenberg PB, Pyarajan S, Lass JH, Peachey NS, Iyengar SK. VA Million Veteran Program. A multi-ancestry GWAS of Fuchs corneal dystrophy highlights the contributions of laminins, collagen, and endothelial cell regulation. Commun Biol 2024; 7:418-[PMID: 38582945].

57. Zwingelberg SB, Lautwein B, Baar T, Heinzel-Gutenbrunner M, von Brandenstein M, Nobacht S, Matthaei M, Bachmann BO. The influence of obesity, diabetes mellitus and smoking on fuchs endothelial corneal dystrophy (FECD). Sci Rep 2024; 14:11596-[PMID: 38773227].

58. Liu C, Miyajima T, Melangath G, Miyai T, Vasanth S, Deshpande N, Kumar V, Ong Tone S, Gupta R, Zhu S, Vojnovic D, Chen Y, Rogan EG, Mondal B, Zahid M, Jurkunas UV. Ultraviolet A light induces DNA damage and estrogen-DNA adducts in Fuchs endothelial corneal dystrophy causing females to be more affected. Proc Natl Acad Sci U S A 2020; 117:573-83. [PMID: 31852820].

59. Kumar V, Deshpande N, Parekh M, Wong R, Ashraf S, Zahid M, Hui H, Miall A, Kimpton S, Price MO, Price FW Jr, Gonzalez FJ, Rogan E, Jurkunas UV. Estrogen genotoxicity causes preferential development of Fuchs endothelial corneal dystrophy in females. Redox Biol 2024; 69:102986[PMID: 38091879].