

Multi-ancestry GWAS of Fuchs corneal dystrophy highlights roles of laminins, collagen, and endothelial cell regulation

Neal Peachey (✉ neal.peachey@va.gov)

Louis Stokes Cleveland VA

Bryan Gorman

VA Boston Healthcare System <https://orcid.org/0000-0002-4239-4672>

Michael Francis

Booz Allen Hamilton <https://orcid.org/0000-0002-1320-7161>

Cari Nealon

VA Northeast Ohio Healthcare System

Christopher Halladay

Providence VA Medical Center

Nalvi Duro

VA Boston Healthcare System

Kyriacos Markianos

VA Boston HealthCare System <https://orcid.org/0000-0003-0214-6014>

Giulio Genovese

Broad Institute <https://orcid.org/0000-0003-3066-5575>

Pirro Hysi

King's College London <https://orcid.org/0000-0001-5752-2510>

Hélène Choquet

Division of Research, Kaiser Permanente Northern California <https://orcid.org/0000-0001-9839-8667>

Natalie Afshari

UC San Diego, Shiley Eye Institute

Yi-Ju Li

Duke University <https://orcid.org/0000-0001-6996-4834>

J. Michael Gaziano

VA Boston Healthcare System

Adriana Hung

Vanderbilt University Medical Center

Wen-Chih Wu

Providence VA Medical Center

Paul Greenberg

Providence VA Medical Center

Saiju Pyarajan

Center for Data and Computational Sciences, Veterans Affairs Boston Healthcare System <https://orcid.org/0000-0002-9047-3762>

Jonathan Lass

Case Western Reserve University and University Hospitals Case Medical Center

Sudha Iyengar


Case Western Reserve University

Article

Keywords:

Posted Date: May 3rd, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2762003/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Fuchs endothelial corneal dystrophy (FECD) is a leading indication for corneal transplantation, but its molecular pathophysiology remains poorly understood. We performed genome-wide association studies (GWAS) of FECD in the Million Veteran Program (MVP) and meta-analyzed with the previous largest FECD GWAS, finding twelve significant loci (eight novel). We further confirmed the *TCF4* locus in admixed African and Hispanic/Latino ancestries, and found an enrichment of European-ancestry haplotypes at *TCF4* in FECD cases. Among the novel associations are low frequency missense variants in laminin genes *LAMA5* and *LAMB1* which, together with previously reported *LAMC1*, form laminin-511 (LM511). AlphaFold 2 protein modeling suggests that mutations at *LAMA5* and *LAMB1* may destabilize LM511 by altering inter-domain interactions or extracellular matrix binding. Finally, phenome-wide association scans and co-localization analyses suggest that the *TCF4* CTG18.1 trinucleotide repeat expansion leads to dysregulation of ion transport in the corneal endothelium and has pleiotropic effects on renal function.

Introduction

Fuchs endothelial corneal dystrophy (FECD) is the most common corneal dystrophy, affecting more than 5% of people older than 40 years of age, and is the leading indication for corneal transplantation (keratoplasty) in the United States¹. Globally, only 1 in 70 people needing a corneal transplant receive one², and a portion of transplants result in graft rejection or failure³, demonstrating the need for early prediction of genetic susceptibility, as well as pharmacologic alternatives, for treating FECD.

FECD is a progressive, bilateral disease^{4,5}. Earliest indications of FECD are the presence of excreted collagenous deposits called guttae⁶⁻⁸. As FECD progresses, guttae grow in numbers and merge, leading to a thickening of Descemet's membrane. These changes put stress on corneal endothelial cells (CECs), which regulate solute transfer and flow of water into the stroma. CECs begin to undergo cell death via apoptosis^{9,10}, accompanied by measurable changes in CEC shape and density¹¹, corneal biomechanics^{12,13}, and central corneal thickness (CCT)^{14,15}. Disruption of endothelium function leads to corneal edema, resulting in blurred vision and eventual severe vision loss.

The etiology of FECD involves complex interactions between incompletely penetrant genetic factors and biological and environmental factors¹⁶. Female sex and advanced age are established risk factors. Risk may also differ across populations; lower rates of FECD diagnosis have been observed in African Americans in both clinical settings¹⁷ and Medicare claims¹⁸. Similarly, examining FECD by genetic ancestry in the Department of Veterans Affairs Million Veteran Program (MVP), we found significantly reduced prevalence in participants of admixed African (AFR) and Hispanic/Latino (HIS) continental ancestries relative to European ancestry (EUR)¹⁹.

The first genetic risk factors for FECD were discovered in family studies²⁰ and included ultra-rare mutations in *COL8A2*²¹ and *SLC4A11*²². Subsequently, genome-wide association studies (GWAS) identified four risk loci for FECD. Of these, the most significant was common variation at 18q21.2²³ tagging the CTG18.1 trinucleotide repeat (TNR) expansion in an intron of *TCF4*²⁴ (transcription factor 4). As many as 75% of European-ancestry FECD cases had at least one expanded CTG18.1 allele²⁵. The previous largest FECD GWAS to date²⁶ confirmed *TCF4* and identified three additional loci: *LAMC1*, *KANK4*, and *ATP1B1*.

Recent genetic studies of FECD in non-European ancestries have largely focused on the genotyping and association of CTG18.1 alleles. CTG18.1 expansions are associated with FECD in African American¹⁷, Chinese²⁷, Indian²⁸, Japanese²⁹, and Thai³⁰ populations. However, CTG18.1 expansions are generally observed at relatively lower frequencies in non-European ancestry FECD patients compared to European-descent FECD patients^{17,31}. It remains unclear whether the population frequency of penetrant CTG18.1 alleles differs by genetic ancestry.

Here, we leverage genetic and clinical data provided by the MVP to conduct the largest GWAS analysis of FECD and first multi-ancestry GWAS meta-analysis. We confirm the four previously reported loci, including the presence of the *TCF4* GWAS locus in AFR and HIS, and present eight novel loci, expanding our knowledge of the genetic drivers of FECD.

Results

Multi-ancestry GWAS for FECD

We identified FECD cases in MVP participants of EUR, AFR, and HIS ancestry (Supplementary Table 1) following a clinically validated phenotyping algorithm¹⁹. Cases were mostly male (88.6%), reflecting the predominantly male composition of the MVP dataset³². As FECD is more common in women, there was a higher frequency of female cases vs. controls in each ancestry group (combined 11.4% cases vs. 8.4% controls). Mean age of FECD cases ranged from 62.8 [95% Confidence Interval (CI)=52.7, 72.9] in AFR to 70.5 [60.9, 80.1] years old in EUR. We performed a mixed-model GWAS for FECD using SAIGE³³ for each ancestry (Fig. 1, Supplementary Fig. 1), including age, age-squared, sex, and ten ancestry-specific principal components as covariates.

The *TCF4* locus reached genome-wide significance (GWS; $P < 5 \times 10^{-8}$) across all three ancestries analyzed in MVP; this was the first time *TCF4* has been significantly associated with FECD in a GWAS in AFR or HIS (Table 1; Supplementary Fig. 1a-c). The lead SNP at *TCF4*, rs11659764 ($r^2=0.21$ and $D'=0.97$ with the previously reported FECD index variant, rs613872), was the same across all three ancestries. Although the marker varied in frequency, the additive effect of each allele of rs11659764 on FECD was highly similar across ancestries, and we concur that the lower frequency of CTG18.1 TNR expansions observed in cases in AFR-descent populations contributes to a reduction in FECD risk¹⁷.

We applied local ancestry admixture mapping models at the *TCF4* locus in AFR and HIS to directly compare risk conferred by haplotype ancestry within the same individuals. In the AFR population, each EUR haplotype was additively associated with FECD (odds ratio (OR) = 1.28 [1.02, 1.61] ($P=0.015$), with 23% frequency of EUR haplotypes in cases vs. 18% in controls. In HIS, we found a similar OR for EUR haplotypes relative to AFR and Native American ancestry (NAT) haplotypes (OR=1.27 [0.91, 1.78]; $P=0.17$), with 64% EUR haplotype frequency in cases vs. 57% in controls, but this was non-significant due to lower power. Consistent with allele frequencies at our lead tagging SNP rs11659764, this result suggests that EUR haplotypes contain a higher frequency of pathogenic alleles compared to AFR and possibly also NAT haplotypes. The sample sizes for the MVP Asian cohorts were too low to obtain reliable estimates, but data from prior studies in Japanese cohorts and allele frequencies from the 1000 Genomes Project suggest that East Asians also have lower frequency of CTG18.1 expansions, contributing to a lower FECD risk³⁴.

Seven loci were GWS in the MVP EUR discovery cohort, including all four known FECD GWAS loci²⁶ (*TCF4*, *KANK4*, *LAMC1*, and *ATP1B1*) and three novel loci at *SSBP3*, *THSD7A*, and *PIDD1* (Supplementary Table 2; Supplementary Fig. 1a). Previously, a SNP at the *PIDD1* gene locus was reported to have reached suggestive significance²⁶ ($P=7 \times 10^{-7}$), while our lead novel variants at *SSBP3* and *THSD7A* were at least nominally significant ($P=2.61 \times 10^{-5}$ and $P=0.025$, respectively).

Table 1. Association of the top SNP at the *TCF4* locus, rs11659764, in MVP cohorts. In European (EUR), admixed African (AFR), and Hispanic/Latino (HIS) cohorts, rs11659764 had the most significant association with FECD. The minor allele increased odds ratio of FECD risk consistently across ancestry groups despite differences in effect allele frequency (EAF). CI confidence interval.

Ancestry	Odds ratio [95% CI]	P-value	EAF cases	EAF controls
EUR	6.41 [5.86, 7.01]	9.4×10^{-360}	0.222	0.045
AFR	7.57 [4.87, 11.75]	1.1×10^{-19}	0.061	0.009
HIS	7.16 [3.93, 13.04]	6.2×10^{-11}	0.131	0.022

We then performed inverse variance-weighted fixed effects meta-analyses, first exclusively across the two European cohorts, MVP EUR and Afshari et al.²⁶ (Supplementary Fig. 1d), and finally a multi-ancestry meta-analysis which added MVP AFR (HIS were excluded due to fewer than 100 cases; Fig. 1). This multi-ancestry meta-analysis tested a total of 18,302,074 variants in up to 3,970 cases and 333,794 controls (Supplementary Table 1), ~2.8 times the case sample size of the previous largest FECD GWAS cohort.

In the multi-ancestry meta-analysis, we replicated the 4 previously reported loci and identified a total of eight novel FECD loci emerging at GWS: *LAMA5*, *LAMB1*, *COL18A1*, *SSBP3*, *THSD7A*, *RORA*, *PIDD1*, and *HS3ST3B1* (Table 2; Fig. 2; Supplementary Fig. 2). Genomic control (λ) was 1.01, indicating minimal systematic inflation. Stepwise conditional and joint association analysis (COJO-slc) of the lead variant in each locus indicated no additional independent signals reaching GWS (*TCF4* was excluded due to the untyped CTG18.1 trinucleotide repeat).

As expected, the largest OR was observed at rs11659764 in *TCF4* (OR=7.15, [6.60, 7.74]; Supplementary Fig. 3). Effect sizes at index SNPs were consistent across MVP EUR and Afshari et al.²⁶, and all meta-analysis index SNPs were at least nominally significant ($P < 0.05$) in the prior report, further validating our phenotyping approach. Six of twelve index variants did not meet the meta-analysis minor allele frequency (MAF) cutoff of $\geq 1\%$ in AFR. Additionally, all index SNPs had consistent effect direction in AFR, with the exception of rs12439253 (*RORA*), which had a non-significant and opposite effect direction (Supplementary Fig. 4). Two AFR meta-analysis SNPs, rs1138714 in *PIDD1* and rs114065856 in *COL18A1*, had consistent direction with EUR cohorts but were not significant.

Linkage disequilibrium score regression (LDSC) analysis indicated the total liability-scale SNP heritability ($SNP-h^2$) for FECD, based on EUR meta-analysis summary statistics, was 0.43 (standard error [SE] = 0.32), assuming a 5% population prevalence. As LDSC generally measures polygenicity³⁵, the uncertainty of the $SNP-h^2$ estimate may reflect the partially monogenic (*TCF4*) architecture of FECD.

Table 2. Genome-wide significant loci in the multi-ancestry meta-analysis of FECD. Genomic risk loci from the meta-analysis of MVP European and African cohorts plus Afshari et al.²⁶. We identified eight novel FECD loci, and replicated all four previously reported loci. rs1138714 previously reached suggestive significance in Afshari et al.²⁶ at $P = 7 \times 10^{-7}$. Genomic coordinates correspond to GRCh38. EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; SE, standard error; OR [L95, U95], odds ratio with lower and upper bounds of 95% confidence interval; Direction, SNP effect direction in MVP EUR, MVP AFR, and Afshari et al.²⁶ meta-analysis cohorts, respectively. “?” indicates the AFR variant did not meet the allele frequency cutoff of 1% and was not included.

rsID	Chr:Pos	Predicted causal gene	EA/NEA	EAF	N case	N	OR [L95, U95]	P-value	Direction
Novel loci ($P < 5 \times 10^{-8}$)									
rs11590557	1:54,324,099	<i>SSBP3</i>	A/G	0.04	3,655	258,564	1.61 [1.43,1.81]	6.86×10^{-15}	+?+
rs74882680	7:11,700,254	<i>THSD7A</i>	G/A	0.02	3,655	258,564	1.72 [1.48, 2.00]	2.78×10^{-12}	+?+
rs150990106	7:107,955,927	<i>LAMB1</i>	A/G	0.02	3,655	258,564	1.75 [1.45, 2.10]	4.33×10^{-9}	+?+
rs1138714	11:825,110	<i>CD151,</i> <i>PIDD1</i>	G/A	0.53	3,970	337,764	1.22 [1.16, 1.28]	3.01×10^{-14}	+++
rs12439253	15:60,764,393	<i>RORA</i>	T/G	0.08	3,970	337,764	1.29 [1.18, 1.40]	4.31×10^{-9}	++
rs9303111	17:14,663,407	<i>HS3ST3B1</i>	C/A	0.32	3,970	337,764	0.81 [0.76, 0.85]	1.17×10^{-13}	--
rs141208202	20:62,322,048	<i>LAMA5</i>	T/C	0.05	3,655	258,564	1.40 [1.25, 1.57]	1.42×10^{-8}	+?+
rs114065856	21:45,432,844	<i>COL18A1</i>	T/C	0.04	3,970	337,764	0.61 [0.52, 0.72]	2.87×10^{-9}	--
Previously reported loci									
rs79742895	1:62,317,189	<i>KANK4</i>	C/T	0.04	3,655	258,564	1.78 [1.59, 1.98]	1.78×10^{-24}	+?+
rs1200114	1:169,091,251	<i>ATP1B1</i>	A/G	0.66	3,970	337,764	0.73 [0.69, 0.77]	5.38×10^{-34}	--
rs2093985	1:183,125,187	<i>LAMC1</i>	T/C	0.54	3,970	337,764	0.80 [0.76, 0.84]	2.58×10^{-18}	--
rs11659764	18:55,668,281	<i>TCF4</i>	A/T	0.05	3,655	258,564	7.15 [6.60, 7.74]	8.60×10^{-509}	+?+

Table 3. Summary of novel candidate genes. Candidate genes for eight novel loci identified in the multi-ancestry meta-analysis for FECD.

Novel candidate gene	Putative function
<i>SSBP3</i>	Likely binds to polypyrimidine promoter of <i>COL1A2</i> , regulating transcription.
<i>THSD7A</i>	Regulation of endothelial cell migration and adhesion via binding of integrin $\alpha\beta3$.
<i>LAMB1</i>	Beta-1 subunit of laminin-511, component of basal lamina.
<i>CDC151; PIDD1</i>	Global regulator of endothelial cell-cell and cell-matrix adhesion; may have a role in regulating corneal endothelial cell death via apoptosis.
<i>RORA</i>	Member of retinoic acid-related orphan receptors (RORs) family, which play a role in embryonic development of the eye.
<i>HS3ST3B1</i>	Regulates heparan sulfate, which may have a role in corneal homeostasis.
<i>LAMA5</i>	Alpha-5 subunit of laminin-511, component of basal lamina.
<i>COL18A1</i>	Collagen type XVIII subunit, is cleaved to form endothelial cell regulator endostatin.

Novel FECD candidate genes

We identified candidate genes for our eight novel GWAS loci in the biological context of FECD; these are summarized in Table 3. Two novel loci emerged with lead variants in laminin genes: *LAMA5* ($\alpha5$) and *LAMB1* ($\beta1$). Together with the previously reported *LAMC1* ($\gamma1$) protein, these subunits form the laminin-511 heterotrimer (LM511; also called laminin-10), implicating an important role for LM511 in CEC maintenance and FECD pathogenesis. In previous studies, LM511 staining patterns were thicker in FECD corneas than controls³⁶; additionally, LM511 facilitated the expansion of CECs in culture³⁷ and promoted recovery of CECs in animal models of CEC transplantation³⁸. At *LAMB1*, our association peak consisted of three low-frequency (1-2% in EUR) variants in LD ($r^2 > 0.9$; Supplementary Fig. 2), each with a posterior inclusion probability (PIP) of 30-35% estimated from sum of single effects (SuSiE) fine-mapping^{39,40}. Of these, the most likely causal variant is the missense mutation at rs80095409 (p.Arg795Gly), which was computationally predicted to have a deleterious impact on protein structure by both SIFT⁴¹ and Polyphen⁴² classifiers, with a Combined Annotation Dependent Depletion (CADD) score of 29.7⁴³ (Supplementary Fig. 5A). Interestingly, the *LAMB1* locus has no pleiotropy with other traits reported in the GWAS Catalog.

At *LAMA5*, the characteristic subunit of LM511, the lead variant rs141208202 is also a low-frequency (4-5% in EUR) missense mutation, p.Gly2156Glu, that was predicted by SIFT to be deleterious and had 78% PIP. The next most significant variant (rs143905087; $P = 6.74 \times 10^{-8}$), is an intronic variant in *CABLES2* in only moderate LD with the lead variant ($r^2 = 0.54$; Supplementary Fig. 5B) and 17% PIP. Thus, we prioritize rs141208202 as a likely causal variant at *LAMA5*, mediated through putative impact on protein structure, which we explore further below. However, rs141208202 is a *LAMA5* splicing quantitative trait locus (QTL) in some tissues in GTEx⁴⁴ and is located within a CTCF binding site⁴⁵, and thus may also have a regulatory impact.

We discovered two novel loci likely driven by collagen genes: *SSBP3* and *COL18A1*. *SSBP3* (single-stranded DNA binding protein 3) is predicted to bind to a polypyrimidine tract in the promoter of *COL1A2* and regulate its expression⁴⁶. Collagen

type I is one of the primary collagens found in corneal tissue, and other subunits of collagen type I have emerged in previous GWAS of corneal traits⁴⁷. *COL18A1* encodes the alpha chain of type XVIII collagen, a ubiquitous component of the basement membrane (BM). In addition to its structural role, cleavage of type XVIII collagen generates the regulatory peptide endostatin, which inhibits proliferation of endothelial cells through G1 arrest⁴⁸ and can induce cell death, implicating anti-tumorigenic and anti-angiogenic properties of this domain⁴⁹.

Another novel locus was identified at *THSD7A*. *THSD7A* interacts with integrin alpha V beta 3 ($\alpha v\beta 3$)^{50,51}, expressed on CECs⁵², to inhibit migration. *THSD7A* has been previously associated in GWAS studies with four ocular traits: glaucoma, intraocular pressure, refractive error, and cataract (Supplementary Table 3). Though the lead variants for these associations are in *THSD7A*, they have low r^2 with our lead SNP rs74882680 ($r^2 \leq 0.015$), due to multiple distinct LD blocks within this gene (Supplementary Fig. 2).

The lead variant of a gene-dense region at 11p15.5, rs1138714, corresponds to several potential candidate genes (Supplementary Fig. 2). *CD151*, whose gene product is a member of the tetraspanin family, is an essential global regulator of endothelial cell-cell and cell-matrix adhesion⁵³. Like *THSD7A*, *CD151* is involved in integrin binding via the formation of tetraspanin-enriched microdomains. Additionally, *CD151*, along with type XVIII collagen and laminins, are members of the collagen chain trimerization pathway. Another gene at 11p15.5, *PNPLA* (AKA desnutrin or TTS-2.2), is a paralogue with similar properties to *hGS2*, which is responsible for transferring fatty acids from triglycerides to retinol, as well as hydrolyzing retinylesters⁵⁴. Adequate retinol is required for corneal development and function, and CECs are involved in the conversion of retinol into retinoic acid⁵⁵. Finally, *PIDD1* also has a potential role in regulating CEC death via apoptosis. *PNPLA2* and *PIDD1* were differentially expressed in CEC in patients with keratoconus (KC) and myopia⁵⁶. The association of rs1138714 with eQTLs at all three of these biologically relevant genes in GTEx indicates that synchronized co-expression of multiple causal genes in this region may also be possible⁴⁴. This locus has been previously associated with multiple ocular traits, and our FECD index variant at rs1138714 is in strong LD with rs10902223 ($r^2=0.99$), reported as the lead variant for KC and intraocular pressure, and is also in moderate LD with rs4963153 ($r^2=0.54$), the lead variant reported for associations with corneal resistance factor (CRF) and CCT (Supplementary Table 3).

We identified a novel association with FECD at *RORA*, which belongs to the family of retinoic acid-related orphan receptors (RORs). RORs are a superfamily of nuclear receptor transcription factors which bind to hormone response units. Although it shares structural features with retinoic acid receptors (RARs), *RORA* does not have known ligand-binding properties with retinol. *RORA* is most commonly associated with regulation of *BMAL1* and circadian rhythm; CECs have a highly robust circadian clock, and FECD and other corneal maladies are known to exhibit diurnal variation⁵⁷. *RORA* is induced by oxidative stress; reduction of *NFE2L2* nuclear factor translocation, which leads to downregulation of antioxidant expression, has previously been observed in FECD cases⁵⁸. *RORA* also regulates the differentiation and maintenance of type-2 innate lymphoid cells^{59,60}, which are among the immune cells resident in the cornea⁶¹. Additionally, our top FECD index SNP at *RORA*, rs12439253, has $r^2=0.59$ with the KC index SNP rs76194223.

Finally, a novel FECD locus was found in an intergenic region ~314 Kbp downstream from the nearest coding gene, *HS3ST3B1*. *HS3ST3B1* is a 3-O-sulfotransferase integral membrane protein, which catalyzes the addition of sulfate groups to heparan sulfate (HS). HS is required for a wide range of cellular processes, including by CECs for maintaining corneal homeostasis⁶². Heparanase (HPSE), which acts as a protease of HS in the BM, was overexpressed in keratoconic corneas, and HPSE catalytic activity was correlated with KC severity⁶³. In addition, a severe impediment to corneal wound healing was observed in a mouse HS knockout model⁶². This locus has been previously associated with CCT in three GWAS studies ($r^2=0.95$ to 1) and CEC size variation coefficient ($r^2=0.63$).

Intriguingly, a locus near *ANAPC1* previously reported to account for 24% of variability in CEC density in an Icelandic population⁶⁴ reached suggestive levels of significance in our multi-ancestry meta-analysis. However, the allele reported to

decrease CEC density (rs78658973-A) was protective for FECD (OR=0.86 [0.80, 0.92]; $P=5.1\times 10^{-6}$). In the same study, this allele was also significantly associated with increased coefficient of cell size variation and decreased percentage of hexagonal cells. The other allele reported to decrease CEC density, the CTG18.1 TNR expansion, greatly increases risk of FECD (Supplementary Table 3). Thus, our results support a complex relationship between CEC density and FECD, with potential bidirectional causality.

Pleiotropy of FECD risk alleles

Index variant directional trends across four corneal traits

We compared the effect size and direction of our lead FECD variants with summary statistics from other corneal traits: KC⁴⁷, CCT^{65,66}, CRF⁶⁷, and corneal hysteresis (CH)⁶⁸. We found consistent directional trends in a variant-level comparison across these traits. Eight of twelve FECD index variants had nominally significant associations ($P<0.05$) in at least one other corneal trait. At the nominally significant variants for each respective trait, all KC and CCT variant effects were in the same direction as FECD, while all CRF variants, and all variants but one in CH (*SSBP3*) were associated with effects in the opposite direction, consistent with previous reports¹² (Fig. 3; Supplementary Table 4). Genetic correlations (r_g) between FECD and these traits were not significant, however they followed the same directional pattern as the variant-level trends.

Association of PGS Catalog-based polygenic scores with FECD status

We calculated polygenic scores (PGS) for every trait in the PGS Catalog⁶⁹, for all MVP EUR subjects, and performed a phenome-wide scan for the association of normalized PGS scores with FECD case-control status to discover shared genetic etiology (Supplementary Table 5). A total of 2,649 scores corresponding to 560 uniquely mapped traits in the Experimental Factor Ontology (EFO) were tested; we considered 24 traits to be significant after adjustment for multiple comparisons ($P<0.05/560$). We found that PGSs for other corneal traits had the strongest associations with FECD, including CH (OR=0.83 [0.79, 0.86]; $P=7.04\times 10^{-20}$) and CRF (OR =0.86 [0.83, 0.90]; $P=4.73\times 10^{-12}$). The negative effect direction of these corneal trait PGS associations is consistent with our variant-level analysis in Fig. 3. After corneal traits, several renal PGSs had significant associations with FECD status, including urinary albumin-to-creatinine ratio (UACR; OR=1.15 [1.10, 1.20]; $P=2.28\times 10^{-10}$), which was reported previously⁷⁰, plus urinary sodium (OR=0.89 [0.86, 0.93]; $P=8.80\times 10^{-8}$) and urinary potassium (OR=0.91 [0.87, 0.95]; $P=6.46\times 10^{-6}$).

PheWAS on FECD index SNPs

We performed phenome-wide association scans (PheWAS) on the index variant of each locus in the FECD meta-analysis (Supplementary Table 6). In up to 458,296 MVP EUR participants, a total of 1,460 phenotypes were tested for each SNP: 1,170 phecodes⁷¹, 64 laboratory and vital signs measurements, and 225 survey questions. We found 31 associations with non-cornea traits that were significant after correction for the total number of tests ($P<0.05/17,520$). Among the significant pleiotropic associations of FECD risk alleles are a protective association with open-angle glaucoma at *SSBP3*, risk-increasing associations with benign colon neoplasms at laminin genes *LAMA5* and *LAMC1*, and an association with increased heart rate at *LAMB1*, which is replicated in the UK Biobank ($P=2.0\times 10^{-8}$)⁷².

Pleiotropy and co-localization analyses of TCF4 demonstrate link to renal function

The most significant PheWAS associations were observed at the *TCF4* risk allele, which was strongly associated with laboratory measurements of increased serum bicarbonate ($P=7.0\times 10^{-62}$), decreased chloride ($P=9.1\times 10^{-24}$), and increased potassium ($P=2.3\times 10^{-9}$) (Supplementary Fig. 6). This pleiotropy at *TCF4* likely explains the robust association with increased serum bicarbonate we previously observed in a phenome-wide comorbidity scan of FECD case-control status¹⁹.

Upon further evaluation of the *TCF4* locus in these significant laboratory measurement traits, we found that the index SNP of each trait (rs11659764) was the same as in FECD, and each displayed a highly similar complex pattern of local associations (Supplementary Fig. 7), which in FECD are thought to be caused by the partial LD of SNPs on different haplotypes with the CTG18.1 TNR. Validating our results, the same pattern was observed with externally derived UACR summary statistics⁷⁰. We further found that the regression coefficients of significant SNPs at the *TCF4* locus were highly correlated across FECD and each of the four renal traits, suggesting co-localization. Positive correlation with FECD was observed in effect direction and magnitude for bicarbonate ($r=0.91$), potassium ($r=0.77$), and UACR ($r=0.95$), while negative correlation was observed with chloride ($r=-0.90$) (Supplementary Fig. 7).

In an effort to further untangle pleiotropic effects, we performed Bayesian colocalization analyses under the assumption of a single causal variant (the untyped CTG18.1 expansion) using coloc⁷³. All four traits showed evidence of colocalization, with posterior probabilities >0.999 (Supplementary Table 7). We consider these observations strong evidence that the CTG18.1 expansion has pleiotropic effects on renal function. Moreover, the strength of the association with serum bicarbonate suggests that the effect of the CTG18.1 expansion on FECD may be mediated through dysregulation of ion transport in CECs.

The gene products of two novel FECD loci at *LAMA5* and *LAMB1*, plus the known locus at *LAMC1* which we replicated, are the three subunits of the LM511 heterotrimer. Each monomer (α , β 1, and γ 1) of LM511 is a multi-domain polypeptide; these interact with each other to form the long arm of a cross-shaped structure, while their non-interacting portions constitute three short arms (Fig. 4A). The short arms are composed of laminin-type EGF-like (LE) domain repeats that terminate in a laminin N-terminal (LN) domain⁷⁴. These short arms interact with other extracellular proteins to assemble and stabilize the BM, while the long arms facilitate interaction with cell surface receptors via globular domains.

The missense mutations at rs141208202 (*LAMA5*) and rs150990106 (*LAMB1*), correspond to a glycine to glutamic acid substitution at position 2156 of α LE22, and an arginine to glycine substitution at position 795 of β 1 LE6, respectively. We examined the potential impact of these mutations on the structure and function of LM511, using SWISS-MODEL⁷⁵ and AlphaFold 2 (AF2)⁷⁶ to model the α LE22 and β 1 LE6 domains (Supplementary Fig. 8A).

The glycine to glutamic acid substitution in α LE22 replaces a small hydrophobic residue with a large acidic one, altering the surface hydrophobicity and topology (Fig. 4B, top). The required orientation of α LE22, with respect to the cross, positions the mutated residue in close proximity to the other chains. This substantial change in LE22 may disrupt inter-chain interactions, and can also potentially destabilize the triple-helix of the long arm, which leads to disrupted interactions with cell surfaces through allosteric modulation of the LG domains.

Replacing the large basic arginine in β 1 LE6 with a smaller hydrophobic glycine induces similar changes in surface hydrophobicity and topology (Fig. 4B, bottom). The wild-type arginine is part of a positive-negative-positive-negative patch on the LE6 domain surface that may present a binding motif. Breaking this motif can disrupt interactions to neighboring β 1 domains or to other extracellular matrix proteins, resulting in binding affinity differences to the basement membrane and altered cell signaling.

While these two mutations have high potential to disrupt inter-domain interactions, it is unlikely that they will induce significant changes to the tertiary structures of α LE22 and β 1 LE6. This is because LE domain backbones are covalently linked through four disulfide bonds that prevent any significant deviations from the native fold, resulting in no change in intra-hydrogen bond count for α LE22, and a loss of only three hydrogen bonds in β 1 LE6 (Supplementary Fig. 8B).

Correspondingly, Duet⁷⁷ predicted that the $\beta 1$ LE6 mutation is more destabilizing than the $\beta 5$ LE22 mutation. Overall, our structural analysis suggests that the variants associated with FECD may destabilize LM511 through altered inter-domain interactions, rather than through structural changes of the mutated domains.

Discussion

In this study we have identified eight novel genomic risk loci for FECD, and replicated the four existing loci, in the largest GWAS of FECD cases to-date ($N_{\text{cases}}=3,970$). Our multi-ancestry analysis confirmed the considerably large effect of the *TCF4* locus across AFR and HIS ancestries; *TCF4* was the exclusive signal reaching GWS in these ancestry groups with smaller case numbers. Our results increase confidence in known FECD mechanisms, and our novel candidate genes expand our understanding of the role of laminins, collagen, integrins, and endothelial cell regulation in FECD pathophysiology and the maintenance of corneal deturgescence.

All three genes encoding subunits of LM511 had GWS associations with FECD in this study. Previously, LM511 has been primarily associated with tumor growth, both *in vitro* and *in vivo*, typically in the context of integrin-mediated adherence to tumor cells⁷⁸. In a recent US cohort study where 68% of FECD cases were female, there were significantly higher risks in FECD cases for comorbidities with breast, thyroid, ovarian, and basal cell carcinomas⁷⁹. Our PheWAS results also indicate the index variants at *LAMC1* and *LAMA5* are significantly associated with colon cancer (Supplementary Table 6). These findings suggest a potential link between LM511 and the increased risk of certain cancers observed in FECD cases. Another novel FECD gene encodes CD151, which modulates binding between laminins and integrins; CD151 overexpression is a marker of poor breast cancer survival and may also contribute to LM511 tumor response⁷⁸.

Collagens are major components of the BM and Descemet's membrane, and the "infiltration" of collagenous secretion (guttae) from Descemet's membrane is a hallmark of FECD. Our GWAS results have associated *SSBP3*, which regulates COL1A2 (type I collagen), and *COL18A1* (type XVIII collagen) with FECD for the first time. Type XVIII collagen, which has been previously associated with Knobloch syndrome⁴⁹, is found in BMs of nearly every tissue as well as the stromal side of Descemet's membrane. Type XVIII collagen is also a HS proteoglycan; the product of another novel gene *HS3ST3B1* is responsible for generating binding sites for proteins on HS chains. FECD CEC samples have been previously shown to contain higher levels of keratan sulfate, a sulfated glycosaminoglycan (GAG) found in the ECM⁵⁸, and our results suggest HS-GAGs may have a similar (lubricating) role in FECD.

Our findings corroborate previous reports of correlation between corneal and renal traits. We found that the PGSs for the renal trait UACR, which has been associated with several retinal disorders, along with urinary sodium and urinary potassium, had highly significant associations with FECD, secondary only to other corneal traits (Supplementary Table 5). The *TCF4* locus has also been associated at GWS with these renal measures in GWAS Catalog. Similarly, previous experimental analysis of corneal endothelium in samples of FECD with the CTG18.1 TNR expansion found increased expression of genes involved in ion transport⁸⁰. Our analysis demonstrated for the first time a high probability of colocalized causal structure between FECD and renal-associated ionic measurements (Supplementary Fig. 7). This included a novel relationship of FECD with serum bicarbonate, which has been previously shown to mediate lactate-coupled water flux in the corneal endothelial pump⁸¹.

Furthermore, the novel FECD gene products of LM511 and COL181A play a role in FECD as well as kidney disease, and this may be due to concurrent embryonic patterning in renal and corneal endothelial cells; the kidney and eye develop around the same time during organogenesis around the fourth through sixth gestational weeks⁸². LM511 and *COL181A* are both highly expressed in tubule BMs. Polycystic kidney disease (PKD) cells have reduced LM511 and overexpressed LM332, and a mutation causing loss of *LAMA5* function induced PKD in a mouse model⁸³. The *LAMC1* locus has been previously associated in GWAS with several renal traits, including estimated glomerular filtration rate, type 2 diabetes, diastolic blood

pressure, and red blood cell count. This study also replicated a known FECD risk locus at *ATP1B1*, which is highly expressed in CECs, and whose gene product regulates sodium balance as a subunit of Na⁺/K⁺ ATPase.

Our analysis contains several limitations. First, the algorithm we used to identify FECD cases¹⁹, while clinically validated, was based solely on electronic health record diagnoses, and not the slit lamp imaging used previously²⁶, which may have diluted the phenotyping in our analysis. We were constrained by the demographics of FECD cases in the MVP dataset; FECD is more common in women, but our sample, and MVP in general, skew heavily male, which may have biased our GWAS towards the identification of male-specific genetic factors. We also did not differentiate between rare early-onset and more common late-onset FECD, whose pathophysiologies may involve separate genetic mechanisms²⁵.

It is well established that the most predictive FECD allele at 18q21.2 is the CTG18.1 TNR expansion⁸⁴. Our GWAS used chip-based genotyping, so we relied on SNPs tagging CTG18.1 alleles instead of direct genotyping. Although our lead *TCF4* SNP rs11659764 is an imperfect proxy for CTG18.1, it nonetheless showed a strong and consistent association signal across multiple ancestry groups.

AlphaFold 2 and SWISS-MODEL are accurate in single-state predictions, but a limitation arises as they provide no information on protein fluctuations, leading to lower confidence in the structure of intrinsically disordered regions (IDR). While crystal structures of homologs suggest that there are no significant IDRs in LM511 LE domains, the question of how mutations can affect their dynamics remains. Although the predicted single-state structures used here do not capture shifts in dynamics, they nonetheless inform that the mutations significantly change surface chemistry and topology, and by extension, interactions to binding partners. Additionally, AlphaFold 2 is trained on wild-type protein structures and therefore has limited ability to predict when missense mutations will cause changes in protein folding⁸⁵; however, the backbone structure of LM511 indicates folding changes are not likely to occur from as a result of our FECD risk alleles, and so this limitation should not impact our functional predictions.

Our GWAS results have added eight novel genomic risk loci to the previous total of sixteen genes known to be associated with FECD⁸⁴. We were able to place these novel loci into biological contexts that are compatible with currently understood mechanisms of FECD disease progression. Additionally, the MVP dataset enabled unprecedented quantitative analyses of non-EUR cohorts³², and this analysis expands our understanding of the shared genetic architecture of FECD in these populations. We hope these results can lead to improved genetic risk prediction, and once experimentally validated, will help inform modern treatment strategies.

Methods

Ethics/study approval. Informed consent was obtained from all participants, and all studies were performed with approval from the Institutional Review Boards at participating centers.

Phenotyping. We previously developed a FECD case-control algorithm based on structured electronic health record data and confirmed its accuracy by chart review at three VA Medical Centers¹⁹. We applied the algorithm to conduct genome-wide association analyses (GWAS) and to analyze associated electronic health record (EHR) data.

QC and imputation. MVP samples were genotyped on the Thermo Fisher MVP 1.0 Axiom array. The design and QC of the array is described in detail elsewhere⁸⁶. Genotypes were phased using SHAPEIT4⁸⁷ and imputed to the TOPMed reference panel (version R2) using Minimac4.

GWAS. Samples were classified according to genetic ancestry using the HARE (Harmonized Ancestry and Race/Ethnicity) method⁸⁸. GWAS analyses were performed on ancestry-stratified subsets in MVP using SAIGE v1.1.6.2³³, adjusting for sex, age, mean-centered age², and 10 ancestry-specific principal components. To ensure accurate effect size estimation, Firth

approximation was applied to SNPs with $P < 0.05$. Association scans were performed on well-imputed SNPs (INFO > 0.5) using an ancestry-specific minor allele frequency (MAF) cutoff of $\geq 0.1\%$.

GWAS meta-analysis. We performed inverse variance-weighted fixed effects meta-analyses of GWAS summary statistics. First, we performed a EUR GWAS meta-analysis of MVP EUR and Afshari et al.²⁶ summary statistics, followed by a multi-ancestry GWAS meta-analysis of MVP EUR, Afshari et al.²⁶, and MVP AFR summary statistics. (MVP HIS was excluded from the multi-ancestry meta-analysis due to containing <100 cases.) Each set of summary statistics was converted into GWAS-VCFs using the +munge plug-in (<https://github.com/freeseek/score>) of bcftools v1.16⁸⁹. The Afshari et al.²⁶ summary statistics were lifted over to the GRCh38 genome build using the +liftover plug-in. Finally, Meta-analyses were performed using the +metal plug-in. For the multi-ancestry meta-analysis, a cohort-specific MAF cutoff $\geq 1\%$ was applied.

Characterizing significant loci. We used the COJO-slc1 method implemented in GCTA⁹⁰ version 1.94.1 to find conditionally independent genome-wide significant secondary signals at significant loci. An LD reference panel was constructed from 100,000 randomly selected MVP EUR subjects. The *TCF4* locus was excluded from COJO analysis due to the association of the untyped CTG18.1 repeat expansion. Variants for each independent genomic risk locus in the multi-ancestry meta-analysis were clumped and lead variants were identified using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) web server (v1.4.2)⁹¹. The maximum P -value cutoff was set to 0.05, and a first LD threshold of $r^2 \geq 0.6$ and second threshold of $r^2 \geq 0.1$ were used to define loci and lead SNPs. The maximum distance between LD blocks to merge loci was 250 Kb. Pleiotropy of significant loci with previous GWAS traits was identified using GWAS Catalog via FUMA, and LDtrait⁹², using a 250 Kb range.

LDSC. Non-partitioned liability score heritability for FECD and pairwise genetic correlations (r_g) between FECD and ocular traits were computed using Linkage Disequilibrium score regression (LDSC) v1.0.1³⁵. Summary statistics for keratoconus⁴⁷, CCT^{65,66}, and CRF⁶⁷, were obtained from GWAS Catalog; corneal hysteresis (CH) was provided by the authors⁶⁸. Prior to computing r_g , all summary statistics were quality controlled and alleles were harmonized to the reference genome using MungeSumstats v1.7.8⁹³.

SuSiE fine-mapping. Genome-wide significant loci in the EUR meta-analysis were fine-mapped using SuSiE^{39,40} version 0.11.42. Pairwise SNP LD matrices were constructed from imputed dosages over the same sample set used in the MVP EUR GWAS (N=254,596) using LDSTORE 2.0. Default options were used, including the maximum number of causal variants at a locus (10). The *TCF4* locus was excluded from this analysis due to the association of the untyped CTG18.1 repeat expansion.

Associations of PGSs with FECD. Phenome-wide polygenic score files were obtained from European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) PGS Catalog⁶⁹. All European-ancestry subjects in MVP were scored across all available PGSs using the +score plugin (<https://github.com/freeseek/score>) of bcftools v1.16⁸⁹. PGSs were then loaded into the dosage format field of VCFs readable by SAIGE for association testing. To determine pleiotropy of genetic predisposition to traits on FECD, logistic regression was used to examine associations of PGSs on MVP EUR FECD cases and controls using SAIGE v1.1.6.2³³, adjusting for the same covariates as in GWAS (sex, age, mean-centered age², and 10 ancestry-specific principal components).

PheWAS of index SNPs. We performed a PheWAS on each individual index SNP using summary statistics generated from the August 2022 beta release of the genome-wide PheWAS (gwPheWAS) project in MVP. Genotypes were imputed using the African Genome Resource and 1000 Genomes imputation panels. Phenotypes were derived from phecodes following standard definitions⁷¹, a baseline survey distributed to all MVP enrollees, as well as EHR-based laboratory and vital signs measurements. A GWAS was performed on each SAIGE using sex, age, age-squared, and 10 principal components as covariates.

Colocalization. Genetic associations in MVP EUR participants at the *TCF4* locus (chr18:50,000,000-60,000,000) for serum bicarbonate, chloride, and potassium were obtained using PLINK 2.0 alpha 4⁹⁴. Traits were rank-based inverse normal transformed (RINT), and linear regression was performed using sex, age, age-squared, and 20 principal components. Genotype QC was performed as in the FECD GWAS described above. RINT urinary albumin-to-creatinine ratio (UACR) summary statistics⁷⁰ for EUR were obtained from GWAS Catalog, and lifted from hg19 to hg38. Effect comparison plots include only variants from chr18:54,500,000-56,500,000 with $P < 0.001$; effect correlation was measured using Pearson's r . Single causal variant colocalization was performed on summary statistics using the `colic.abf()` function in Coloc v5.2.1⁹⁵. A posterior probability > 0.9 for Hypothesis 4 (both traits are associated and share a single causal variant) was used as the criteria for colocalization.

Structural analysis. Because no known crystal structures of human $\alpha 5$ LE22 and $\beta 1$ LE6 exist, we modeled them using two structure prediction tools. SWISS-MODEL⁷⁵ was used to homology model the domains; the template with the highest GMQE score was selected. AI-based AlphaFold 2⁷⁶ (AF2) was used to supplement SWISS-MODEL for the missing portions in the homology-based template (Supplementary Fig. 8A). Structural differences between the SWISS-MODEL and the rat homolog, and those between SWISS-MODEL and AF2 predictions were both within the range of thermal fluctuations, lending confidence to the AF2 predictions. DUET was used to predict the change in protein stability due to the mutations⁷⁷.

Declarations

Data Availability

The full summary level association data from the meta-analysis and individual population association analyses in MVP will be available via the dbGaP study accession number phs001672.

Acknowledgements

We recognize the late Dr. Robert P. Igo (1965-2020) for his contributions in generating the summary statistics from Afshari et al. (2017) used in this work. This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by award I01 BX003364. This work was also supported by the Cleveland Institute for Computational Biology, NIH Core Grants (P30 EY025585, P30 EY011373), the Clinical and Translational Science Collaborative of Cleveland (UL1TR002548) from the National Center for Advancing Translational Sciences (NCATS) component of the NIH and NIH Roadmap for Medical Research, unrestricted grants from Research to Prevent Blindness to Case Western Reserve University, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, and the University of Buffalo, the Harper-Inglis Memorial for Eye Research, The Peierls Foundation, and That Man May See.

We thank the Veteran participants in MVP and MVP staff. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. We acknowledge the VA Million Veteran Program (MVP) and the VA-DOE genome-wide PheWAS core analytic team for generating the corresponding PheWAS summary statistics that were used in this manuscript. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

Competing interests

The authors have declared no competing interest.

Author contributions

Drafted the manuscript: B.R.G., M.F., N.S.P., S.K.I.

Analyzed the data: B.R.G., M.F., N.D., K.M., G.G.

Acquired the data: B.R.G., C.L.N., C.W.H., P.H., H.C., N.A.A., Y.-J. L., J.M.G., S.P., N.S.P., S.K.I.

Critically revised the manuscript for important intellectual content: all authors

References

1. Eye Bank Association of America. 2021 Eye Banking Statistical Report. (2022).
2. Gain, P. et al. Global Survey of Corneal Transplantation and Eye Banking. *JAMA Ophthalmol.* 134, 167–173 (2016).
3. Deng, S. X. et al. Descemet Membrane Endothelial Keratoplasty: Safety and Outcomes: A Report by the American Academy of Ophthalmology. *Ophthalmology* 125, 295–310 (2018).
4. Wilson, S. E. & Bourne, W. M. Fuchs' dystrophy. *Cornea* 7, 2–18 (1988).
5. Ong Tone, S. et al. Fuchs endothelial corneal dystrophy: The vicious cycle of Fuchs pathogenesis. *Prog. Retin. Eye Res.* 80, 100863 (2021).
6. Burns, R. R., Bourne, W. M. & Brubaker, R. F. Endothelial function in patients with cornea guttata. *Invest. Ophthalmol. Vis. Sci.* 20, 77–85 (1981).
7. Chiou, A. G. et al. Confocal microscopy in cornea guttata and Fuchs' endothelial dystrophy. *Br. J. Ophthalmol.* 83, 185–189 (1999).
8. Zoega, G. M. et al. Prevalence and risk factors for cornea guttata in the Reykjavik Eye Study. *Ophthalmology* 113, 565–569 (2006).
9. Borderie, V. M. et al. Corneal endothelial cell apoptosis in patients with Fuchs' dystrophy. *Invest. Ophthalmol. Vis. Sci.* 41, 2501–2505 (2000).
10. Li, Q. J. et al. The role of apoptosis in the pathogenesis of Fuchs endothelial dystrophy of the cornea. *Arch. Ophthalmol.* 119, 1597–1604 (2001).
11. McLaren, J. W., Bachman, L. A., Kane, K. M. & Patel, S. V. Objective assessment of the corneal endothelium in Fuchs' endothelial dystrophy. *Invest. Ophthalmol. Vis. Sci.* 55, 1184–1190 (2014).
12. del Buey, M. A., Cristóbal, J. A., Ascaso, F. J., Lavilla, L. & Lanchares, E. Biomechanical properties of the cornea in Fuchs' corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.* 50, 3199–3202 (2009).
13. Reinprayoon, U., Jermjutitham, M. & Kasetuwan, N. Rate of Cornea Endothelial Cell Loss and Biomechanical Properties in Fuchs' Endothelial Corneal Dystrophy. *Front. Med.* 8, 757959 (2021).
14. Kopplin, L. J. et al. Relationship of Fuchs endothelial corneal dystrophy severity to central corneal thickness. *Arch. Ophthalmol.* 130, 433–439 (2012).
15. Patel, S. V., Hodge, D. O., Treichel, E. J., Spiegel, M. R. & Baratz, K. H. Predicting the Prognosis of Fuchs Endothelial Corneal Dystrophy by Using Scheimpflug Tomography. *Ophthalmology* 127, 315–323 (2020).
16. Louttit, M. D. et al. A multicenter study to map genes for Fuchs endothelial corneal dystrophy: baseline characteristics and heritability. *Cornea* 31, 26–35 (2012).
17. Eghrari, A. O., Vahedi, S., Afshari, N. A., Riazuddin, S. A. & Gottsch, J. D. CTG18.1 Expansion in TCF4 Among African Americans With Fuchs' Corneal Dystrophy. *Invest. Ophthalmol. Vis. Sci.* 58, 6046–6049 (2017).
18. Mahr, M. A., Baratz, K. H., Hodge, D. O. & Erie, J. C. Racial/Ethnic Differences in Rates of Penetrating or Endothelial Keratoplasty for Fuchs Endothelial Corneal Dystrophy Among US Medicare Beneficiaries. *JAMA Ophthalmol.* 134, 1178–1180 (2016).

19. Nealon, C. L. et al. Association between Fuchs Endothelial Corneal Dystrophy, Diabetes Mellitus and Multimorbidity. *bioRxiv* (2022) doi:10.1101/2022.12.14.22283472.
20. Krachmer, J. H., Purcell, J. J., Jr, Young, C. W. & Bucher, K. D. Corneal endothelial dystrophy. A study of 64 families. *Arch. Ophthalmol.* 96, 2036–2039 (1978).
21. Gottsch, J. D. et al. Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of fuchs corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.* 46, 1934–1939 (2005).
22. Vithana, E. N. et al. SLC4A11 mutations in Fuchs endothelial corneal dystrophy. *Hum. Mol. Genet.* 17, 656–666 (2008).
23. Baratz, K. H. et al. E2-2 protein and Fuchs's corneal dystrophy. *N. Engl. J. Med.* 363, 1016–1024 (2010).
24. Wieben, E. D. et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One* 7, e49083 (2012).
25. Thaug, C. & Davidson, A. E. Fuchs endothelial corneal dystrophy: current perspectives on diagnostic pathology and genetics-Bowman Club Lecture. *BMJ Open Ophthalmol* 7, (2022).
26. Afshari, N. A. et al. Genome-wide association study identifies three novel loci in Fuchs endothelial corneal dystrophy. *Nat. Commun.* 8, 14898 (2017).
27. Xing, C. et al. Transethnic replication of association of CTG18.1 repeat expansion of TCF4 gene with Fuchs' corneal dystrophy in Chinese implies common causal variant. *Invest. Ophthalmol. Vis. Sci.* 55, 7073–7078 (2014).
28. Nanda, G. G., Padhy, B., Samal, S., Das, S. & Alone, D. P. Genetic association of TCF4 intronic polymorphisms, CTG18.1 and rs17089887, with Fuchs' endothelial corneal dystrophy in an Indian population. *Invest. Ophthalmol. Vis. Sci.* 55, 7674–7680 (2014).
29. Nakano, M. et al. Trinucleotide Repeat Expansion in the TCF4 Gene in Fuchs' Endothelial Corneal Dystrophy in Japanese. *Invest. Ophthalmol. Vis. Sci.* 56, 4865–4869 (2015).
30. Okumura, N. et al. Trinucleotide repeat expansion in the transcription factor 4 (TCF4) gene in Thai patients with Fuchs endothelial corneal dystrophy. *Eye* 34, 880–885 (2020).
31. Fautsch, M. P. et al. TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. *Prog. Retin. Eye Res.* 81, 100883 (2021).
32. Gaziano, J. M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* 70, 214–223 (2016).
33. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* 50, 1335–1341 (2018).
34. Nanda, G. G. & Alone, D. P. REVIEW: Current understanding of the pathogenesis of Fuchs' endothelial corneal dystrophy. *Mol. Vis.* 25, 295–310 (2019).
35. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295 (2015).
36. Byström, B. et al. Laminins in normal, keratoconus, bullous keratopathy and scarred human corneas. *Histochem. Cell Biol.* 127, 657–667 (2007).
37. Okumura, N. et al. Laminin-511 and -521 enable efficient in vitro expansion of human corneal endothelial cells. *Invest. Ophthalmol. Vis. Sci.* 56, 2933–2942 (2015).
38. Zhao, C. et al. Laminin 511 Precoating Promotes the Functional Recovery of Transplanted Corneal Endothelial Cells. *Tissue Eng. Part A* 26, 1158–1168 (2020).
39. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* 82, 1273–1300 (2020).
40. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. *PLoS Genet.* 18, e1010299 (2022).

41. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081 (2009).
42. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249 (2010).
43. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* 13, 1–12 (2021).
44. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330 (2020).
45. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122 (2016).
46. Bayarsaihan, D., Soto, R. J. & Lukens, L. N. Cloning and characterization of a novel sequence-specific single-stranded-DNA-binding protein. *Biochem. J* 331 (Pt 2), 447–452 (1998).
47. Hardcastle, A. J. et al. A multi-ethnic genome-wide association study implicates collagen matrix integrity and cell differentiation pathways in keratoconus. *Commun Biol* 4, 266 (2021).
48. Hanai, J.-I. et al. Endostatin causes G1 arrest of endothelial cells through inhibition of cyclin D1. *J. Biol. Chem.* 277, 16464–16469 (2002).
49. Heljasvaaraab R, Aikiocd M, Ruotsalainen H, Pihlajaniemia T. Collagen XVIII in tissue homeostasis and dysregulation – Lessons learned from model organisms and human patients. *Matrix Biol.* 57-58, 55–75 (2017).
50. Wang, C.-H. et al. Thrombospondin type I domain containing 7A (THSD7A) mediates endothelial cell migration and tube formation. *J. Cell. Physiol.* 222, 685–694 (2010).
51. Kuo, M.-W., Wang, C.-H., Wu, H.-C., Chang, S.-J. & Chuang, Y.-J. Soluble THSD7A is an N-glycoprotein that promotes endothelial cell migration and tube formation in angiogenesis. *PLoS One* 6, e29000 (2011).
52. Rayner, S. A., Gallop, J. L., George, A. J. T. & Larkin, D. F. P. Distribution of integrins $\alpha\beta 5$, $\alpha\beta 3$ and αv in normal human cornea: Possible implications in clinical and therapeutic adenoviral infection. *Eye* 12, 273–277 (1998).
53. Zhang, F. et al. Tetraspanin CD151 maintains vascular stability by balancing the forces of cell adhesion and cytoskeletal tension. *Blood* 118, 4274–4284 (2011).
54. Gao, J. G. & Simon, M. A comparative study of human GS2, its paralogues, and its rat orthologue. *Biochem. Biophys. Res. Commun.* 360, 501–506 (2007).
55. Nezzar, H. et al. Molecular and metabolic retinoid pathways in the human ocular surface. *Mol. Vis.* 13, 1641–1650 (2007).
56. McComish, B. J. et al. Association of Genetic Variation With Keratoconus. *JAMA Ophthalmol.* 138, 174–181 (2020).
57. Nakai, H. et al. Comprehensive Analysis Identified the Circadian Clock and Global Circadian Gene Expression in Human Corneal Endothelial Cells. *Investigative Ophthalmology & Visual Science* vol. 63 16 Preprint at <https://doi.org/10.1167/iovs.63.5.16> (2022).
58. Zhang, J. & Patel, D. V. The pathophysiology of Fuchs' endothelial dystrophy – A review of molecular and cellular insights. *Exp. Eye Res.* 130, 97–105 (2015).
59. Wong, S. H. et al. Transcription factor ROR α is critical for nuocyte development. *Nat. Immunol.* 13, 229–236 (2012).
60. Halim, T. Y. F. et al. Retinoic-acid-receptor-related orphan nuclear receptor alpha is required for natural helper cell development and allergic inflammation. *Immunity* 37, 463–474 (2012).
61. Liu, J. & Li, Z. Resident Innate Immune Cells in the Cornea. *Front. Immunol.* 12, 620284 (2021).
62. Coulson-Thomas, V. J. et al. Loss of Corneal Epithelial Heparan Sulfate Leads to Corneal Degeneration and Impaired Wound Healing. *Invest. Ophthalmol. Vis. Sci.* 56, 3004–3014 (2015).
63. García, B. et al. Heparanase Overexpresses in Keratoconic Cornea and Tears Depending on the Pathologic Grade. *Dis. Markers* 2017, (2017).

64. Ivarsdottir, E. V. et al. Sequence variation at ANAPC1 accounts for 24% of the variability in corneal endothelial cell density. *Nat. Commun.* 10, 1284 (2019).
65. Bonnemaier, P. W. M. et al. Multi-trait genome-wide association study identifies new loci associated with optic disc parameters. *Commun Biol* 2, 435 (2019).
66. Choquet, H. et al. A multiethnic genome-wide analysis of 44,039 individuals identifies 41 new loci associated with central corneal thickness. *Commun Biol* 3, 301 (2020).
67. Jiang, X. et al. Fine-mapping and cell-specific enrichment at corneal resistance factor loci prioritize candidate causal regulatory variants. *Commun. Biol.* 3, 762 (2020).
68. Simcoe, M. J., Khawaja, A. P., Hysi, P. G., Hammond, C. J. & UK Biobank Eye and Vision Consortium. Genome-wide association study of corneal biomechanical properties identifies over 200 loci providing insight into the genetic etiology of ocular diseases. *Human Molecular Genetics* vol. 29 3154–3164 Preprint at <https://doi.org/10.1093/hmg/ddaa155> (2020).
69. Lambert, S. A. et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* 53, 420–425 (2021).
70. Teumer, A. et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat. Commun.* 10, 4130 (2019).
71. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26, 1205–1210 (2010).
72. Elsworth, B. et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv* 2020.08.10.244293 (2020) doi:10.1101/2020.08.10.244293.
73. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383 (2014).
74. Aumailley, M. The laminin family. *Cell Adh. Migr.* 7, 48–55 (2013).
75. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303 (2018).
76. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
77. Pires, D. E. V., Ascher, D. B. & Blundell, T. L. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* 42, W314–9 (2014).
78. Pouliot, N. & Kusuma, N. Laminin-511: a multi-functional adhesion protein regulating cell migration, tumor invasion and metastasis. *Cell Adh. Migr.* 7, 142–149 (2013).
79. Xu, T. T., Baratz, K. H., Fautsch, M. P., Hodge, D. O. & Mahr, M. A. Cancer Risk in Patients With Fuchs Endothelial Corneal Dystrophy. *Cornea* 41, 1088 (2022).
80. Wieben, E. D. et al. Gene Expression and Missplicing in the Corneal Endothelium of Patients With a TCF4 Trinucleotide Repeat Expansion Without Fuchs' Endothelial Corneal Dystrophy. *Invest. Ophthalmol. Vis. Sci.* 60, 3636–3643 (2019).
81. Li, S., Shyam, R., Ogando, D. G. & Bonanno, J. A. Bicarbonate activates glycolysis and lactate production in corneal endothelial cells by increased pH. *Exp. Eye Res.* 199, 108193 (2020).
82. Bodaghi, B., Massamba, N. & Izzedine, H. The eye: a window on kidney diseases. *Clin. Kidney J.* 7, 337–338 (2014).
83. Zhang, Y., Reif, G. & Wallace, D. P. Extracellular matrix, integrins, and focal adhesion signaling in polycystic kidney disease. *Cell. Signal.* 72, 109646 (2020).
84. Zhang, J., McGhee, C. N. J. & Patel, D. V. The Molecular Basis of Fuchs' Endothelial Corneal Dystrophy. *Mol. Diag. Ther.* 23 97–112 (2019).
85. Buel, G. R. & Walters, K. J. Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* 29, 1–2 (2022).

86. Hunter-Zinck, H. et al. Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am. J. Hum. Genet.* 106, 535–548 (2020).
87. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10, 5436 (2019).
88. Fang, H. et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am. J. Hum. Genet.* 105, 763–772 (2019).
89. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* 10, (2021).
90. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 44, 369–375 (2012).
91. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* vol. 8 Preprint at <https://doi.org/10.1038/s41467-017-01261-5> (2017).
92. Lin, S.-H., Brown, D. W. & Machiela, M. J. LDtrait: An Online Tool for Identifying Published Phenotype Associations in Linkage Disequilibrium. *Cancer Res.* 80, 3443–3446 (2020).
93. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* 37, 4593–4596 (2021).
94. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).
95. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* 17, e1009440 (2021).

Figures

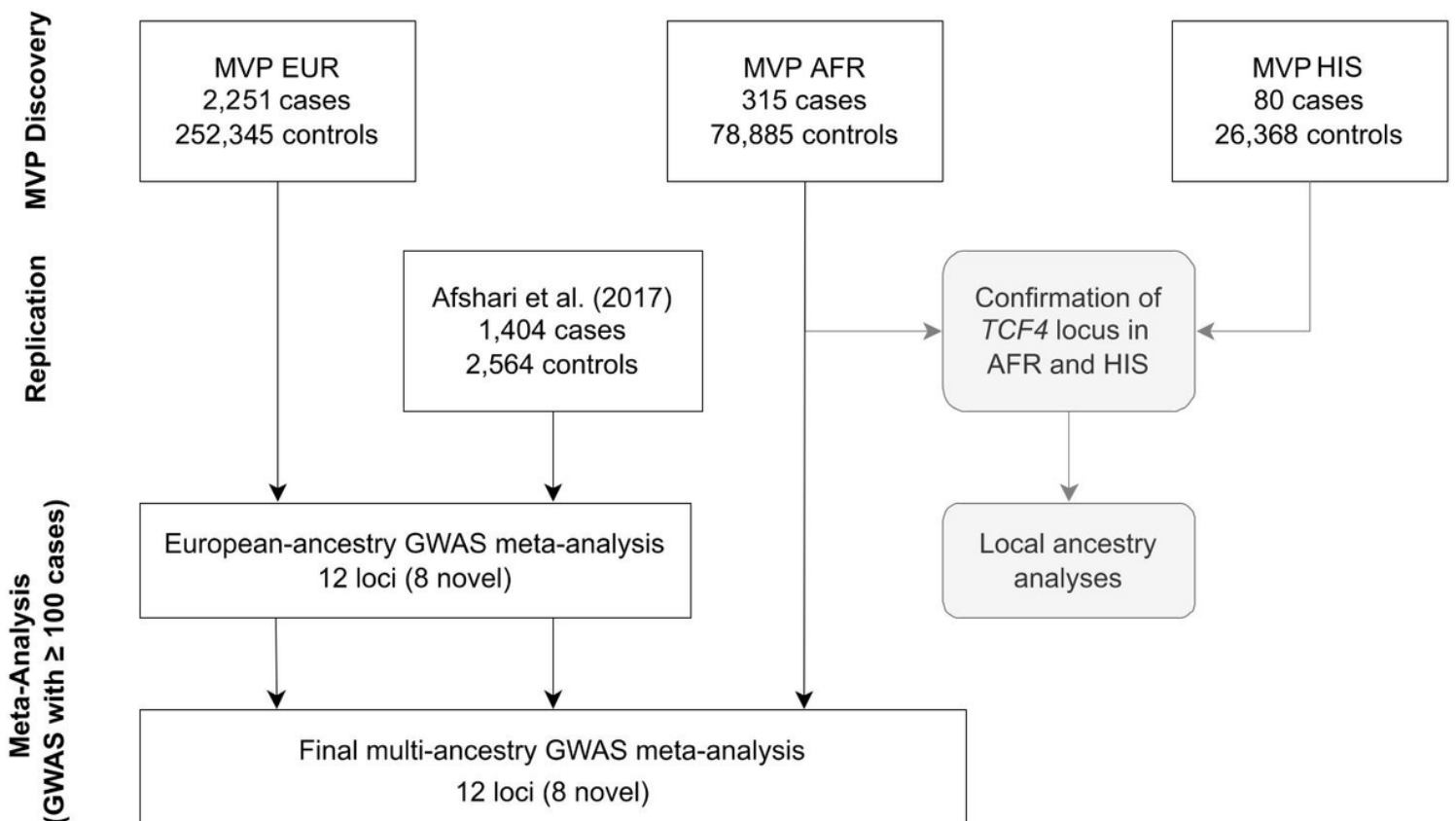


Figure 1

GWAS overview. Discovery analyses were performed in MVP European (EUR), admixed African (AFR), and Hispanic/Latino (HIS) cohorts. Numbers of FECD cases and controls are shown. Local ancestry analyses were performed in AFR and HIS cohorts. HIS were not included in multi-ancestry meta-analysis due to the low number of FECD cases. Afshari et al.²⁶ was included as a replication cohort.

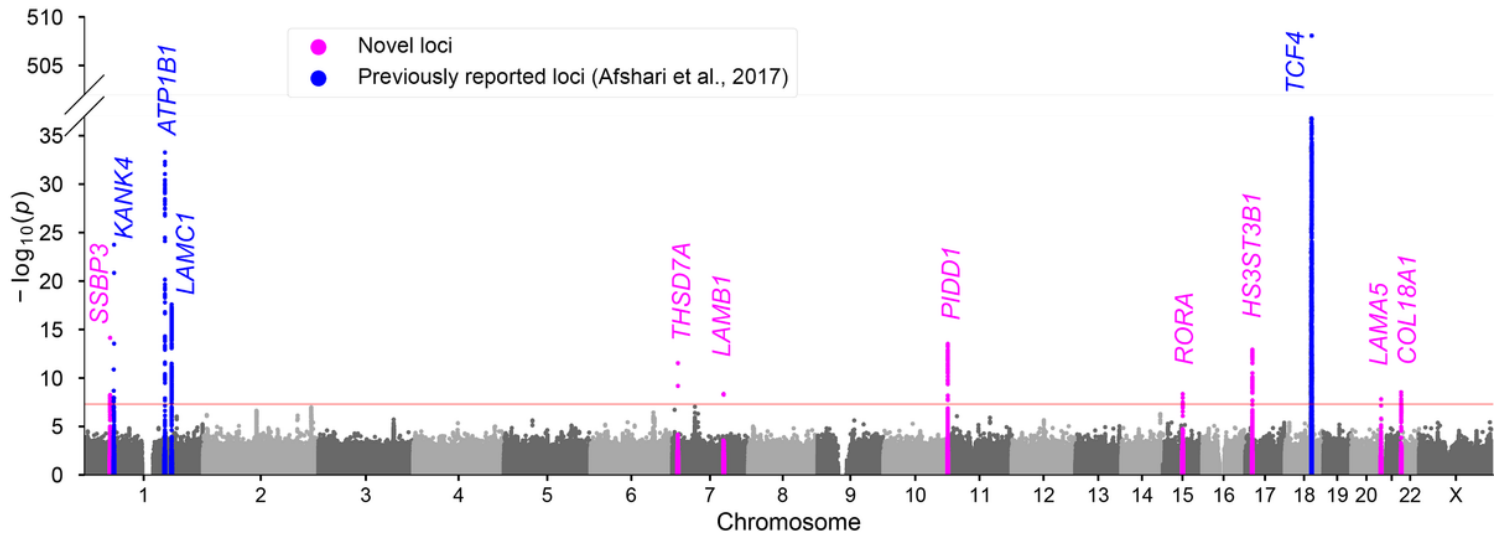


Figure 2

Manhattan plot of the multi-ancestry meta-analysis. Plot shows the $-\log_{10}(P)$ for associations of genetic variants with FECD across 22 autosomal chromosomes plus chromosome X. Significant loci are highlighted in magenta (novel) and blue (previously reported). The red line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$).

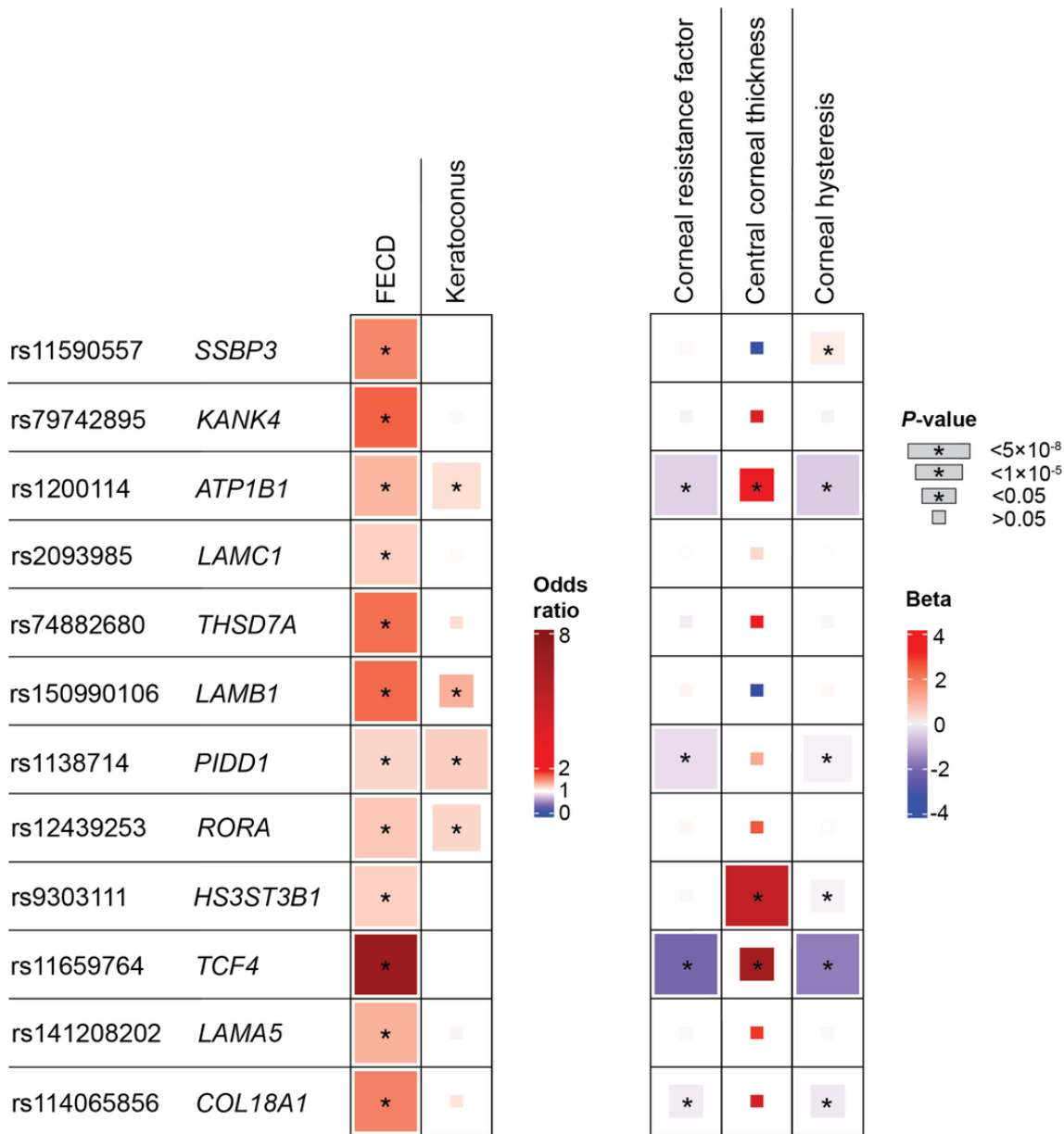


Figure 3

Comparing effects of FECD with four corneal traits. Variant-level comparison of lead FECD variants with four other corneal traits: keratoconus (KC), central corneal thickness (CCT), corneal resistance factor (CRF), and corneal hysteresis (CH). Box sizes correspond to P -value tiers, and * indicates $P < 0.05$. Units: FECD, odds ratio; KC, odds ratio; CCT, μm ; CRF, mm Hg; CH, mm Hg.

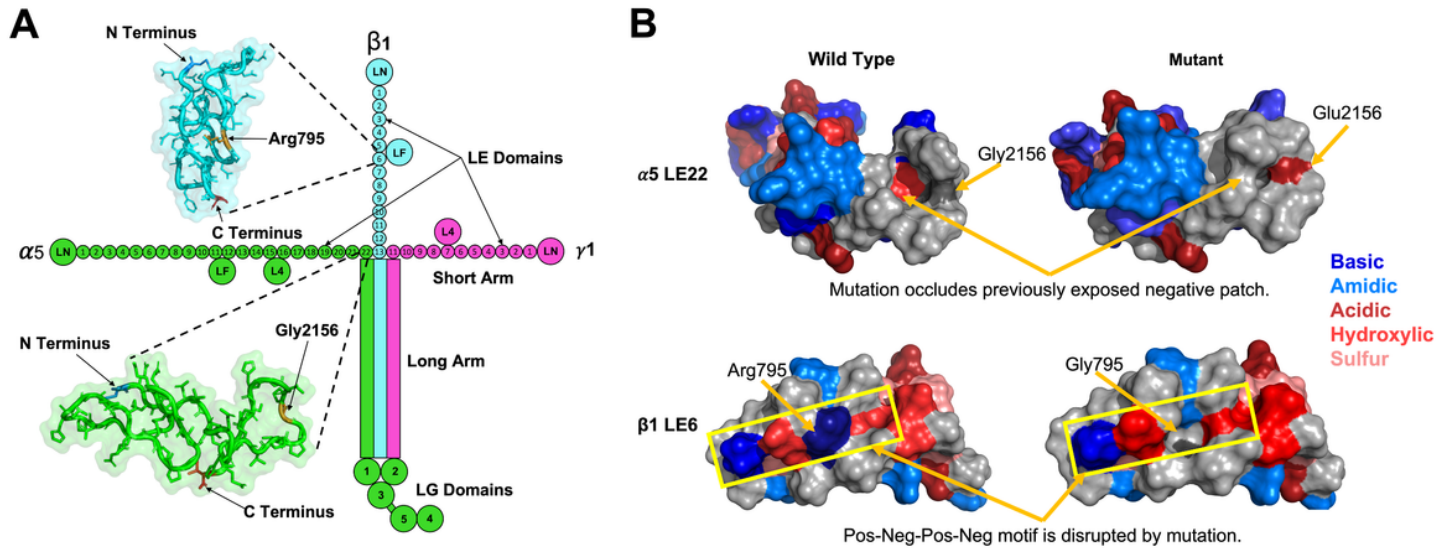


Figure 4

Structure of LM511 and predicted impact of missense variants in laminin genes *LAMA5* ($\alpha 5$) and *LAMB1* ($\beta 1$). (a) Structural organization of the laminin-511 (LM511) heterotrimer. The green color denotes the *LAMA5* subunit, blue denotes *LAMB1*, and pink denotes *LAMC1*. Significant FECD variants are located on the short arms of $\alpha 5$ and $\beta 1$, in LE (laminin-type epidermal growth factor (EGF) like) domains LE22 and LE6, respectively. The insets depict AlphaFold 2 predictions of these domains, and the locations of the residues are shown in orange. (b) Top: Predicted surface structure of the $\alpha 5$ LE22 domain with and without the Gly2156Glu variant. Bottom: Predicted surface structure of the $\beta 1$ LE22 domain with and without the Arg795Gly variant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplTable1.pdf](#)
- [SupplTable2.pdf](#)
- [SupplTable3.pdf](#)
- [SupplTable4.pdf](#)
- [SupplTable5.pdf](#)
- [SupplTable6.pdf](#)
- [SupplTable7.pdf](#)
- [FuchsSupplementaryFigures20230206.pdf](#)