Deciphering novel TCF4-driven molecular origins and mechanisms underlying a common triplet repeat expansion-mediated disease

Authors

Nihar Bhattacharyya[1], Nathaniel J Hafford-Tear[1*], Amanda N Sadan[1*], Anita Szabo[1*], Niuzheng Chai[1], Christina Zarouchlioti[1], Jana Jedlickova[2], Szi Kay Leung[3], Tianyi Liao[1], Lubica Dudakova[2], Pavlina Skalicka[2,4], Mohit Parekh[1], Aaron R Jeffries[3], Michael E Cheetham[1], Kirithika Muthusamy[5], Alison J Hardcastle[1,5], Nikolas Pontikos[1,5], Petra Liskova[2,4], Stephen J Tuft[1,5], Alice E Davidson[1,5†]

Affiliations

[1] University College London Institute of Ophthalmology, London, ECIV 9EL, UK

[2] Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[3] Faculty of Health and Life Sciences, University of Exeter, Exeter, EX4 4PY, UK

[4] Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[5] Moorfields Eye Hospital, London, EC1V 2PD, UK

[†] To whom correspondence should be addressed:

Dr Alice Davidson, UCL Institute of Ophthalmology, 11-43 Bath Street, London, UK

Email: alice.davidson@ucl.ac.uk

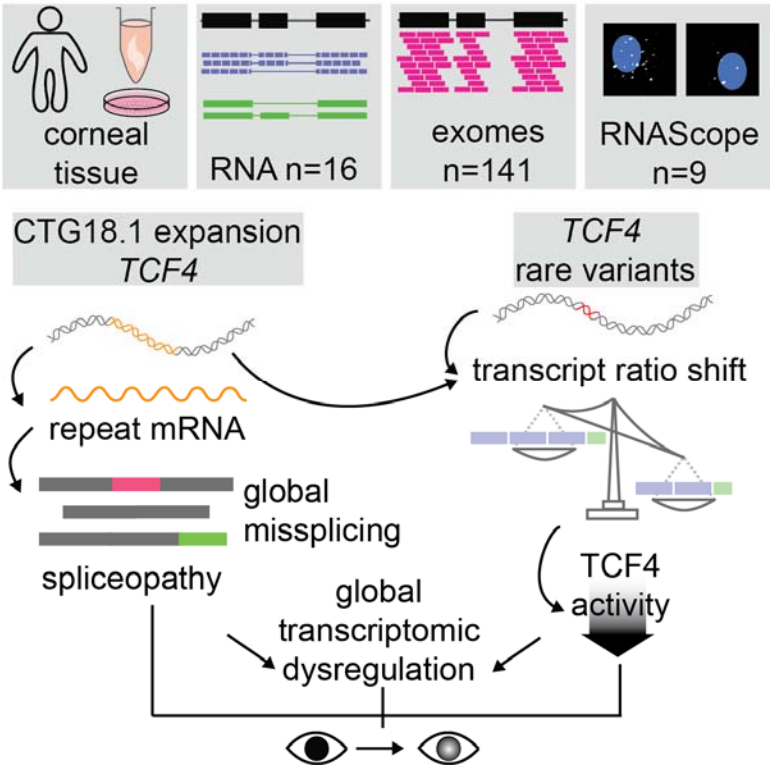
*These authors contributed equally to the manuscript

Keywords: TCF4 dysregulation, RNA toxicity, spliceopathy, triplet-repeat, repeat expansion-mediated disease, corneal dystrophy, Fuchs endothelial corneal dystrophy, *TCF4,* long-read sequencing, RNA-Seq, Iso-Seq, rare *TCF4* variants, isoform-specific dysregulation

Abstract:

The predominant cause of Fuchs endothelial corneal dystrophy (FECD) is a CTG repeat expansion (termed CTG18.1) situated within an intron of the transcription factor encoding gene, *TCF4*. Here we use a primary FECD case-derived corneal endothelial cell (CEC) system to enhance our understanding of multiple pathogenic processes underlying CTG18.1-mediated FECD. We define differential gene expression and alternative splice events using long- and short-read RNA-seq datasets generated from 15 biologically independent primary CEC lines, comprehensively characterizing aberrant splicing-related aspects of the disease. Further targeted analysis of these data, alongside a complementary spatial transcriptomics approach, reveals a unique and distinctive pattern of *TCF4*-specific dysregulation that underpins expansion-positive FECD and isolates downstream consequences of this shift as an important pathogenic component of disease. To explore whether *TCF4* dysregulation, irrespective of CTG18.1 expansions, could cause FECD we also interrogated exome data generated from a large (n=141) genetically refined cohort of CTG18.1 expansion-negative FECD cases. We identify four rare and predicted deleterious (minor allele frequency <0.005; CADD>15) *TCF4* variants all encompassed by only a small fraction (10/93) of total *TCF4* transcripts suggesting that, in rare instances, disruption of a specific subset of TCF4 isoforms may also confer CEC-specific disease independent of CTG18.1 expansions. In summary, our study supports the hypothesis that at least two distinct pathogenic mechanisms, RNA toxicity and *TCF4* isoform-specific dysregulation, underpin the pathophysiology of FECD. We anticipate these data will inform and guide the development of translational interventions for this common triplet-repeat mediated disease.

Graphical abstract



Introduction:

Fuchs endothelial corneal dystrophy (FECD) is a common, age-related, and visually disabling disease estimated to affect ~4.5% of people over the age of 50[1]. It primarily affects the corneal endothelium, a monolayer of post-mitotic cells on the posterior corneal surface that maintains relative corneal dehydration. As the disease progresses, collagenous excrescences (guttae) form on the thickened corneal endothelial basement membrane (Descemet membrane) and corneal endothelial cells (CECs) undergo accelerated cell death, which results in corneal oedema and opacity[2]. Affected individuals typically start to experience symptoms in their 5th to 6th decade, with glare, reduced contrast sensitivity, and blurred vision[3]. Transplantation of donor endothelial tissue is the preferred treatment of choice for cases with visually significant symptoms. However, given an increasing societal burden of age-related disease, coupled with a global shortage of suitable donor tissue, there is a pressing need for alternative treatments[2,4].

Up to 80% of FECD cases of European descent have one or more expanded copie(s) of a CTG trinucleotide repeat element (termed CTG18.1; MIM #613267) located within an intronic region of the transcription factor encoding gene *TCF4*[3,5]. We have previously demonstrated, in our large (n=450) FECD cohort, that an expanded copy of this CTG repeat (defined as ≥50 copies) confers >76-fold increased risk for developing FECD[6]. Expansion of non-coding repeat elements within the human genome have been identified to cause thirty distinct human diseases to date[7], including fragile X tremor ataxia syndrome (MIM #300623), myotonic dystrophy types 1 and 2 (MIM #160900 and #602668) and *C9orf72*-associated amyotrophic lateral sclerosis (ALS) and/or frontotemporal dementia (MIM #105550)[7]. Expansion-positive (Exp+) FECD represents by far the most common disease within this category. However, given it has so far only been established to only affect the cornea in a late-onset fashion, it occurs in contrast to the majority of much rarer non-coding repeat expansion disorders that typically have devastating outcomes on neurological and/or neuromuscular systems. Despite these differences, strong mechanistic molecular parallels exist between CTG18.1-mediated FECD and other non-coding repeat expansion diseases, whereby multiple pathogenic mechanisms occur concurrently to drive pathological outcomes[7,8].

We, and others, have demonstrated that the presence of one or more expanded CTG18.1 alleles consistently results in the accumulation of repeat-containing transcripts within the nuclei of CECs, sequestering splicing factors MBNL1/2 and dysregulating an array of splicing events[5,6,9]. However, a comprehensive understanding of the transcriptome-wide consequences of this process and its relevance to FECD pathophysiology and personalized therapies has yet to be realized[6,10,11]. Alongside RNA toxicity, repeat-associated non-AUG (RAN) dependant translation of repetitive $(CUG)_n$ and $(CAG)_n$ transcripts[12], dysregulation of *TCF4* itself has also been hypothesized to contribute to the underlying pathology of CTG18.1 Exp+ mediated FECD, but exactly how this non-coding repeat expansion drives dysregulation remains elusive[13,14]. This is largely because the gene comprises >90 alternative spatially and temporally expressed isoforms[13–15]. Such diversity is reflected in the varied functional roles TCF4 plays in human development, alongside the diverse range of human disease in addition to FECD associated with *TCF4* variants. For example, *TCF4* haploinsufficiency results in Pitt-

Hopkins syndrome (MIM #610954)[16,17] a severe neurodevelopmental disorder, whereas somatic *TCF4* mutations are commonly identified in Sonic hedgehog medulloblastoma[18] and common germline variants have been associated with Schizophrenia risk[19,20].

In this study, we used primary FECD case-derived CEC systems to enhance our understanding of the multiple and non-mutually exclusive pathogenic processes underlying CTG18.1-mediated FECD and identify novel patterns of *TCF4*-specific dysregulation that may, in part, explain the tissue specificity of the disease. In a minority of expansion-negative FECD cases we also identify rare *TCF4* variants that are predicted to induce aberrant expression or function of non-canonical *TCF4* isoforms that may also underlie disease in the absence of CTG18.1 expansions.

Materials and Methods:

Subject Recruitment, phenotyping and bio-sample processing

The study adhered to the tenets of the Declaration of Helsinki and was approved by the Research Ethics Committees of University College London (UCL) (22/EE/0090), Moorfields Eye Hospital (MEH) London (13/LO/1084), or the General University Hospital (GUH) Prague (2/19 GACR). All participants provided written informed consent and had either clinical signs of FECD or had a previous corneal transplant for FECD. For each participant who had planned endothelial keratoplasty, we collected the diseased corneal endothelium and Descemet membrane (DM) removed as part of the surgical procedure (MEH and GUH). We obtained control tissue as donor corneoscleral discs classed as unsuitable for clinical transplantation due to non-CEC related reasons (Miracles in Sight Eye Bank, Texas, USA) and removed a disc of tissue of similar dimensions to the clinical samples. We stored all excised tissues in either Eusol-C (Alchimia) or Lebovitz L-15 media (Life Technologies) before transferring to the laboratory for processing. Participant genomic DNA was extracted from whole blood (QIAgen Gentra Puregene Blood Kit), whereas control DNA was extracted from the remaining tissue of the corneoscleral disc (QIAgen Blood and Tissue kits).

CTG18.1 genotyping

We determined the CTG18.1 repeat length using a short tandem repeat (STR) genotyping assay as previously described in Zarouchlioti et al. 2018[6], originally adapted from Wieben et al. 2012[21]. Briefly, STR-PCR was performed using a 5'-FAM-conjugated primer (5' CAGATGAGTTTGGTGTAAGAT 3') upstream of the repeat and an unlabeled primer (5' ACAAGCAGAAAGGGGGCTGCAA 3') downstream. Post-PCR, products were combined with GeneMarker ROX500 ladder and separated on an ABI 3730 Electrophoresis capillary DNA analyzer (Applied Biosystems). Sizing and data analysis was performed on GeneMarker software (SoftGenetics). To distinguish between samples with presumed bi-allelic non-expanded alleles of equal length and samples with a larger repeat length beyond the detection limit of the STR-PCR, we performed a further triplet-primed PCR assay to confirm the presence or absence of a CTG18.1 allele above the detection range of the STR assay (approximately 125 repeats), following previously published methods[22].

Primary corneal endothelial cell (CEC) culture

We isolated and cultured FECD and control-derived CECs as described by Zarouchlioti et al. 2019[6], adapted from Peh et al. 2015[23]. Briefly, to dissolve the DM and detach the CECs, the tissue sample was incubated for 2 to 4 hours in 0.2% collagenase type I. After incubation, the CECs were resuspended in stabilization media (containing Human Endothelial-SFM (Life technologies) supplemented with 5% fetal bovine serum (FBS), 1% antibiotic/antimycotic, and 0.1% selective ROCK inhibitor Y-27632 (AdooQ BioSciences)), and then seeded into 6-well plates pre-coated with FNC coating mixture (Stratech). After 24 hours, we replaced the culture media with expansion media (Ham's F-12 Nutrient Mix with GlutaMAX Supplement (Life Technologies)/Medium 199 GlutaMAX Supplement (Life Technologies), 20 µg/mL ascorbic acid, 1% insulin-transferrin-selenium (Life Technologies), 5% FBS, 1% antibiotic/antimycotic, 10 ng/mL bFGF (Life Technologies) and 0.1% selective ROCK inhibitor Y-27632 (AdooQ BioSciences)). All cells were cultured at 37ºC and 5% CO_2 with twice-weekly media changes.

RNA extraction

Total RNA was extracted from the primary CEC lines using a NucleoSpin RNA XS kit (Macherey-Nagel), in accordance with the manufacturer's protocol. The RNA was eluted and stored at -80°C prior to sequencing.

Short-read RNA-seq

Stranded RNA-seq libraries were prepared by BGI Genomics using the TruSeq RNA Library Prep Kit (Illumina) following the manufacturer's protocol. We sequenced these with either a HiSeq 4000 platform (Illumina) with short-read strand-specific library prep with poly A enrichment (Illumina). Paired-end reads ([insert accession number here]) were filtered to remove low-quality reads (Q ≤5) or those with >10% of N base calls. Adaptor sequences were also removed during data quality analysis. We checked the quality of the resulting fastq files for the short-read data using FastQC Quantification[24]. Reads were aligned to the human genome (hg38) using STAR[25] (v2.7.10a). After alignment, hierarchical clustering was performed using PoiClaClu[26]. We also performed principal component analysis (PCA) using PCAtools[27] (v2.2.0) to determine global sample clustering.

We analyzed differential gene expression using DESeq2[28] (version 1.8.2), with independent hypothesis weighting (IHW) testing of the results to generate FDR values[29,30], and $\log_2$fold change was calculated using the apeglm (v 1.2) package[31]. We then determined significance as FDR corrected-p value (padj) <0.05. In addition, we analyzed different exon usage amongst the subgroups with DEXSeq2[32] (v 1.17.28). Significant results were filtered using an FDR-corrected P-value of <0.05.

We assessed differential splicing at a junction-based level with rMATS-turbo 4.1.2[33] with the Ensembl hg38 release 105 GTF file as reference. Default rMATS settings were used, including the 0.01% splicing difference cut-off in testing the null hypothesis. For this study, we focused on splicing events that are fully/partly annotated by Ensembl. Significant alternative

splicing events were identified as an absolute delta percent spliced in (psi) magnitude ≥ 0.1 and FDR ≤ 0.05. rMATs output was subsequently analyzed using rMAPS2[34] and maser 1.12.1[35].

Long-read RNA-seq

PacBio Sequel long-read transcriptomic sequencing (Iso-Seq)[36] was applied to generate full-length cDNA sequences from control and FECD Exp+ primary CEC-derived RNA samples. Raw reads ([insert database accession number here]) were processed and de-multiplexed with Iso-Seq (v3) and merged for transcript collapse using Cupcake[36] (parameters: "-c 0.85 -i 0.95 --dun-merge-5-shorter"). High-quality, full-length transcripts from the merged dataset were then aligned to the human reference genome (hg38) using Minimap2[37] (v2.17, parameters: "-ax splice -uf --secondary=no -C5 -O6,24 -B4") and annotated using SQANTI3[38] (v7.4). SQANTI filtering of artifacts (intra-priming and reverse transcriptase template switching) was then applied.

RNAScope detection of TCF4 transcripts

We seeded primary FECD case-derived CECs on chamber slides (Lab-tek) and confluent CEC cultures were stained using standard RNAScope protocols with minor adjustments (ACD Bio). We used the Hs-TCF4-C2 (cat. 557411) probe set to detect all *TCF4* isoforms, as well as a custom set of *TCF4* probes to detect all 5' exons in the *TCF4* transcript NM_001083962.2 excluding all 3' exons contained in the shorter transcript NM_001243234.2. Akoya TSA Cy3 fluorophores were used to visualize probe staining. Negative control probes (cat. 320871) were provided by ACD-Bio to verify specificity. All CEC lines analyzed by RNAScope were also tested for RNA foci in parallel via fluorescence *in situ* hybridization (FISH) using a CUG-specific probe (Cy3-$(CAG)_7$)[6]. A minimum of 100 cells and four images were acquired for each cell line and each probe.

Exome Sequencing

Exome sequencing libraries were generated using a SureSelect Human All Exome V6 capture kit (Agilent, USA) or SeqCap EZ MedExome Enrichment Kit (Roche) and sequenced on a HiSeq 4000 or HiSeq 2500 platform (Illumina). All raw sequencing data were aligned and annotated following our previously published methods[39]. In brief, we aligned sequencing reads using Novoalign (version 3.02.08), and variants and indels were called according to Genome Analysis Toolkit (GATK) (version 4.1.7.0) best practices (joint variant calling followed by variant quality score recalibration[40]). We annotated the variants using the Variant Effect Predictor[41]. The resulting variants were filtered on predicted deleteriousness using CADD scores (CADD>15)[42], minor allele frequency (MAF) <0.005 in the Genome Aggregation Database (v3.1.2) [43], Kaviar Genomic Variant Database[44] and the internal UCLex allele frequency[39] to exclude artifactual calls generated by the genomics pipeline. Splice site predictions were subsequently made using SpliceAI[45] and SpliceRover[46]. FECD cases identified to harbor rare *TCF4* variants were additionally interrogated for rare and potentially disease-associated variants (MAF<0.005 in gnomAD and Kaviar Genomic Variant Database; CADD>15) in previously reported FECD-associated genes, including *COL8A2*, *SLC4A11*, *AGBL1* and *ZEB1*[3].

Results

RNA-seq data reveal CTG18.1 expansion-mediated FECD to be transcriptomically distinct from non-expanded FECD

All primary CEC monolayer cultures selected for downstream analysis displayed a consistent morphology characteristic of CECs *in vivo.* We generated short-read RNA-seq data from unpassaged primary CEC cultures derived from healthy control (n=4, mean age 58.5 years), Exp+ FECD (mono-allelic CTG18.1 allele repeat length ≥50) (n=3, mean age 63.3 years) and Exp- FECD (bi-allelic CTG18.1 allele repeat lengths ≤30) (n=3, mean age 63.3 years) tissues **(Table S1).** Exome data generated from the Exp- FECD cases were analyzed for the presence of variants in any previously FECD-associated genes, such as *TCF4, COL8A2, SLC4A11, ZEB1,* and *AGBL1*[3], to identify any potential underlying genetic cause. No rare (MAF <0.005) or potentially deleterious (CADD >15) variants were identified.

Hierarchical clustering using heatmap visualization and PCA analysis revealed that Exp+ samples grouped separately from not only the controls, but also the Exp- samples **(Figure 1A, B)**. Notably, Exp- samples also clustered together on the PCA plot, but the heatmap derived from hierarchical clustering highlighted that greater intra-group variability existed between Exp- samples compared to variability within control and Exp+ groups (**Figure 1A**). We hypothesize that the increased transcriptomic diversity within the Exp- group is due to potentially distinct, as yet unidentified, genetic causes driving disease for each FECD case-derived sample. We included all three groups in the downstream comparative pipelines using the following pairwise comparisons; **(PWC1)** control versus Exp+ FECD, **(PWC2)** Exp+ FECD versus Exp- FECD, and **(PWC3)** control versus Exp- FECD.

To further verify splicing dysregulation events identified in the short-read transcriptomic dataset and better understand the relationship between individual events to transcript-specific isoforms, we also generated long-read RNA-seq data from additional, biologically independent unpassaged primary CEC cultures derived from healthy control (n=4. mean age 65.25 years) and Exp+ FECD (mono-allelic CTG18.1 allele repeat length ≥50) (n=4, mean age 66.75 years) tissues **(Table S1).** After alignment and stringent quality control, reads that displayed evidence of 5' degradation, intrapriming and reverse transcription template switching were removed. Exp+ and control samples were subsequently analyzed with SQANTI3 to assess isoform diversity (**Table S2**).

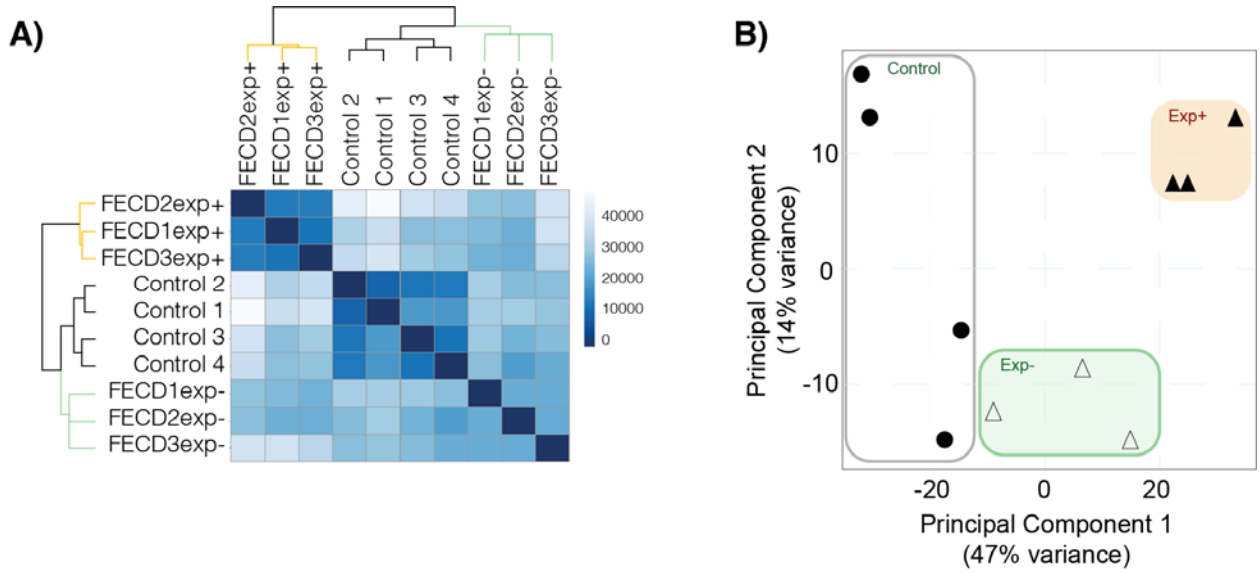


Figure 1: Similarities and differences of short-read transcriptomes of diseased and control CECs show three distinct subgroups. A) Hierarchical clustering heatmap of sample-to-sample distances within short-read RNA-Seq data. Darker colors denote higher similarities between samples, demonstrating that Exp+ and control are more similar within each subgroup than with other samples. Exp- shows a similar result but is less pronounced, likely due to differences in unidentified underlying genetic causes. **B)** Principal component plot of all short-read samples. All three subsets of data also cluster separately in the PCA plot.

Alternative splicing analysis of long-read and short-read transcriptomic data demonstrate the complexity of aberrant splicing events underlying CTG18.1 expansion-mediated disease

To characterize and quantify alternative splicing events within our short-read dataset, we utilized rMATS, a junction-based method that employs a hierarchical framework, to model unpaired datasets and evaluate and quantify all major alternative splice types[33,47,48]. Comparing Exp+ to control samples (PWC1) we identified 2,118 significant alternative splicing events in 1,437 genes (**Figure 2A, Figure S1, Table S2)**. Alternative splicing comparisons including Exp+ (PWC1 and PWC2) yielded the most alternative splice types in every category (**Figure 2A**) supporting the hypothesis that this genetic subtype of disease is a spliceopathy.

Iso-seq and subsequent SQANTI3 analysis (**Table S3**) revealed a higher proportion of isoforms per gene in Exp+ samples (1.9 unique isoforms per unique gene in Exp+ vs 1.5 in control). Additionally, there was a larger proportion of novel genes identified (5.35% vs 4.7% of total genes identified). When characterizing transcripts based on splice junctions, we saw a higher proportion of isoforms per gene classified as novel in catalog (NIC) and novel not in catalog (NNC) in Exp+ vs control (1.2 ctr vs 1.4 Exp+ for NIC, and 1.3 ctr vs 1.5 Exp+ for NNC) (**Table S3**). This is consistent with the rMATS results (**Figure 2**), where Exp+ samples showed more isoform diversity compared to control and Exp- samples.

We then compared the rMATS and Iso-seq results to a subset of published datasets exploring differential splicing in FECD patient-derived samples (**Table S4**) and selected 24

published differential splicing events (from 23 genes) that had previously been identified by bioinformatic or *in vitro* methods to be strongly associated with CTG18.1 expansion-mediated FECD[3,5,6]. Alternative splicing analysis of our transcriptomic dataset identified that 75.0% (18/24) of these events were significant in PWC1 (control vs Exp+, **Table S4**). In addition to these 24 selected events, rMATS identified another 10 novel and significant differential splicing events within Exp+ in these genes, highlighting the increasing complexity of splicing dysregulation within Exp+ FECD. Long-read sequencing confirmed the presence of these novel splicing events in Exp+ FECD (**Table S5**). Also, it was interesting to note that *FN1* was identified to have robust alternative splicing in Exp+, and further investigation showed that alternatively spliced exons in Exp+ corresponded with skipping of the Extra Domain A (EDA) and Extra Domain B (EDB) protein domains that play a critical role in signal transduction and fibrillar formation[49] (**Table S9**).

Skipped exon alternative splicing events were the most common in all three pairwise comparisons (**Figure 2A**). Prior work regarding alternative splicing in FECD has focused on MBNL1/2 and their sequestration by the repeat transcripts, leading to the hypothesis that the localized depletion of available MBNL splicing factors would cause missplicing[5,6,10]. MBNL proteins are key regulators of splicing, acting alone to regulate exon splicing (positively and negatively, dependent on binding site), while also acting synergistically with other splicing factors[50,51]. Given the complex role of MBNL proteins in transcript splicing, when we explore skipped exon events in PWC1, it is evident that Exp+ skipped exon events are evenly distributed in magnitude and direction (**Figure 2B)**.

We further explored this observation to determine if the flanking intronic sequences were enriched for common RNA-binding protein (RBP) motifs and binding sites. We focussed our RBP motif enrichment analysis (rMAPS2[34]) on the 250bp region upstream of alternatively spliced exons and isolated those areas where there was enrichment in PWC1 and PWC2, but not PWC3 (i.e. events enriched within Exp+ CEC lines). By filtering rMAPS results in this manner, a trend whereby skipped exons are significantly ($p < 0.05$) upregulated by RNA-binding proteins with TT-containing binding motifs and downregulated by RNA-binding proteins with CG-containing binding motifs was observed (**Table S6**). As MBNL proteins favor YCGY binding motifs, this suggests that other RNA-binding proteins with TT-containing motifs could also play a role in driving the spliceopathy underpinning a key mechanism of disease in Exp+ FECD.

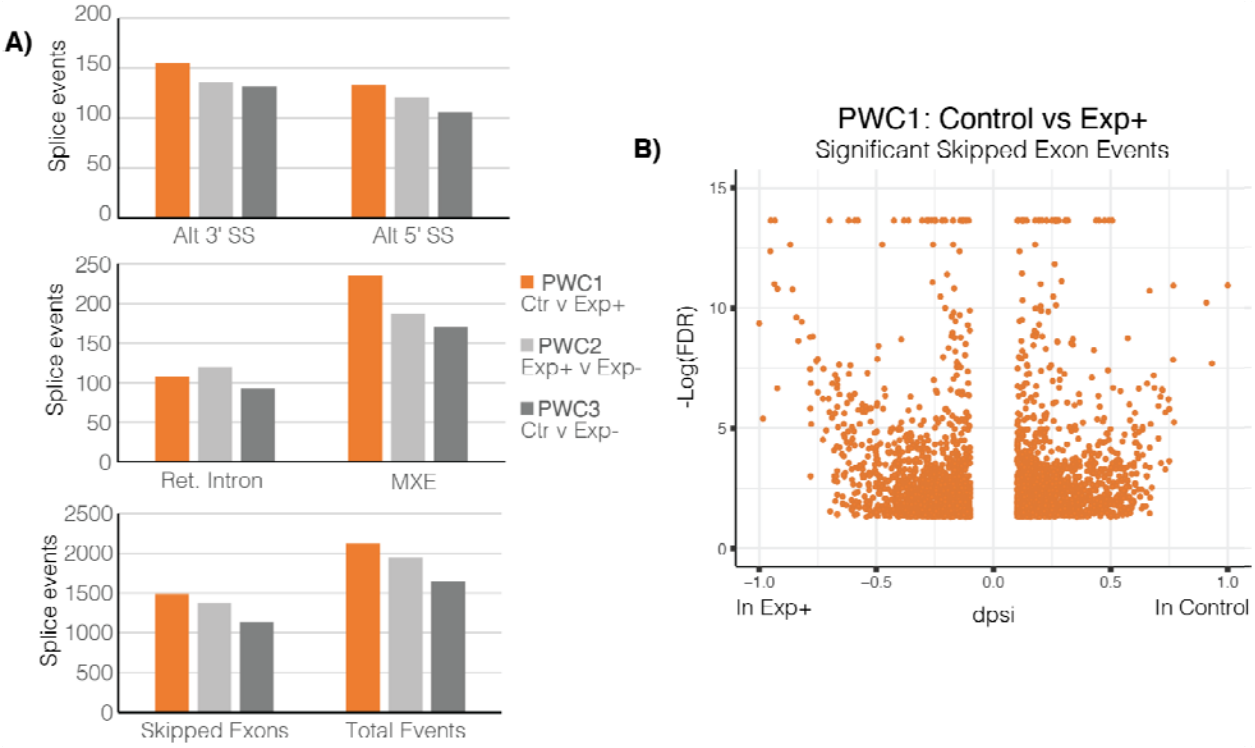


Figure 2: Alternative splicing analyses of short-read CEC RNA-seq data demonstrate increased levels of alternative splicing in Exp+ FECD. A) Summary of global differences in splicing events categorized by rMATS**.** N=4 for controls, N=3 each for FECD Exp+ and FECD Exp-. The most common splice category of alternative splicing observed between all pairwise comparisons was skipped exon, representing ~70% of all significant alternative splicing events detected. Alt 3' SS: alternative 3' splice site, Alt 5' SS: alternative 5' splice site, Ret. Intron: retained intron, MXE: mutually exclusive exon **B)** Volcano plot of statistically significant skipped exon events in PWC1 (Control vs Exp+). The dpsi value denotes the magnitude of change for each dysregulated skipped exon event identified. A positive dspi denotes decreased levels of exon inclusion in Exp+, whereas a negative dspi denotes increased levels of exon inclusion in Exp+.

Evidence of TCF4 dysregulation appears enriched in CTG18.1-mediated FECD, highlighting specific candidate biomarkers for this genetic subtype of disease

Differential gene expression was assessed with DESeq2[28] in all three groups. For all comparisons an FDR corrected P-value (padj) of <0.05 was used as a threshold for significance (**Table S7**). A total of 4,399 genes (16.7% of all mapped genes) were identified to be differentially expressed between the Exp+ and controls (PWC1; **Figure 3A**). Of these, 1,905 genes had a shrunkLFC value (accurate $\log_2$fold values of expression accounting for high dispersion and low expression counts) of greater than 1 or less than -1. Overall, PWC2 and PWC3 groups resulted in 2,273 and 2,118 significantly differentially expressed genes, respectively **(Figure 3A; Table 1)**.

Pairwise analysis revealed considerable overlap between dysregulated genes identified by all three pairwise comparisons **(Figure 3A),** however, highest convergence was observed by PWC1 (Exp+ versus control) and PWC2 (Exp+ v Exp-), suggesting these events are driven by CTG18.1 expansions. In total, both pairwise comparisons reveal 1,150 dysregulated genes uniquely and significantly dysregulated in Exp+ FECD case-derived samples. Conversely, by comparing the overlap between PWC1 and PWC3 (Exp+ vs control and Exp- vs control) we identified a further 1,105 genes that are dysregulated in FECD, irrespective of the underlying genetic cause, and thus representing more ubiquitous signatures of the disease. Dysregulated genes in PWC1, including Exp+ unique and common to FECD, did not skew directionally and were evenly up or down-regulated within our analysis (**Figure 2B**). We isolated 3,183 genes that are uniquely dysregulated in Exp+ FECD, 2,033 genes dysregulated solely in PWC1 and a further 1,150 dysregulated genes shared with PWC2; (highlighted in orange **Figure 3A**) and proceeded with downstream analysis (**Table S7**).

To determine if other RNA-binding proteins may be implicated in the alternative splicing seen in Exp+ FECD, we cross-referenced our short-read data with a list of ~1500 human RNA-binding proteins[52], and discovered that there were 207 RBPs uniquely dysregulated in Exp+ (59 downregulated and 148 upregulated), including *MBNL1, CELF1, ESRP2,* and a variety of hnRNP and DDX encoding genes (**Table S8**). Both the hnRNP and DDX family of RNA-binding proteins have previously been implicated in other repeat expansion diseases[53–56]. It is likely that a combination of these dysregulated RBPs plays a role in driving the alternative splicing seen in Exp+ FECD.

We conducted pathway analysis using g:profiler[54] (querying gene ontology, KEGG pathway, and Reactome pathways). Gene ontology and KEGG pathway analysis of PWC1 dysregulated genes was consistent with previously published transcriptomes[5,57,58]. Intriguingly, we see strong parallels to pathway enrichment analyses highlighted from a study of *TCF4* neural progenitor knockdown cell model[59] within the 3,183 genes uniquely dysregulated in Exp+, such as protein processing in endoplasmic reticulum, MAPK signaling, actin cytoskeletal regulation, endocytosis and apoptosis (**Table S10**). Critically, these pathways are not enriched in PWC3, suggesting that these are unique to Exp+ CECs. Furthermore, several enriched pathways involving Wnt signaling were seen in the 3,183 dysregulated genes unique to Exp+ that are not present in PWC3, again suggesting *TCF4* function may be modified within Exp+ FECD CECs (**Table S11**). Furthermore, enrichment of dysregulated genes in pathways involving other repeat expansion-mediated diseases and misfolded protein aggregate diseases suggest mechanistic parallels which reinforce Exp+ CECs as a model system with broad relevance to repeatome biology (**Table S11**).

The consistent presence of PI3K-Akt pathway enrichment within GO, KEGG, and Reactome results in Exp+ dysregulated genes (and its consistent absence in PWC3 pathway enrichment) is a novel observation and suggests that Exp+ FECD CECs may interact with growth factors and the extracellular matrix in a unique manner. Hence, components of the PI3K-Akt pathway may serve as informative biomarkers for Exp+ FECD in future clinical trial settings, and thus warrant further investigation. Additional pathways are enriched in this set of genes (and not in PWC3) that involve RNA processing and binding, other repeat-mediated disease related mechanisms, epithelial to mesenchymal transition, and multiple signal transduction cascades with extracellular matrix interactions (**Table S11**).

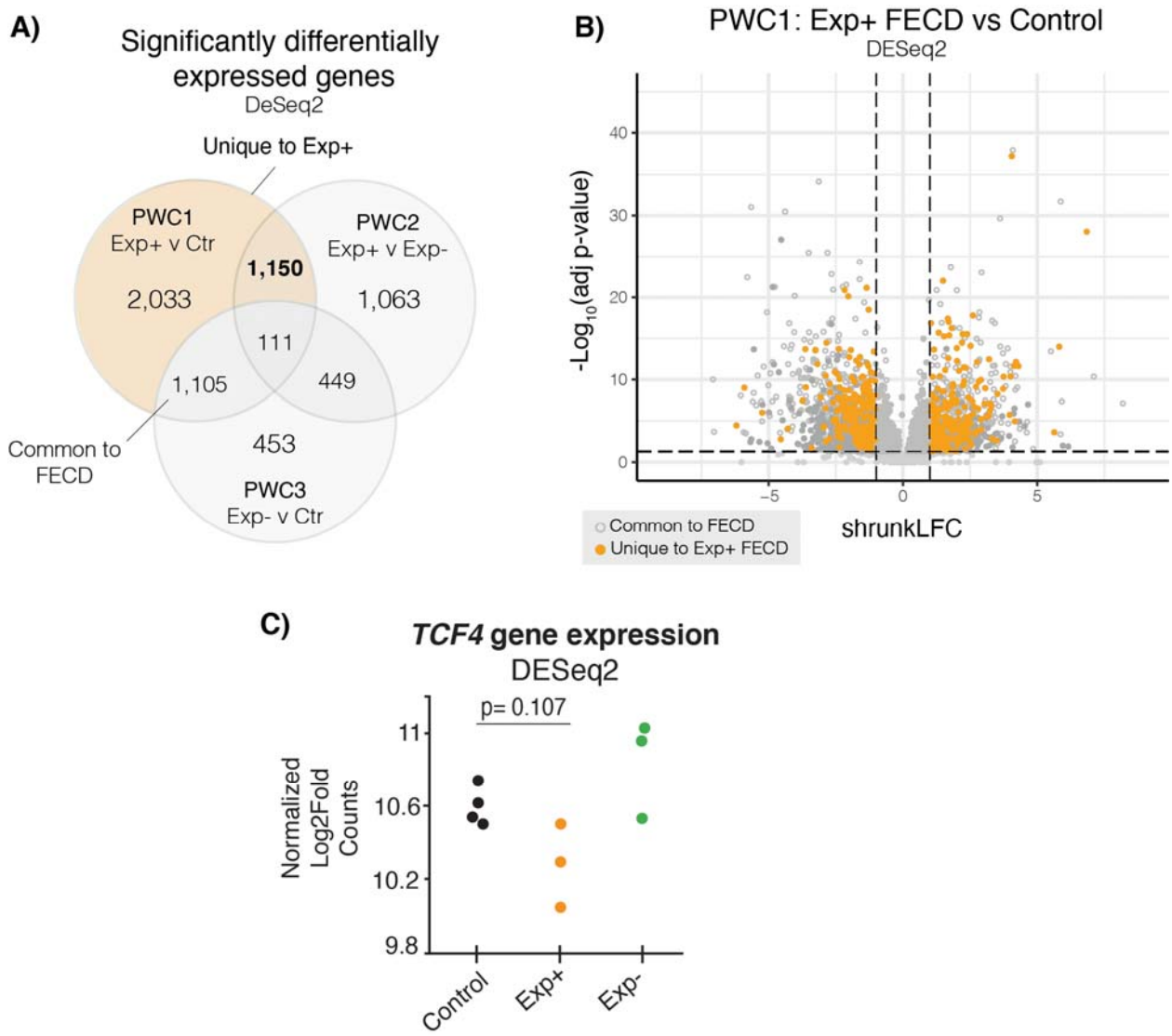


Figure 3: Differential gene expression analysis on Exp+ FECD, Exp- FECD, and control primary CEC transcriptomes. **A)** Differential gene expression was assessed via DESeq2. Three pairwise comparisons were conducted: PWC1 (Exp+ vs control), PWC2 (Exp+ vs Exp-), and PWC3 (Exp- vs control). Venn diagrams show overlap of significantly differentially expressed genes (adjusted p-value < 0.05) within these three comparisons. **B)** Volcano plot shows the distribution of gene expression in Exp+ FECD compared to the control group. Horizontal dotted lines denote the boundary between significance (adjusted p-value < 0.05), whereas vertical dotted lines show a shrunkLFC cutoffs for -1 and 1. Dark gray solid markers show significantly expressed genes with a fold change magnitude greater than 1. Solid orange markers delineate uniquely Exp+ differentially expressed genes. **C)** Total *TCF4* expression showed no significant dysregulation between the different groups.

Differential TCF4 exon usage suggests isoform ratio disruption of TCF4 in CTG18.1-mediated FECD

Given signatures of *TCF4* dysregulation were apparent in the pathway enrichment analysis of Exp+ CECs (**Table S10**), we then aimed to determine if any tangible patterns of *TCF4* transcript-specific dysregulation was observed in the Exp+ CECs, despite total *TCF4* levels not being significantly different between Exp+, Exp- and control CECs **(Figure 3C).** On this basis, DEXSeq2[32] was used to assess differential exon usage.

DEXSeq2 utilizes generalized linear models to detect differential exon usage while controlling for biological variation. This analysis allowed us to query gene expression and splicing at the exon level. In concordance with our rMATs and DESeq2 analysis, higher levels of transcriptome-wide differential exon usage events were observed in PWC1 and PWC2 compared to PWC3 (**Figure S2**), suggesting these patterns of differential exon usage are likely attributed to CTG18.1-related pathology.

Multiple *TCF4* exons (n=15) were significantly dysregulated in Exp+ compared to control and Exp- FECD samples (PWC1 results shown in **Figure 4A, Table S12**), whereas no significant differences were observed between Exp- and control (PWC3, **Figure 4B, Table S12**), demonstrating that differential *TCF4* exon usage is driven by CTG18.1 expansions. A subset of exons contained within a small group of *TCF4* transcripts (8 out of 93) are upregulated in Exp+, while an increased number of more broadly expressed exons are downregulated (appearing in 64 out of 93 transcripts) (**Figure 4A, Table S12**). Transcripts containing downregulated exons are primarily included in longer isoforms of the gene which contain 3 AD domains, while the upregulated exons (excluding those surrounding the repeat) appear in shorter *TCF4* isoforms containing only two AD domains. This suggests that Exp+ CECs may be employing certain compensatory measures in response to the transcript-specific dysregulation that is induced by the presence of CTG18.1 repeat expansions. Interestingly, our rMATS analysis independently detects this same phenomenon (**Table S13**), which showed a significant mutually exclusive exon splicing event whereby the last exon in common with most 3 AD domain-containing isoforms (Exon 6 in ENST000000354452.8) was favored in the control group, while the first exon in common with all isoforms (Exon 7 in ENST000000354452.8) was favored in the Exp+ group.

We also detect significant retention of the intron flanking the CTG18.1 repeat, previously described by Sznajder et al. 2018[60] in Exp+ FECD **(Figure 4A).** The intronic region is significantly upregulated in Exp+ CECs (E173 and E174) in both PWC1 and PWC2 excluding PWC3 **(Figure 4A, Table S10)**. While this intron represents the 5' untranslated region (UTR) of several transcript of *TCF4*, it is also included within an intronic region in 32 out of 93 currently annotated *TCF4* transcripts.

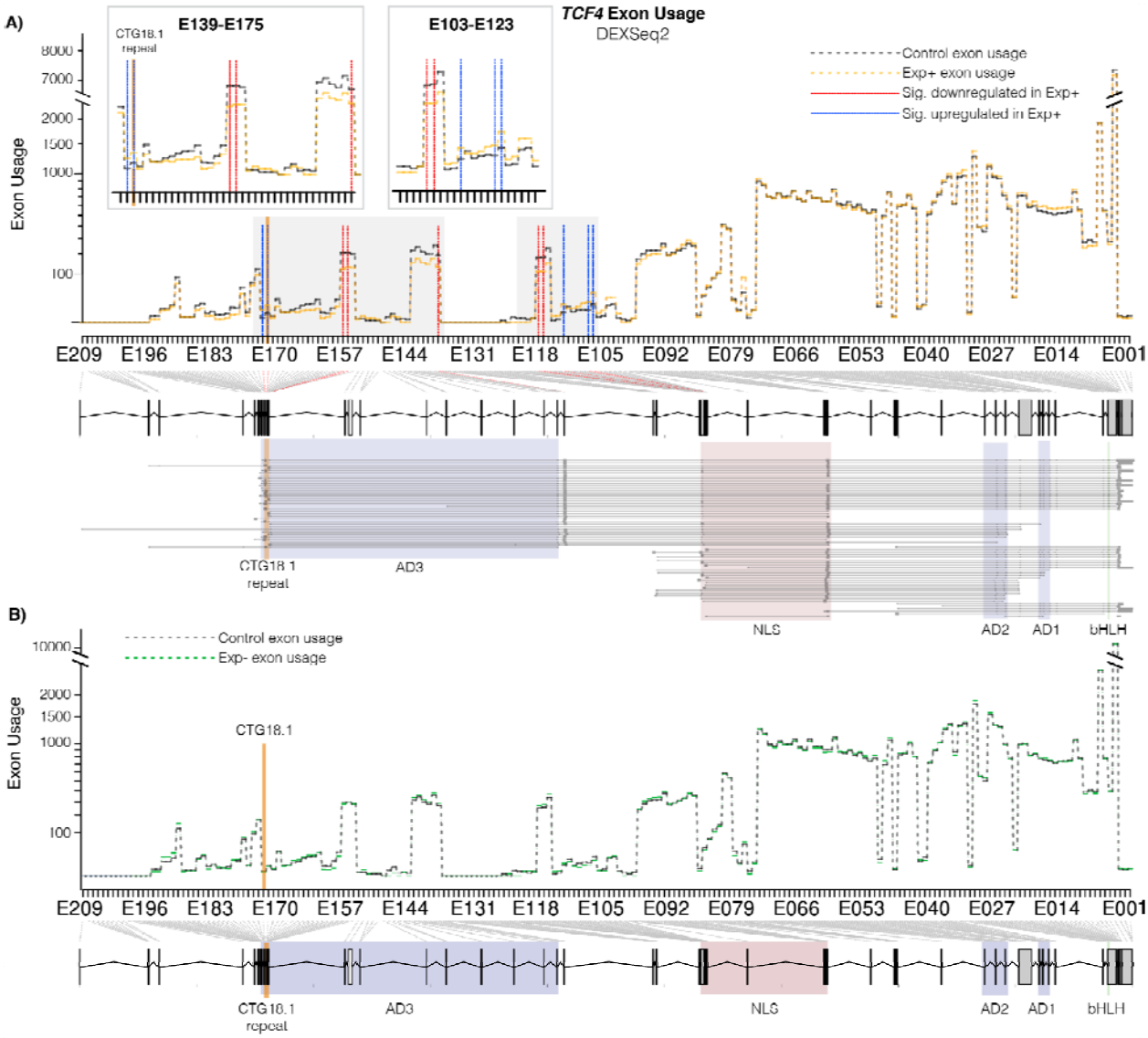


Figure 4: CTG18.1 expansion-mediated FECD results in alternative *TCF4* exon usage within corneal endothelial cells. A) Differences in exon usage between control (gray dotted line) and Exp+ (orange dotted line) CECs. Significantly upregulated and downregulated exons are highlighted with blue and red vertical lines, respectively. The location of CTG18.1 is depicted with an orange line. Inset boxes magnify dysregulated regions. A genomic schematic of *TCF4* with transcripts aligned is presented below the axis. Key protein domains highlighted (green - basic helix loop helix (bHLH) domain, blue - activation domains (ADs), red - main nuclear localization signal (NLS). **B)** Differences in exon usage between control (gray dotted line) and Exp- (green dotted line) CECs. No exons were statistically dysregulated.

RNAScope shows robust detection of altered TCF4 isoform expression in CTG18.1-mediated FECD case-derived corneal endothelial cells

To further validate our hypothesis that the subset of *TCF4* isoforms containing AD3 was being selectively downregulated due to the presence of an expanded CTG18.1 (≥50 repeats), we performed RNAScope. We visualized *TCF4* transcripts in Exp+ FECD, Exp- FECD, and control CECs using two distinct sets of probes, termed A and B (**Figure 5A**). Probe set A comprises a set of probes that allow us to visualize all *TCF4* transcripts in the cell (probe covers all exons in the transcript Isoform B+, TCF4-201, ENST00000354452.8, NM_001083962.2). Probe set B enables exclusive detection of the longer subset of transcripts as this probe set targets exons in the canonical *TCF4* transcript containing all three AD domains which do not overlap with the canonical *TCF4* transcript containing AD1 and AD2 only (Isoform A+/J, TCF4-204, ENST00000457482.7, NM_001243234.2). Hence, these probe sets enable us to detect and visualize the ratio of the longer AD3-containing cohort of transcripts, identified to be downregulated in Exp+ CECs, via our differential exon usage analysis of RNA-seq data (**Figure 4A**).

The specified subsets of *TCF4* transcripts were visualized in control, Exp+ and Exp- FECD case-derived CECs (n=3 for each) (**Figure 5B, Table S14)**. Each independent primary CEC culture was also probed with a negative control (targeting bacterial genes) and a $(CAG)_7$-Cy3 probe, using a standard FISH protocol, to validate the presence or absence of CUG-specific RNA foci derived from expanded copies of CTG18.1 (**Figure S14**). RNA foci were consistently detected in all CTG18.1 Exp+ CECs, and absent from the control and Exp- CECs. Using CellProfiler (version 4) puncta representing *TCF4* transcripts, detected with both probe set A and B, were quantified and puncta/cell values were established for each image. Puncta/cell values were averaged across images and a percentage was calculated representing the proportion of total *TCF4* puncta representing the subset of transcripts with AD3 for each biological replicate. This showed that, on average, Exp+ CECs had a significantly smaller proportion of the longer AD3-containing transcripts (16.33% ± 4.53) compared to unaffected control CECs (36% ± 6.9, p=0.0145) and non-expanded FECD CECs (30.9% ± 5.17, p=0.0213). This result suggests that the presence of the expanded repeat alone consistently skews *TCF4* transcript ratios to favor shorter isoforms. As this subgroup of isoforms typically contains only two out of three activation domains, a higher prevalence of shorter TCF4 isoforms would likely affect potential dimerization and TCF4 transcriptional function in Exp+ CECs.

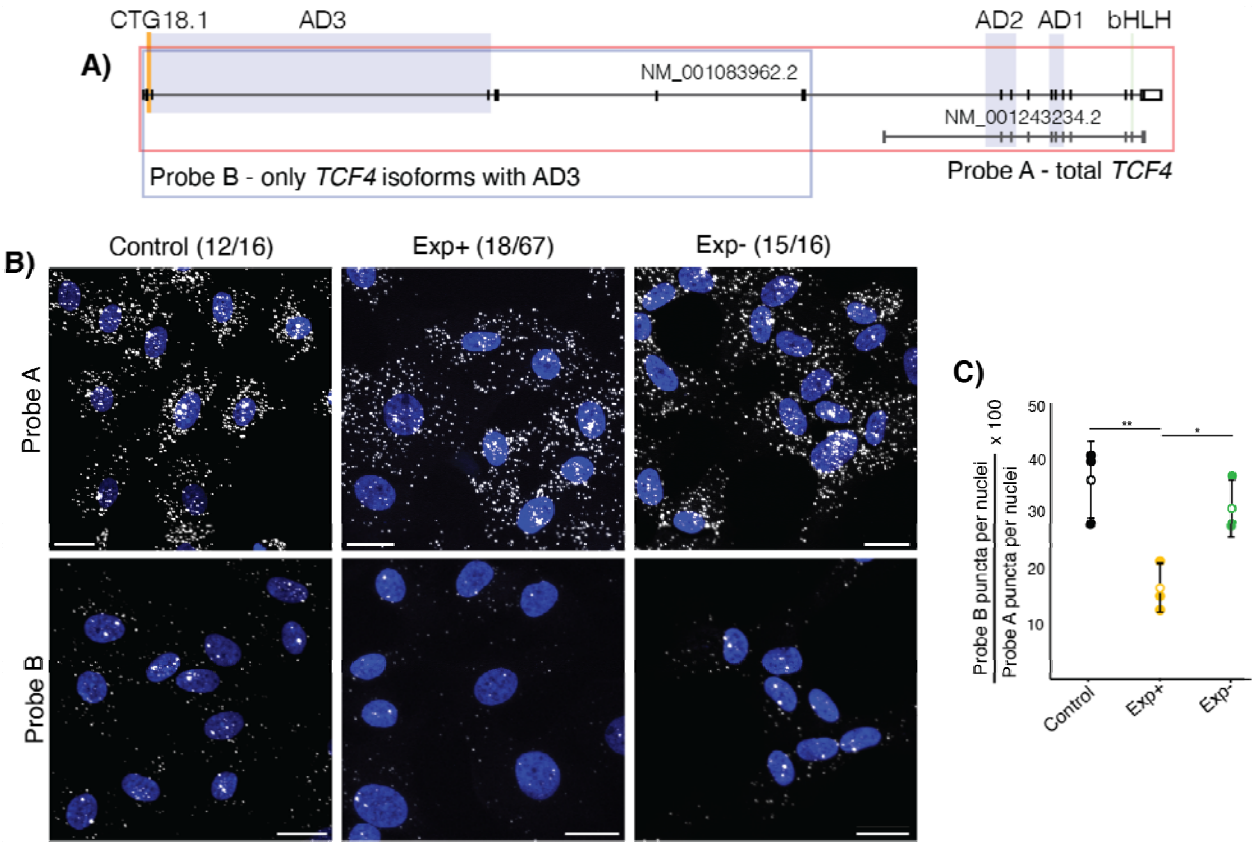


Figure 5: RNAScope probes targeting *TCF4* transcripts demonstrate that longer transcripts containing AD3 account for a smaller proportion of total transcripts in Exp+ FECD CECs. A) Probe targeting schematic for *TCF4*. Probe A targets all exons in NM_001083962.2 and will visualize total *TCF4* transcripts, while Probe B will target only the 5' exons, excluding those that are contained within shorter transcripts (NM_001243234.2). Probe B will allow us to visualize the longer subset of AD3-containing transcripts within the cell. CTG18.1 is highlighted in orange. Green bar denotes TCF4 bHLH region, blue regions highlight TCF4 activation domains, red box shows the bipartite TCF4 NLS signal and orange shows the genomic region containing the CTG18.1 repeat. **B)** Representative confocal images of Exp+, Exp- and control CECs with nuclei (blue, DAPI) and *TCF4* puncta (visualized with Cy3 but represented here with white). Scale bars represent 20 µm. **C)** Plotting proportion of longer AD3-containing *TCF4* transcripts in Exp+ (orange), control (black), and Exp- (green) shows that Exp+ have significantly lowered proportion of longer transcripts compared to both control ($p = 0.0145$) and Exp- ($p = 0.0213$) CECs, while there is no significant difference in proportion between control and Exp- ($p = 0.362$). Open circle denotes mean and error bars show standard deviation. Significance determined with student T-test.

Rare TCF4 variants were identified within a genetically refined CTG18.1 expansion-negative FECD cohort, illuminating the potential importance of isoform-specific TCF4 dysregulation in FECD

We next wanted to explore if rare and potentially deleterious *TCF4* variants could be contributing to disease in FECD cases without a CTG18.1 expansion, given that Exp+ CECs display alternative *TCF4* exon usage (**Figure 4; Table S12**), leading to skewed ratios of *TCF4* isoforms (**Figure 5**), with detectable downstream features of TCF4-driven dysregulation (**Table S10**). To test this hypothesis we interrogated exome data generated from 141 Exp- FECD cases, acquired from our ongoing inherited corneal disease research program, representing approximately 22% of total FECD cases recruited to date. Filtering for rare (MAF <0.005) and potentially deleterious (CADD score threshold of >15) *TCF4* variants, we identified four distinct variants in 3 Exp- FECD cases (probands A, B and C; **Table 1, S16**). Sanger sequencing confirmed the presence of all four *TCF4* variants and exome data from these cases were also interrogated using the same thresholds (MAF <0.005, CADD >15) to determine if any other rare variants in genes previously associated with FECD could also be contributing to disease[3] (**Table S15**).

Proband A, a female first diagnosed with FECD at 57-years-of-age (**Table S16; Figure S3**), harbored two unique, and consecutive, in *cis TCF4* variants encompassed by six *TCF4* transcripts **(Table 1; Figure 6).** For ENST00000566286.5, the variants occur within a coding region and result in consecutive *in cis* heterozygous missense and nonsense variants; c.[57G>T; 58A>T]; p.[(Arg19Ser; p.Lys20*)]. For all other transcripts, the variants are located within 5' untranslated (UTR) regions and could therefore also potentially exert regulatory effects on their respective expression within the corneal endothelium[61–63] (**Figure 6B**). Proband A was not identified to harbor any additional rare, and predicted deleterious, variants in any additional FECD-associated genes.

Proband B, a male first diagnosed with FECD at the age of 42-years-old (**Table S16**), harbored a single nucleotide variant, positioned 7-8bp downstream of the variants identified in Proband A **(Table 1; Figure 6).** Similarly, this change occurs within a coding region of a single protein-coding transcript expressed within CECs, ENST00000566286.5. Here it results in a synonymous change, c.66G>A, p.(Glu22=). However, given that it alters the last nucleotide of exon 1 within the transcript, we performed in silico analysis to predict if it is likely to alter splicing[45,46]. Splice AI predicts the variant introduces loss of the splice donor site for exon 1 (SpliceAI Δ score 0.78). SpliceRover also predicts that c.66G>A weakens the native splice donor site for exon1 ENST00000566286.5 (from 0.320 to 0.004) and that this could result in the activation of a cryptic splice donor site downstream (from 0.098 to 0.233) of the wildtype donor site, which would introduce a short frameshift insertion followed by a premature termination codon (PTC) c.66_67insGTGCTCGATGAATTTTC, p.(Arg23Valfs*12). For an additional five transcripts, the variant is located within non-coding regions, and due to its exon boundary location for three of these (**Figure 6C**) it is also predicted to result in aberrant splicing and the insertion of a 17bp insertion within the 5' UTR region (**Figure 6C**). Hence, it may also plausibly affect the regulation of these transcripts within the corneal endothelium. Proband B was also

identified to carry a single heterozygous *SLC4A11* missense variant of uncertain clinical significance, not previously associated with FECD; c.1040G>A, p.(Arg347Gln) (**Table S15**). Unfortunately, tissue was not available from the affected proband to test the functional relevance of either the *TCF4* or *SLC4A11* variants *in vitro*.

It is intriguing that *TCF4* variants identified in both probands A and B are predicted to affect only a small fraction of total *TCF4* isoforms encompassed by the region of interest (6/71), either via potential haploinsufficiency of a single transcript (ENST00000566286.5) and/or potential disruption of 5'UTR regulatory elements for a further five (**Figure 6**). This observation draws parallels to our transcriptomic and RNAScope data in Exp+ CECs (**Figures 4 and 5**) as it similarly suggests that only a fraction of *TCF4* isoforms are dysregulated in CECs derived from CTG18.1-mediated FECD cases.

A further rare missense variant was identified in Proband C, c.26T>C, p.(Ile9Thr) (**Table 1, Table S15**), first referred to our center at the age of 74-years, with a strong family history of FECD. This exon is only included in four annotated *TCF4* transcripts that do not show expression in our CEC system (**Figure 6A**). No other rare variants in FECD-associated genes were identified in this individual.

Table 1: Summary of rare, potentially deleterious *TCF4* variants identified in a FECD CTG18.1 Exp- cohort (n=141).

Variant (hg38)	HGVSc, HGVSp	CADD, MAF gnomAD MAF Kaviar	FECD Exp-Cases (n=141)
chr18:55,588,479-C-A	c.57G>T p.(Arg19Ser)	18.3 0 0	1/141 (Proband A)
chr18:55,588,478-T-A	c.58A>T p.(Lys20*)	16.45 0 0	1/141 (Proband A)
chr18:55,588,470-C-T	c.66G>A p.(Glu22=)	18.39 2.09×10^{-3} 0	1/141 (Proband B)
chr18:55,403,689-A-G	c.26T>C p.(Ile9Thr)	17.29 0 6.40×10^{-6}	1/141 (Proband C)

Rare (MAF < 0.005 in gnomAD genomes, Kaviar and internal UCLex allele frequency) and potentially deleterious (CADD score threshold of >15) *TCF4* variants identified within a FECD CTG18.1 Exp- cohort (n=141). The annotation from Variant Effect Predictor is filtered to the "most severe" transcripts which are ENST00000544241.6 for Proband C, and ENST00000566286.5 for Proband A and B. Abbreviations are as follows: Kaviar, Kaviar Genomic Variant database; gnomAD, The Genome Aggregation Database; UTR, untranslated region; MAF, minor allele frequency, HGVSc, coding DNA sequence based on the Human Genome Variation Society; HGVSp**,** protein sequence based on the Human Genome Variation Society.

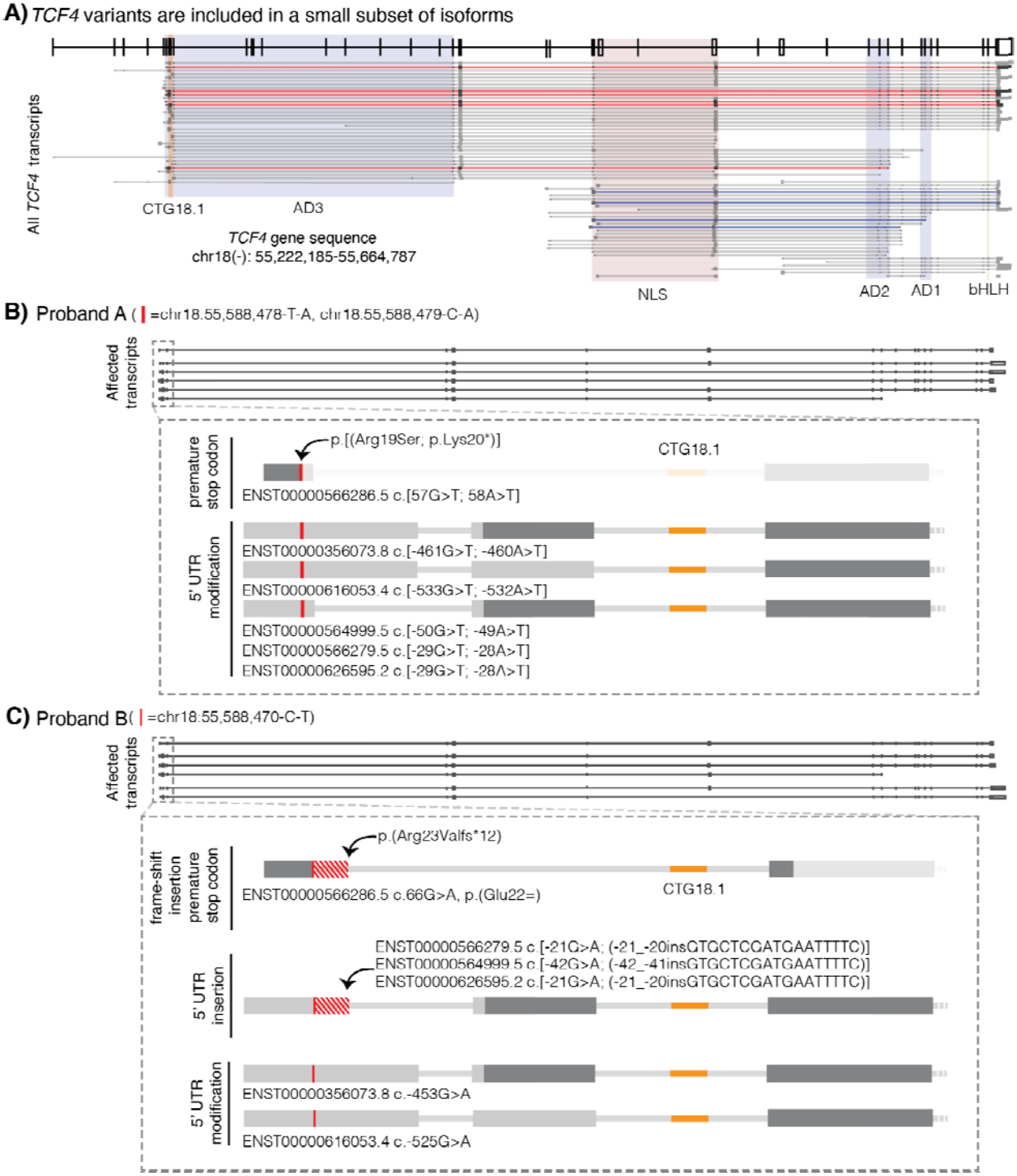


Figure 6: Schematic of *TCF4* and rare variants identified in molecularly unsolved FECD cases, without CTG18.1 expansions. **A)** All 93 annotated *TCF4* transcripts depicted, and only 10 include exons containing rare variants identified in *TCF4* expansion negative FECD cases. The six transcripts that encompass variants present within Proband A and B are highlighted in red (ENST00000616053.4, ENST00000356073.8, ENST00000564999.5, ENST00000566279.5, ENST00000566286.5, ENST00000626595.2). The four isoforms that encompass variants identified in Proband C are highlighted in blue (ENST00000630828.2, ENST00000628078.2,

ENST00000564228.5, ENST00000544241.6). Green bar denotes TCF4 bHLH region, blue regions highlight TCF4 activation domains, red box shows the bipartite TCF4 NLS signal and orange shows the genomic region containing the CTG18.1 repeat. **B)** Proband A harbors two heterozygous, *in cis*, rare variants affecting consecutive nucleotides included within 6 distinct transcripts. For protein coding transcript ENST00000566286.5, the variants are predicted to introduce a missense variant immediately followed by a premature stop codon after 20 amino acids. For the remaining 5 transcripts, the rare variants are located within 5' untranslated regulatory regions (UTR). **C)** Proband B harbors a single rare heterozygous variant encompassed within the same 6 transcripts as proband A. For protein coding transcript ENST00000566286.5, the variant introduces a synonymous variant altering the last nucleotide of exon 1. This is predicted to weaken the native splice donor site resulting in activation of a cryptic downstream splice donor site and subsequently a frameshift insertion, followed by a premature termination codon (PTC). The variant is located within 5'UTR regions for 5 additional transcripts. For 3/5 of these (ENST00000566279.5, ENST00000564999.5, ENST00000626595.2) the variant is similarly predicted to affect splicing and introduce a 17bp insertion into the 5'UTR region.

Discussion:

FECD is now recognized to be the most common repeat expansion-mediated disease in humans, with approximately 75% of European FECD cases harboring CTG18.1 expansions[3]. However, many aspects of its underlying pathophysiology remain elusive, despite the strong mechanistic parallels existing with other repeat-mediated diseases, and the urgent clinical need for new treatments for this common and visually disabling disease. Here, we have utilized a large, biologically independent set of primary FECD case-derived CEC lines to enhance our understanding of the multiple pathogenic processes underlying CTG18.1-mediated FECD. Our extensive bulk RNA-Seq and spatial transcriptomic analysis of FECD case-derived CEC lines demonstrates that, in addition to being a spliceopathy, expansion-positive FECD is concurrently associated with isoform-specific patterns of *TCF4* dysregulation, hence providing novel mechanistic insights into the pathophysiology of this triplet-repeat mediated disease. Furthermore, we have discovered rare and predicted deleterious *TCF4* variants in FECD cases without CTG18.1 expansions. This suggests that, in atypical instances, disruption of a specific subset of *TCF4* transcripts may underlie FECD, independent of CTG18.1 expansions. This, in part, could explain the missing heritability of the disease and reinforces the hypothesis that dysregulation of TCF4 itself is pivotal to FECD pathophysiology.

Our robust transcriptomic analysis, presented within a biologically relevant cellular and genomic context, highlights an array of novel molecular signatures driven by CTG18.1 expansions, as well as validating a range of previously described Exp+-associated transcriptomic events. We anticipate that many of the novel events identified here could serve as informative biomarkers to aid the development of CTG18.1 targeted therapeutic approaches. To date, CUG-containing RNA foci and MBNL1/2 missplicing have been adopted as the primary therapeutic outcome measures[6,10,64,65]. Data presented here support and expand the hypothesis that accumulation of repeat expansion-derived RNA structures induce toxic gain-of-function effects that disrupt cellular homeostasis and induce aberrant RNA metabolism[6,10]. For instance, we have discovered a novel array of RBP encoding transcripts, including *MBNL3*, *CELF1* and a range of *hnRNP* and *DDX* family members, uniquely dysregulated in Exp+ CECs that warrant future follow up to elucidate their respective contributions to aberrant RNA metabolism related mechanisms. However, such biomarkers only relate to the spliceopathy aspects of the disease and our data go on to also illustrate that translational efforts would also benefit from using more diverse outcome measures. In addition to RNA metabolism biomarkers, global Exp+-mediated differential gene expression profiles, alongside *TCF4*-specific dysregulation itself, should also be included as relevant outcome measures for Exp+ therapies being developed, given we hypothesize these processes are also pivotal to the pathology. Novel cell surface biomarkers reported here that are specifically upregulated in Exp+ (e.g. *ITGA5* and *ITGB1*) could have particular utility in a clinical trial setting due to their accessible locations. Other membrane encoding proteins interfacing with the extracellular matrix or signal transduction could also be a rich source of potential external markers, such as cell surface receptors involved in the P13K-Akt pathway (e.g. *EGFR, MET,* and *LTK*) that we show are specifically upregulated in Exp+ CECs.

Genetic stratification of patient-derived CECs in this study has also highlighted that Exp+ FECD is transcriptionally distinct from Exp- disease. The data show that 72% of all differentially expressed genes, 81% of skipped exon events, and 72% of all enriched pathways detected in Exp+ are unique to the repeat expansion disease. For example, in accordance with previous studies[57,66], we observe that pronounced upregulation of fibronectin (*FN1*) is a robust transcriptomic feature of FECD, irrespective of genetic underlying cause. However, our comprehensive transcriptomic analysis has enabled us to determine that the EDA/B+ forms of fibronectin[49] are significantly favored in Exp+ FECD. This illustrates that *FN1* expression and splicing can act as both a general FECD molecular biomarker and an Exp+ specific biomarker. Similarly, *AQP1* previously documented to be downregulated in FECD[67], has here been demonstrated to only display marked downregulation in Exp+ CECs, illustrating the importance of stratifying cases based on CTG18.1 status. These examples, alongside the emerging evidence that the Exp+ FECD displays distinctive pathological features and epidemiological characteristics[4,6], warrants Exp+ FECD being categorized as a distinct clinical entity . Integrating CTG18.1 genotyping into routine FECD clinical care pathways is warranted, especially given the numerous international efforts underway to develop gene-directed therapies for CTG18.1-mediated FECD[10,11,64,65].

Our complementary RNA-seq approaches, in combination with the spatial resolution of *TCF4* transcripts subgroups, has highlighted a unique and distinctive pattern of TCF4-specific dysregulation that underpins Exp+ FECD. Additionally, our pathway analysis has isolated features of dysregulation that we hypothesize represent downstream consequences of this isoform shift and an important pathogenic component of the disease. Notably, we have observed commonality between signaling transduction pathways dysregulated in our Exp+ CECs (P13K-Akt and MAPK) and those dysregulated in neural progenitor cell *TCF4* knockdown transcriptomes[59]. We postulate that despite total overall levels of *TCF4* not displaying significant levels of downregulation in Exp+ FECD, the shift in isoform distributions to proportionally increase levels of shorter (1-2 AD domains) versus longer (2-3 AD domains) isoforms could underlie the overall reduction of TCF4 functionality detected via our gene ontology approach, given that longer TCF4 isoforms contain three activation domains and have been shown to have an order of magnitude more transcription activation function than the isoforms containing 1 or 2[15]. Furthermore, given that domains in the AD3-containing isoforms can also dictate intracellular location and dimerization partners of TCF4, we also hypothesize that additional TCF4-driven functional implications of this shift occur within affected CECs that warrant future investigation. Interestingly, a recent single-cell (sc)RNA-seq study has identified that varying levels of *TCF4* expression exist within discrete populations of healthy CECs[68]. Future application of scRNA-seq to FECD-case derived CEC samples has the potential to further delineate tissue-specific mechanisms of how *TCF4* isoform dysregulation driving disease within the corneal endothelium.

By interrogating exome data generated from a subset of genetically refined Exp- FECD cases, we have also gone on to discover rare and predicted deleterious *TCF4* variants in 3/141 Exp- FECD cases. Exp- FECD cases represent approximately 22% of our total FECD cohort and therefore rare *TCF4* variants can only function as very rare genetic causes or risk factors

for disease within this group. Nonetheless, this is still within the same order of magnitude attributed to *COL8A2*-associated disease (MIM#136800) within our cohort (3/141), a very rare but well-established genetic cause of FECD. We suggest future FECD genetic screening approaches should include sequencing coding and regulatory regions of *TCF4,* to further elucidate the potential role that rare *TCF4* variants may play in the absence of CTG18.1 expansions. The rare *TCF4* variants we identified in Exp- FECD cases are all located within exons that are not included by the vast majority of TCF4 isoforms, leading us to hypothesize that they would exert loss-of-function and/or regulatory effects on TCF4 functionality limited to only 10/93 annotated isoforms. This observation is in keeping with the relatively mild and tissue specific nature of FECD, which is in stark contrast to the severe neurodevelopmental disorder Pitt-Hopkins associated with total *TCF4* haploinsufficiency[69]. Hence, we propose that the variants identified in this study provide significant insight into which TCF4 isoforms, when dysregulated, may induce CEC-specific disease. These data also independently support our observation that Exp+ CECs display TCF4 isoform-specific dysregulation and this mechanism is key to the pathophysiology of CTG18.1 expansion-mediated disease. Parallels can be drawn to other repeat expansion-mediated disorders, such as fragile X syndrome (*FRM1*)[70], spinocerebellar ataxia type 7 (*ATXN7,* MIM #164500)[71] and *C9orf72*-mediated ALS[72], whereby rare loss-of-function mutations have been proposed to induce comparable phenotypic outcomes to non-coding repeat expansions. However, in these instances total haploinsufficiency is thought to underlie the respective pathogenic mechanisms, in contrast to this scenario whereby we hypothesize only a fraction of total *TCF4* isoforms are susceptible to the effect of either CTG18.1 expansions or rare *TCF4* variants within the corneal endothelium.

In conclusion, this study supports the hypothesis that at least two distinct pathogenic mechanisms, RNA toxicity and *TCF4* isoform-specific dysregulation, underpin the pathophysiology of the transcriptionally distinct CTG18.1-mediated FECD. In our established primary corneal endothelial system, we see evidence of *TCF4* isoform dysregulation at the *TCF4* transcript level and can detect the potential downstream effects of the isoform-specific pattern of TCF4 dysregulation. Our data also suggests that, in rare instances, irrespective of CTG18.1 expansions, FECD cases can harbor *TCF4* variants that may independently induce isoform-specific TCF4 dysregulation. This supports the hypothesis that *TCF4* dysregulation is a key mechanism underlying this common, age-related and visually disabling disease. We anticipate that concepts and biomarkers identified by this study will inform and guide the development of future translational interventions targeting this triplet-repeat mediated disease, and these in turn may act as exemplars for much rarer and currently untreatable and severe repeat expansion disorders that typically affect less accessible neurological/neuromuscular systems.

Declaration of Interest: The authors declare no competing interests.

Acknowledgments: We would like to thank all affected individuals for participating in this research. We would also like to acknowledge Miracles for Sight Eye bank for providing control corneal tissue. This work was funded by a UKRI Future Leader Fellowship MR/S031820/1 (A.E.D, C.Z, A.S), Moorfields Eye Charity GR000060, GR001395, GR001337 (A.E.D, N.B, N.H.T, A.S, N.P) and Sight Research UK SAC 036 (A.E.D, A.J.S), the Rosetrees Trust (A.E.D, A.J.S), and the The National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology (A.E.D, S.J.T, K.M, A.J.H, M.E.C, N.P). J.J, L.D., P.S., and P.L. were supported by GACR 20-19278S, UNCE/MED/007 and SVV 2016/260148. N.P was supported by an NIHR AI Award (AI_AWARD02488). This project utilized equipment funded by the UK Medical Research Council (MRC) Clinical Research Infrastructure Initiative (award number MR/M008924/1).

References:

1. Lorenzetti, D.W., Uotila, M.H., Parikh, N., and Kaufman, H.E. (1967). Central cornea guttata. Incidence in the general population. Am. J. Ophthalmol. *64*, 1155–1158.
2. Matthaei, M., Hribek, A., Clahsen, T., Bachmann, B., Cursiefen, C., and Jun, A.S. (2019). Fuchs Endothelial Corneal Dystrophy: Clinical, Genetic, Pathophysiologic, and Therapeutic Aspects. Annu. Rev. Vis. Sci. *5*, 151–175. 10.1146/annurev-vision-091718-014852.
3. Fautsch, M.P., Wieben, E.D., Baratz, K.H., Bhattacharyya, N., Sadan, A.N., Hafford-Tear, N.J., Tuft, S.J., and Davidson, A.E. (2021). TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. Prog. Retin. Eye Res. *81*, 100883. 10.1016/j.preteyeres.2020.100883.
4. Thaung, C., and Davidson, A.E. (2022). Fuchs endothelial corneal dystrophy: current perspectives on diagnostic pathology and genetics—Bowman Club Lecture. BMJ Open Ophthalmol. *7*, e001103. 10.1136/bmjophth-2022-001103.
5. Wieben, E.D., Baratz, K.H., Aleff, R.A., Kalari, K.R., Tang, X., Maguire, L.J., Patel, S.V., and Fautsch, M.P. (2019). 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. 10.1167/iovs.19-27689.
6. Zarouchlioti, C., Sanchez-Pintado, B., Hafford Tear, N.J., Klein, P., Liskova, P., Dulla, K., Semo, M., Vugler, A.A., Muthusamy, K., Dudakova, L., et al. (2018). Antisense Therapy for a Common Corneal Dystrophy Ameliorates TCF4 Repeat Expansion-Mediated Toxicity. Am. J. Hum. Genet. *102*, 528–539. 10.1016/j.ajhg.2018.02.010.
7. Depienne, C., and Mandel, J.-L. (2021). 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am. J. Hum. Genet. *108*, 764–785. 10.1016/j.ajhg.2021.03.011.
8. Balendra, R., and Isaacs, A.M. (2018). C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat. Rev. Neurol. *14*, 544–558. 10.1038/s41582-018-0047-2.
9. Du, J., Aleff, R.A., Soragni, E., Kalari, K., Nie, J., Tang, X., Davila, J., Kocher, J.-P., Patel, S.V., Gottesfeld, J.M., et al. (2015). RNA Toxicity and Missplicing in the Common Eye Disease Fuchs Endothelial Corneal Dystrophy*. J. Biol. Chem. *290*, 5979–5990. 10.1074/jbc.M114.621607.
10. Hu, J., Rong, Z., Gong, X., Zhou, Z., Sharma, V.K., Xing, C., Watts, J.K., Corey, D.R.,

and Mootha, V.V. (2018). Oligonucleotides targeting TCF4 triplet repeat expansion inhibit RNA foci and mis-splicing in Fuchs' dystrophy. Hum. Mol. Genet. *27*, 1015–1026. 10.1093/hmg/ddy018.

11. Angelbello, A.J., Benhamou, R.I., Rzuczek, S.G., Choudhary, S., Tang, Z., Chen, J.L., Roy, M., Wang, K.W., Yildirim, I., Jun, A.S., et al. (2021). A Small Molecule that Binds an RNA Repeat Expansion Stimulates Its Decay via the Exosome Complex. Cell Chem. Biol. *28*, 34-45.e6. 10.1016/j.chembiol.2020.10.007.
12. Soragni, E., Petrosyan, L., Rinkoski, T.A., Wieben, E.D., Baratz, K.H., Fautsch, M.P., and Gottesfeld, J.M. (2018). Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy. Invest. Ophthalmol. Vis. Sci. *59*, 1888–1896. 10.1167/iovs.17-23265.
13. Westin, I.M., Viberg, A., Byström, B., and Golovleva, I. (2021). Lower Fractions of TCF4 Transcripts Spanning over the CTG18.1 Trinucleotide Repeat in Human Corneal Endothelium. Genes *12*, 2006. 10.3390/genes12122006.
14. Sirp, A., Leite, K., Tuvikene, J., Nurm, K., Sepp, M., and Timmusk, T. (2020). The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters. Sci. Rep. *10*, 18424. 10.1038/s41598-020-75437-3.
15. Sepp, M., Kannike, K., Eesmaa, A., Urb, M., and Timmusk, T. (2011). Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. PloS One *6*, e22138. 10.1371/journal.pone.0022138.
16. Zweier, C., Peippo, M.M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I., et al. (2007). Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am. J. Hum. Genet. *80*, 994–1001. 10.1086/515583.
17. Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A., et al. (2007). 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. *80*, 988–993. 10.1086/515582.
18. Hellwig, M., Lauffer, M.C., Bockmayr, M., Spohn, M., Merk, D.J., Harrison, L., Ahlfeld, J., Kitowski, A., Neumann, J.E., Ohli, J., et al. (2019). TCF4 (E2-2) harbors tumor suppressive functions in SHH medulloblastoma. Acta Neuropathol. (Berl.) *137*, 657–673. 10.1007/s00401-019-01982-5.
19. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B., et al. (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744–747. 10.1038/nature08186.
20. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976. 10.1038/ng.940.
21. Wieben, E.D., Aleff, R.A., Tosakulwong, N., Butz, M.L., Highsmith, W.E., Edwards, A.O., and Baratz, K.H. (2012). A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. PLOS ONE *7*, e49083. 10.1371/journal.pone.0049083.
22. Vasanth, S., Eghrari, A.O., Gapsis, B.C., Wang, J., Haller, N.F., Stark, W.J., Katsanis, N., Riazuddin, S.A., and Gottsch, J.D. (2015). Expansion of CTG18.1 Trinucleotide Repeat in TCF4 Is a Potent Driver of Fuchs' Corneal Dystrophy. Invest. Ophthalmol. Vis. Sci. *56*, 4531–4536. 10.1167/iovs.14-16122.
23. Peh, G.S.L., Chng, Z., Ang, H.-P., Cheng, T.Y.D., Adnan, K., Seah, X.-Y., George, B.L., Toh, K.-P., Tan, D.T., Yam, G.H.F., et al. (2015). Propagation of human corneal endothelial cells: A novel dual media approach. Cell Transplant. *24*, 287–304. 10.3727/096368913X675719.

24. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
25. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. 10.1093/bioinformatics/bts635.
26. Witten, D.M. (2011). Classification and clustering of sequencing data using a Poisson model. Ann. Appl. Stat. *5*, 2493–2518. 10.1214/11-AOAS493.
27. Blighe, K. (2023). PCAtools: everything Principal Component Analysis.
28. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.
29. Ignatiadis, N., and Huber, W. (2021). Covariate powered cross-weighted multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. *83*, 720–751. 10.1111/rssb.12411.
30. Ignatiadis, N., Klaus, B., Zaugg, J.B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods *13*, 577–580. 10.1038/nmeth.3885.
31. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084–2092. 10.1093/bioinformatics/bty895.
32. Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-Seq data. Nat. Preced., 1–1. 10.1038/npre.2012.6837.2.
33. Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. *111*, E5593–E5601. 10.1073/pnas.1419161111.
34. Hwang, J.Y., Jung, S., Kook, T.L., Rouchka, E.C., Bok, J., and Park, J.W. (2020). rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. Nucleic Acids Res. *48*, W300–W306. 10.1093/nar/gkaa237.
35. Veiga, D.T. (2022). maseR: Mapping Alternative Splicing Events to pRoteins.
36. Gordon, S.P., Tseng, E., Salamov, A., Zhang, J., Meng, X., Zhao, Z., Kang, D., Underwood, J., Grigoriev, I.V., Figueroa, M., et al. (2015). Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. PLOS ONE *10*, e0132628. 10.1371/journal.pone.0132628.
37. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. 10.1093/bioinformatics/bty191.
38. Tardaguila, M., Fuente, L. de la, Marti, C., Pereira, C., Pardo-Palacios, F.J., Risco, H. del, Ferrell, M., Mellado, M., Macchietto, M., Verheggen, K., et al. (2018). SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. *28*, 396–411. 10.1101/gr.222976.117.
39. Pontikos, N., Yu, J., Moghul, I., Withington, L., Blanco-Kelly, F., Vulliamy, T., Wong, T.L.E., Murphy, C., Cipriani, V., Fiorentino, A., et al. (2017). Phenopolis: an open platform for harmonization and analysis of genetic and phenotypic data. Bioinformatics *33*, 2421–2423. 10.1093/bioinformatics/btx147.
40. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. 10.1101/gr.107524.110.
41. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. 10.1186/s13059-016-0974-4.
42. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014).

A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315. 10.1038/ng.2892.

43. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. 10.1038/s41586-020-2308-7.
44. Glusman, G., Caballero, J., Mauldin, D.E., Hood, L., and Roach, J.C. (2011). Kaviar: an accessible system for testing SNV novelty. Bioinformatics *27*, 3216–3217. 10.1093/bioinformatics/btr540.
45. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535-548.e24. 10.1016/j.cell.2018.12.015.
46. Zuallaert, J., Godin, F., Kim, M., Soete, A., Saeys, Y., and De Neve, W. (2018). SpliceRover: interpretable convolutional neural networks for improved splice site prediction. Bioinformatics *34*, 4180–4188. 10.1093/bioinformatics/bty497.
47. Shen, S., Park, J.W., Huang, J., Dittmar, K.A., Lu, Z., Zhou, Q., Carstens, R.P., and Xing, Y. (2012). MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. *40*, e61. 10.1093/nar/gkr1291.
48. Park, J.W., Tokheim, C., Shen, S., and Xing, Y. (2013). Identifying Differential Alternative Splicing Events from RNA Sequencing Data Using RNASeq-MATS. In Deep Sequencing Data Analysis Methods in Molecular Biology., N. Shomron, ed. (Humana Press), pp. 171–179. 10.1007/978-1-62703-514-9_10.
49. Dalton, C.J., and Lemmon, C.A. (2021). Fibronectin: Molecular Structure, Fibrillar Structure and Mechanochemical Signaling. Cells *10*, 2443. 10.3390/cells10092443.
50. Konieczny, P., Stepniak-Konieczna, E., and Sobczak, K. (2014). MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res. *42*, 10873–10887. 10.1093/nar/gku767.
51. Konieczny, P., Stepniak-Konieczna, E., and Sobczak, K. (2018). MBNL expression in autoregulatory feedback loops. RNA Biol. *15*, 1–8. 10.1080/15476286.2017.1384119.
52. Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845. 10.1038/nrg3813.
53. Li, M., Zhuang, Y., Batra, R., Thomas, J.D., Li, M., Nutter, C.A., Scotti, M.M., Carter, H.A., Wang, Z.J., Huang, X.-S., et al. (2020). HNRNPA1-induced spliceopathy in a transgenic mouse model of myotonic dystrophy. Proc. Natl. Acad. Sci. *117*, 5472–5477. 10.1073/pnas.1907297117.
54. Mori, K., Lammich, S., Mackenzie, I.R.A., Forné, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., et al. (2013). hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. (Berl.) *125*, 413–423. 10.1007/s00401-013-1088-7.
55. Laurent, F.-X., Sureau, A., Klein, A.F., Trouslard, F., Gasnier, E., Furling, D., and Marie, J. (2012). New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. Nucleic Acids Res. *40*, 3159–3171. 10.1093/nar/gkr1228.
56. Pettersson, O.J., Aagaard, L., Andrejeva, D., Thomsen, R., Jensen, T.G., and Damgaard, C.K. (2014). DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. Nucleic Acids Res. *42*, 7186–7200. 10.1093/nar/gku352.
57. Chu, Y., Hu, J., Liang, H., Kanchwala, M., Xing, C., Beebe, W., Bowman, C.B., Gong, X., Corey, D.R., and Mootha, V.V. (2020). Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. Nucleic Acids Res. *48*, 6740–6758. 10.1093/nar/gkaa422.

58. Wieben, E.D., Aleff, R.A., Tang, X., Kalari, K.R., Maguire, L.J., Patel, S.V., Baratz, K.H., and Fautsch, M.P. (2018). Gene expression in the corneal endothelium of Fuchs endothelial corneal dystrophy patients with and without expansion of a trinucleotide repeat in TCF4. PLOS ONE *13*, e0200005. 10.1371/journal.pone.0200005.
59. Doostparast Torshizi, A., Armoskus, C., Zhang, H., Forrest, M.P., Zhang, S., Souaiaia, T., Evgrafov, O.V., Knowles, J.A., Duan, J., and Wang, K. (2019). Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. Sci. Adv. *5*, eaau4139. 10.1126/sciadv.aau4139.
60. Sznajder, Ł.J., Thomas, J.D., Carrell, E.M., Reid, T., McFarland, K.N., Cleary, J.D., Oliveira, R., Nutter, C.A., Bhatt, K., Sobczak, K., et al. (2018). Intron retention induced by microsatellite expansions as a disease biomarker. Proc. Natl. Acad. Sci. U. S. A. *115*, 4234–4239. 10.1073/pnas.1716617115.
61. Liskova, P., Dudakova, L., Evans, C.J., Rojas Lopez, K.E., Pontikos, N., Athanasiou, D., Jama, H., Sach, J., Skalicka, P., Stranecky, V., et al. (2018). Ectopic GRHL2 Expression Due to Non-coding Mutations Promotes Cell State Transition and Causes Posterior Polymorphous Corneal Dystrophy 4. Am. J. Hum. Genet. *102*, 447–459. 10.1016/j.ajhg.2018.02.002.
62. Davidson, A.E., Liskova, P., Evans, C.J., Dudakova, L., Nosková, L., Pontikos, N., Hartmannová, H., Hodaňová, K., Stránecký, V., Kozmík, Z., et al. (2016). Autosomal-Dominant Corneal Endothelial Dystrophies CHED1 and PPCD1 Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of OVOL2. Am. J. Hum. Genet. *98*, 75–89. 10.1016/j.ajhg.2015.11.018.
63. Wright, C.F., Quaife, N.M., Ramos-Hernández, L., Danecek, P., Ferla, M.P., Samocha, K.E., Kaplanis, J., Gardner, E.J., Eberhardt, R.Y., Chao, K.R., et al. (2021). Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. Am. J. Hum. Genet. *108*, 1083–1094. 10.1016/j.ajhg.2021.04.025.
64. ProQR Therapeutics (2022). Open-Label, Single-Dose, Exploratory Study With QR-504a to Evaluate Safety, Tolerability, and Corneal Endothelium Molecular Biomarker(s) in Subjects With Fuchs Endothelial Corneal Dystrophy With Trinucleotide Repeat Expansion in the TCF4 Gene (FECD3) (clinicaltrials.gov).
65. Powers, A., Rinkoski, T.A., Cheung, K., Schehr, H., Osgood, N., Livelo, C., Levin, N., Safadi, M., Kerr, J., Zhang, C., et al. (2022). GeneTAC™ small molecules reduce toxic nuclear foci and restore normal splicing in corneal endothelial cells derived from patients with Fuchs endothelial corneal dystrophy (FECD) harboring repeat expansions in transcription factor 4 (TCF4). Invest. Ophthalmol. Vis. Sci. *63*, 2753-A0242.
66. Goyer, B., Thériault, M., Gendron, S.P., Brunette, I., Rochette, P.J., and Proulx, S. (2018). Extracellular Matrix and Integrin Expression Profiles in Fuchs Endothelial Corneal Dystrophy Cells and Tissue Model. Tissue Eng. Part A *24*, 607–615. 10.1089/ten.tea.2017.0128.
67. Thériault, M., Gendron, S.P., Brunette, I., Rochette, P.J., and Proulx, S. (2018). Function-Related Protein Expression in Fuchs Endothelial Corneal Dystrophy Cells and Tissue Models. Am. J. Pathol. *188*, 1703–1712. 10.1016/j.ajpath.2018.03.014.
68. Wang, Q., Dou, S., Zhang, B., Jiang, H., Qi, X., Duan, H., Wang, X., Dong, C., Zhang, B.N., Xie, L., et al. (2022). Heterogeneity of human corneal endothelium implicates lncRNA NEAT1 in Fuchs endothelial corneal dystrophy. Mol. Ther. - Nucleic Acids *27*, 880–893. 10.1016/j.omtn.2022.01.005.
69. Sirp, A., Roots, K., Nurm, K., Tuvikene, J., Sepp, M., and Timmusk, T. (2021). Functional consequences of TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual disability, and schizophrenia. J. Biol. Chem. *297*, 101381. 10.1016/j.jbc.2021.101381.
70. Tekendo-Ngongang, C., Grochowsky, A., Solomon, B.D., and Yano, S.T. (2021).

Beyond Trinucleotide Repeat Expansion in Fragile X Syndrome: Rare Coding and Noncoding Variants in FMR1 and Associated Phenotypes. Genes *12*, 1669. 10.3390/genes12111669.

71. Nath, S., Caron, N.S., May, L., Gluscencova, O.B., Kolesar, J., Brady, L., Kaufman, B.A., Boulianne, G.L., Rodriguez, A.R., Tarnopolsky, M.A., et al. (2022). Functional characterization of variants of unknown significance in a spinocerebellar ataxia patient using an unsupervised machine learning pipeline. Hum. Genome Var. *9*, 1–12. 10.1038/s41439-022-00188-8.

72. Liu, F., Liu, Q., Lu, C.X., Cui, B., Guo, X.N., Wang, R.R., Liu, M.S., Li, X.G., Cui, L., and Zhang, X. (2016). Identification of a novel loss-of-function C9orf72 splice site mutation in a patient with amyotrophic lateral sclerosis. Neurobiol. Aging *47*, 219.e1-219.e5. 10.1016/j.neurobiolaging.2016.07.027.