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

Short title (69/70 characters): Transcriptomic signatures of a common triplet repeat disease

Authors

Nihar Bhattacharyya[1], Niuzheng Chai[1*], Nathaniel J Hafford-Tear[1*], Amanda N Sadan[1*], Anita Szabo[1*], Christina Zarouchlioti[1], Jana Jedlickova[2], Szi Kay Leung[3], Tianyi Liao[1], Lubica Dudakova[2], Pavlina Skalicka[2,4], Mohit Parekh[1], Ismail Moghul[1,5], 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

21 Email: alice.davidson@ucl.ac.uk

22 *These authors contributed equally to the manuscript

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

Abstract: (Max 300 words)

(274/300)

Fuchs endothelial corneal dystrophy (FECD) is an age-related cause of vision loss, and the most common repeat expansion-mediated disease in humans characterised to date. Up to 80% of European FECD cases have been attributed to expansion of a non-coding CTG repeat element (termed CTG18.1) located within the ubiquitously expressed transcription factor encoding gene, *TCF4*. The non-coding nature of the repeat and the transcriptomic complexity of *TCF4* have made it extremely experimentally challenging to decipher the molecular mechanisms underlying this disease. Here we comprehensively describe CTG18.1 expansion-driven molecular components of disease within primary patient-derived corneal endothelial cells (CECs), generated from a large cohort of CTG18.1-expanded (Exp+) and CTG 18.1-independent (Exp-) FECD cases. We employ long-read, short-read, and spatial transcriptomic techniques to interrogate expansion-specific transcriptomic biomarkers. Interrogation of long-read sequencing and alternative splicing analysis of short-read transcriptomic data together reveals the global extent of altered splicing occurring within Exp+ FECD, and unique transcripts associated with CTG18.1-expansions. Similarly, differential gene expression analysis highlights the total transcriptomic consequences of Exp+ FECD within CECs. Furthermore, differential exon usage, pathway enrichment and spatial transcriptomics reveal *TCF4* isoform ratio skewing solely in Exp+ FECD with potential downstream functional consequences. Lastly, exome data from 134 Exp- FECD cases identified rare (minor allele frequency <0.005) and potentially deleterious (CADD>15) *TCF4* variants in 7/134 FECD Exp- cases, suggesting that *TCF4* variants independent of CTG18.1 may increase FECD risk. In summary, our study supports the hypothesis that at least two distinct pathogenic mechanisms, RNA toxicity and *TCF4* isoform-specific dysregulation, both underpin the pathophysiology of FECD. We anticipate these data

51 will inform and guide the development of translational interventions for this common triplet-

52 repeat mediated disease.

53

54 **Author's summary:**

55

56 **200/200**

57 Fuchs endothelial corneal dystrophy (FECD) leads to vision loss and is the most common
58 repeat expansion-mediated disease characterised to date; most patients harbour a non-coding
59 CTG repeat expansion within *TCF4*. FECD represents an important paradigm for other rare and
60 devastating neurological repeat-mediated diseases, given its relatively mild, age-related, and
61 tissue-specific nature. Intriguingly, despite *TCF4* being ubiquitously expressed, FECD patients
62 only experience corneal disease, and the biological reason for this tissue-specificity remains
63 elusive. Here, we use tissue from 25 FECD patients to perform complementary long-read, short-
64 read and spatial transcriptomic analyses of diseased cells to enhance our understanding of
65 mechanisms underpinning this disease. These data highlight that at least two mechanisms,
66 RNA toxicity and *TCF4* isoform-specific dysregulation, concurrently underlie the disease state in
67 affected corneal cells. Furthermore, *TCF4* isoform skewing, with evidence of downregulation,
68 suggests this mechanism in part may explain the unique vulnerability of the cornea. In addition,
69 7/134 molecularly unsolved FECD expansion negative cases were identified to harbour rare and
70 potentially deleterious *TCF4* variants, further supporting the hypothesis that dysregulation of
71 TCF4 may be key to FECD pathophysiology. Biological insights presented here will guide the
72 development of personalised FECD therapies and may inform the development of repeat-
73 expansion mediated therapies more broadly.

74 **Introduction:**

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

87

88 Up to 80% of FECD cases of European descent have one or more expanded copies of a
89 CTG trinucleotide repeat element (termed CTG18.1; MIM #613267) located within an intronic
90 region of the transcription factor encoding gene *TCF4* (3,5). We have previously demonstrated,
91 in our large (n=450) FECD cohort, that an expanded copy of this CTG repeat (defined as ≥50
92 copies) confers >76-fold increased risk for developing FECD (6). Expansion of non-coding
93 repeat elements within the human genome have been identified to cause more than 30 human
94 diseases to date (7), including fragile X tremor ataxia syndrome (MIM #300623), myotonic
95 dystrophy types 1 and 2 (MIM #160900 and #602668) and *C9orf72*-associated amyotrophic
96 lateral sclerosis (ALS) and frontotemporal dementia (MIM #105550)(7). Expansion-positive
97 (Exp+) FECD represents by far the most common disease within this category. However, to

date, it has only been established that it affects the cornea, which is in contrast to the majority of the other 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 pathology (7,8).

We, and others, have demonstrated that splicing dysregulation is a key feature of FECD in the presence of one or more expanded CTG18.1 alleles (5,6,9–11). Alongside this RNA induced toxicity, repeat-associated non-AUG (RAN) dependent translation of the expanded repeat, and dysregulation of *TCF4* itself have also been hypothesised to contribute to the underlying pathology of CTG18.1 Exp+ mediated FECD. However, a comprehensive understanding of consequences and role of both the spliceopathy and downstream effects of the *TCF4* dysregulation have yet to be characterised in parallel in the same model system. Furthermore, because *TCF4* comprises >90 alternative spatially and temporally expressed isoforms, deciphering transcript-specific level dysregulation has proven to be extremely challenging and a consistent pattern of *TCF4* dysregulation has yet to be associated with CTG18.1 expansions in CECs (12–14). Such diversity is reflected in the varied functional roles TCF4 plays in human development, alongside the range of human disease associated with *TCF4* variants in addition to FECD. For example, *TCF4* haploinsufficiency results in Pitt-Hopkins syndrome (MIM #610954)(15,16), a severe neurodevelopmental disorder, whereas somatic *TCF4* mutations are commonly identified in Sonic hedgehog medulloblastoma (17) and common germline variants have been associated with Schizophrenia risk (18,19).

In this study, we used a large cohort of biologically independent primary FECD case-derived CECs to isolate the effects of the multiple parallel pathogenic processes underlying

124 CTG18.1-mediated FECD. We present long-read transcriptomic data to enhance our
125 understanding of the increased isoform diversity observed in Exp+ FECD, and spatial
126 transcriptomics to demonstrate *TCF4*-isoform dysregulation. For the first time, we also isolate
127 downstream transcriptomic signatures of dysregulation that we hypothesise to be driven by
128 *TCF4*-isoform dysregulation that may, in part, explain the tissue specificity of the disease. We
129 go on to explore the potential relevance of TCF4 dysregulation to FECD, independent of the
130 CTG18.1 expansion, by interrogating expansion-negative FECD cases for rare *TCF4* variants
131 through exome sequencing and subsequently applying a gene-burden style approach.
132
133

Results

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

To further increase our understanding of CTG18.1 expansion-associated splicing dysregulation events, and better understand the relationship between individual events to transcript-specific isoforms, we generated long-read RNA-seq data from biologically independent unpassaged primary CEC cultures derived from healthy control (n=4. mean age 65.3 years) and Exp+ FECD (mono-allelic CTG18.1 allele repeat length ≥50) (n=4, mean age 66.8 years) tissues (**Table 1**). 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 analysed with SQANTI3 to assess isoform diversity (**Figure 1A,B and Table 2, S1**). We also 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 1**). Exome data generated from the Exp- FECD cases were analysed 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 visualisation and PCA analysis revealed that Exp+ samples grouped separately from not only the controls, but also the Exp- samples **(Figure 1C, D)**. Notably, Exp-

156 samples also clustered together on the PCA plot, but the heatmap derived from hierarchical
157 clustering highlighted that greater intra-group variability existed between Exp- samples
158 compared to variability within control and Exp+ groups (**Figure 1C**). We hypothesise that the
159 increased transcriptomic diversity within the Exp- group is due to potentially distinct, as yet
160 unidentified, genetic causes driving disease for each FECD case-derived sample. We included
161 all three groups in the downstream comparative pipelines using the following pairwise
162 comparisons; **(PWC1)** control versus Exp+ FECD, **(PWC2)** Exp+ FECD versus Exp- FECD, and
163 **(PWC3)** control versus Exp- FECD.
164
165

166 **Table 1: Summary of FECD case and control-derived corneal endothelial cell (CEC)**
167 **cultures analysed by short and long-read RNA-seq, and spatial transcriptomics.**
168

Sample	Sex	Age*	Ethnicity**
Long-read			
Control 1	Female	68	White-European
Control 2	Male	>50	Undeclared
Control 3	Female	74	Undeclared
Control 4	Male	69	White-European
FECD1exp+ (43/90)	Female	74	White-European
FECD2exp+ (14/80)	Male	78	White-European
FECD3exp+ (18/86)	Female	70	White-European
FECD4exp+ (26/78)	Female	45	Undeclared
Short-read			
Control 1	Female	68	White-European
Control 5 Control 6	Male	55	White-European
Control 7	Female	56	Black-African
FECD5exp+ (16/86)	Female	67	White-European
FECD6exp+ (12/106)	Male	69	White-European
FECD7exp+ (16/94)	Female	54	White-European
FECD1exp- (13/18)	Female	64	European/Black Caribbean
FECD2exp- (18/24)	Female	73	White-European
FECD3exp- (22/28)	Male	53	Black-African
Spatial transcriptomics			
Control 8 (12/16)	Female	45	Black-African
Control 9 (15/33)	Female	46	White-European
Control 10 (25/26)	Male	70	White-European
FECD4exp- (15/15)	Female	48	White-European
FECD5exp- (16/23)	Female	57	Undeclared
FECD6exp- (15/16)	Female	58	Undeclared
FECD8exp+ (63/88)	Female	75	Undeclared
FECD9exp+ (12/76)	Male	61	Undeclared
FECD10exp+ (18/67)	Male	65	Undeclared

CTG18.1 expansion status is denoted in brackets for all FECD-case derived CEC lines analysed. *Age at time of corneal tissue excision. **Self-reported by affected individual

169
170
171
172
173
174
175
176

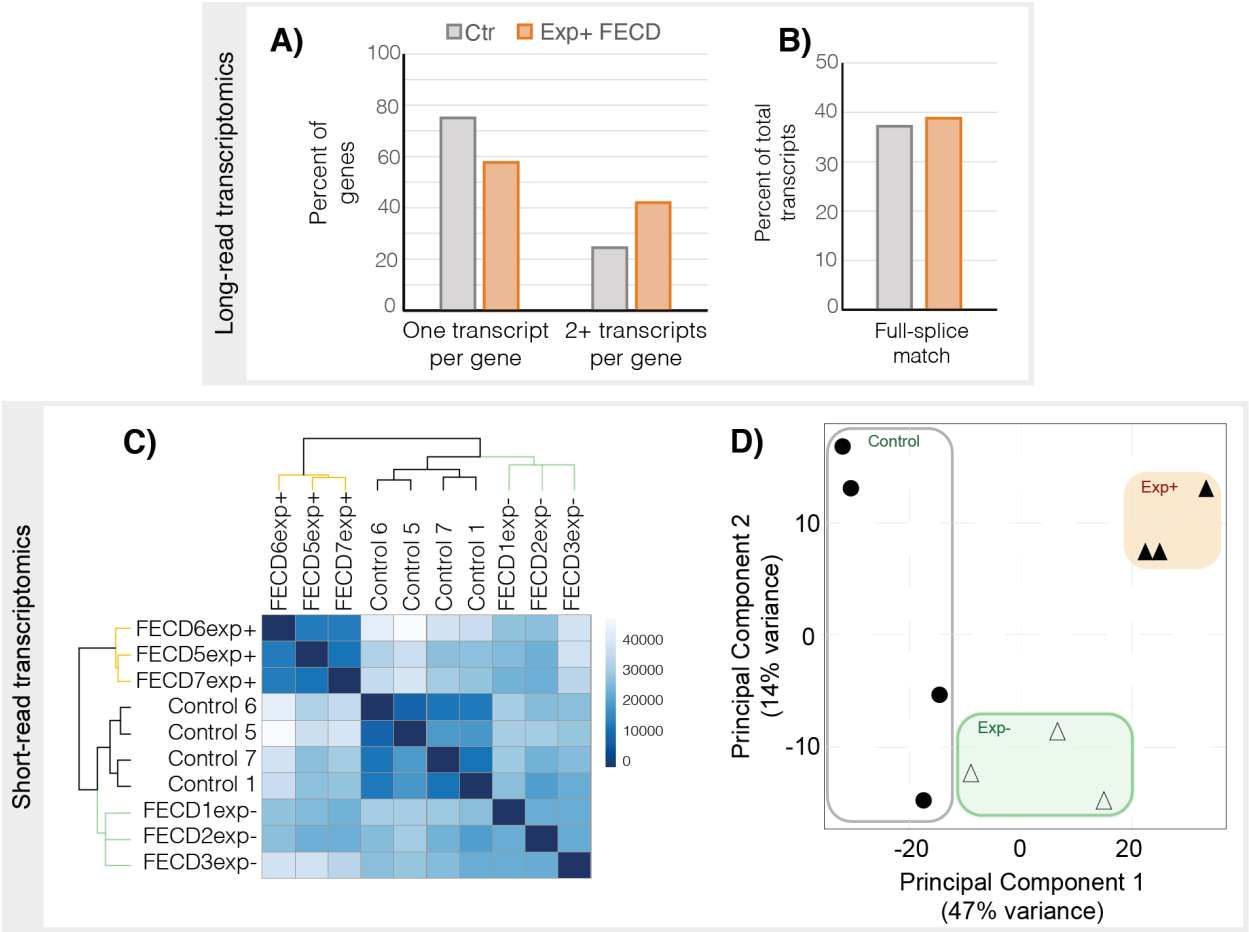


Figure 1: Exp+ FECD samples are transcriptomically distinct in both short- and long-read datasets when compared to control and Exp- samples. A) Bar chart showing proportion of annotated genes in long read transcriptomic dataset with only one transcript detect vs 2 or more transcripts in Control and Exp+ samples. Exp+ long-read samples have a higher proportion of genes with more than one isoform identified. **B)** Proportion of annotated transcripts with full splice matches. Exp+ FECD and Control samples have similar percentages of full splice matches suggesting additional splicing diversity seen in Exp+ is not predominantly due to aberrant splicing. **C)** Hierarchical clustering heatmap of sample-to-sample distances within short-read RNA-Seq data. Darker colours 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. **D)** Principal component plot of all short-read samples. All three subsets of data cluster separately in the PCA plot.

193 ***Alternative splicing analysis of long-read and short-read transcriptomic data***
194 ***demonstrate the complexity of increased splicing events underlying CTG18.1 expansion-***
195 ***mediated disease***
196
197
198 Iso-seq and subsequent SQANTI3 analysis (**Table 2, S1**) revealed a higher proportion of
199 isoforms per gene in Exp+ samples (1.7 unique isoforms per unique gene in Exp+ vs 1.3 in
200 control). Additionally we see a higher proportion of known non-canonical splice junctions in
201 Exp+ samples. When characterising transcripts based on splice junctions, we see similar
202 proportions of isoform subtypes (**Table S1**). This is consistent with previous work (5,9–11) and
203 validated by our short read data (**Figure S2**), whereby Exp+ samples showed greater isoform
204 diversity compared to control and Exp- samples, but not an overall skew in the types of
205 alternative splicing observed.
206
207 **Table 2: Summary of SQANTI3 analysis of IsoSeq data.** In-depth characterisation of isoforms
208 in control and Exp+ long-read RNA-seq data utilising SQANTI3 pipeline following stringent
209 quality control. SQANTI3 subdivides splice junctions into known (present in reference
210 transcriptome) and novel (not in reference transcriptome). Canonical splice junctions refer to the
211 most common type of splice junction (with GT donor and AG acceptor sites), while non-
212 canonical refers to other more rare and tissue specific splice junctions with different donor and
213 acceptor sites

	Control (n=4)		**Exp+** FECD (n=3)	
Genes	2,318		4,652	
Isoforms	3,113		8,032	
Gene Classification				
	# of genes		# of genes	
Annotated Genes	2,309		4,642	
Novel Genes	9		10	
Splice Junction Classification				
	# splice junctions	Percent	# splice junctions	Percent
Known canonical	27,552	97.18	52,153	97.35
Known non-canonical	1	0.00001	21	0.04
Novel canonical	798	2.81	1,400	2.61
Novel non-canonical	0	0	0	0

214

215 To further characterise and quantify alternative splicing events within our short-read
216 dataset, we utilised rMATS, a junction-based method that employs a hierarchical framework, to
217 model unpaired datasets and evaluate and quantify all major alternative splice types (20–22).
218 Comparing Exp+ to control samples (PWC1) we identified 2,118 significant alternative splicing
219 events in 1,437 genes (**Figures S1, S2, Table S2)**. Alternative splicing comparisons including
220 Exp+ (PWC1 and PWC2) yielded the most alternative splice types in every category (**Figure S1,**
221 **S2**) further reinforcing the spliceopathy aspects of this genetic subtype of disease.
222 We then compared the rMATS and Iso-seq results to a subset of published datasets
223 exploring differential splicing in FECD patient-derived samples (**Table S3**) and selected 24
224 published differential splicing events (from 23 genes) that had previously been identified by
225 bioinformatic or *in vitro* methods to be strongly associated with CTG18.1 expansion-mediated
226 FECD(3,5,6). Alternative splicing analysis of our transcriptomic dataset identified that 75.0%
227 (18/24) of these events were significant in PWC1 (control vs Exp+, **Table S3**). In addition to
228 these 24 selected events, rMATS identified another 10 novel and significant differential splicing
229 events within Exp+ in these genes, highlighting the increasing complexity of splicing
230 dysregulation within Exp+ FECD. Long-read sequencing confirmed the presence of these novel
231 splicing events in Exp+ FECD (**Table S4**). Also, it was interesting to note that *FN1* was
232 identified to have robust alternative splicing in Exp+, and further investigation showed that
233 alternatively spliced exons in Exp+ corresponded with skipping of the Extra Domain A (EDA)
234 and Extra Domain B (EDB) protein domains that play a critical role in signal transduction and
235 fibrillar formation (23) (**Table S5**).
236 Skipped exon alternative splicing events were the most common in all three pairwise
237 comparisons (**Figure S1**). Prior work regarding alternative splicing in FECD has focused on
238 MBNL1/2 sequestration(5,6,24) as MBNL proteins are key regulators of exon splicing(25,26).
239 Given their complex role, we attempted to determine if the flanking intronic sequences of
240 alternatively spliced exons were enriched for common RNA-binding protein (RBP) binding

241 motifs. We focussed our RBP motif enrichment analysis (rMAPS2(27)) on the 250bp region
242 upstream of alternatively spliced exons and isolated those areas where there was enrichment in
243 PWC1 and PWC2, but not PWC3 (i.e. events enriched within Exp+ CEC lines). By filtering
244 rMAPS results in this manner, a trend whereby skipped exons are significantly ($p < 0.05$)
245 upregulated by RNA-binding proteins with TT-containing binding motifs and downregulated by
246 RNA-binding proteins with CG-containing binding motifs was observed (**Table S6**). As MBNL
247 proteins favour YCGY binding motifs, this suggests that other RNA-binding proteins with TT-
248 containing motifs could also play a role in driving the spliceopathy underpinning a key
249 mechanism of disease in Exp+ FECD.
250
251 ***Evidence of TCF4 dysregulation appears enriched in CTG18.1-mediated FECD,***
252 ***highlighting specific candidate biomarkers for this genetic subtype of disease***
253
254 Differential gene expression was assessed with DESeq2(28) in all three groups. For all
255 comparisons an FDR corrected P-value (padj) of <0.05 was used as a threshold for significance
256 (**Table S7**). A total of 4,399 genes (16.7% of all mapped genes) were identified to be
257 differentially expressed between the Exp+ and controls (PWC1; **Figure 2A**). Of these, 1,905
258 genes had a shrunkLFC value (accurate log_2fold values of expression accounting for high
259 dispersion and low expression counts) of greater than 1 or less than -1. Overall, PWC2 and
260 PWC3 groups resulted in lowered differential gene expression, with 2,273 and 2,118
261 significantly differentially expressed genes, respectively **(Figure 2A; Table S7)**.
262 Pairwise analysis revealed considerable overlap between dysregulated genes identified
263 by all three pairwise comparisons **(Figure 2A),** however, highest convergence was observed by
264 PWC1 (Exp+ versus control) and PWC2 (Exp+ v Exp-), suggesting these events are driven by
265 CTG18.1 expansions. In total, both pairwise comparisons reveal 1,150 dysregulated genes
266 uniquely and significantly dysregulated in Exp+ FECD case-derived samples. Conversely, by

267 comparing the overlap between PWC1 and PWC3 (Exp+ vs control and Exp- vs control) we
268 identified a further 1,105 genes that are dysregulated in FECD, irrespective of the underlying
269 genetic cause, and thus representing more ubiquitous signatures of the disease. Dysregulated
270 genes in PWC1, including Exp+ unique and common to FECD, did not skew directionally and
271 were evenly up or down-regulated within our analysis (**Figure 2C**). We isolated 3,183 genes
272 that are uniquely dysregulated in Exp+ FECD, 2,033 genes dysregulated solely in PWC1 and a
273 further 1,150 dysregulated genes shared with PWC2; (highlighted in orange **Figure 2A**) and
274 proceeded with downstream analysis (**Table S7**).

275 We conducted pathway analysis using g:profiler(29) (querying gene ontology, KEGG
276 pathway, and Reactome pathways). Gene ontology and KEGG pathway analysis of PWC1
277 dysregulated genes was consistent with previously published transcriptomes(5,10,11). While
278 total *TCF4* gene expression was not significantly downregulated in Exp+ (**Figure 2B**),
279 intriguingly, we observed strong parallels to pathway enrichment analyses highlighted from a
280 study of *TCF4* neural progenitor knockdown cell model(30) within the 3,183 genes uniquely
281 dysregulated in Exp+, such as protein processing in endoplasmic reticulum, MAPK signalling,
282 actin cytoskeletal regulation, endocytosis and apoptosis (**Figure 2D**). Critically, these pathways
283 are not enriched in PWC3, suggesting that these are unique to Exp+ CECs. Furthermore,
284 several enriched pathways involving Wnt signalling were seen in the 3,183 dysregulated genes
285 unique to Exp+ that are not present in PWC3, again suggesting *TCF4* function may be modified
286 within Exp+ FECD CECs (**Table S9**). Furthermore, enrichment of dysregulated genes in
287 pathways involving other repeat expansion-mediated diseases and misfolded protein aggregate
288 diseases suggest mechanistic parallels which reinforce Exp+ CECs as a model system with
289 broad relevance to repeatome biology (**Table S9**).

290 The consistent presence of PI3K-Akt pathway enrichment within GO, KEGG, and
291 Reactome results in Exp+ dysregulated genes (and its consistent absence in PWC3 pathway
292 enrichment) suggests that Exp+ FECD CECs may interact with growth factors and the

293 extracellular matrix in a unique manner. Hence, components of the PI3K-Akt pathway may
294 serve as informative biomarkers for Exp+ FECD in future clinical trial settings, and thus warrant
295 further investigation. Additional pathways are enriched in this set of genes (and not in PWC3)
296 that involve RNA processing and binding, other repeat-mediated disease related mechanisms,
297 epithelial to mesenchymal transition, and multiple signal transduction cascades with
298 extracellular matrix interactions (**Table S9**).

299 To determine if other RNA-binding proteins may be implicated in the alternative splicing
300 seen in Exp+ FECD, we cross-referenced our short-read data with a list of ~1500 human RNA-
301 binding proteins(31), and discovered that there were 207 RBPs uniquely dysregulated in Exp+
302 (59 downregulated and 148 upregulated), including *MBNL1, CELF1, ESRP2,* and a variety of
303 hnRNP and DDX encoding genes (**Table S8**). Both the hnRNP and DDX family of RNA-binding
304 proteins have previously been implicated in other repeat expansion diseases(29,32–34). It is
305 likely that a combination of these dysregulated RBPs plays a role in driving the alternative
306 splicing seen in Exp+ FECD.
307

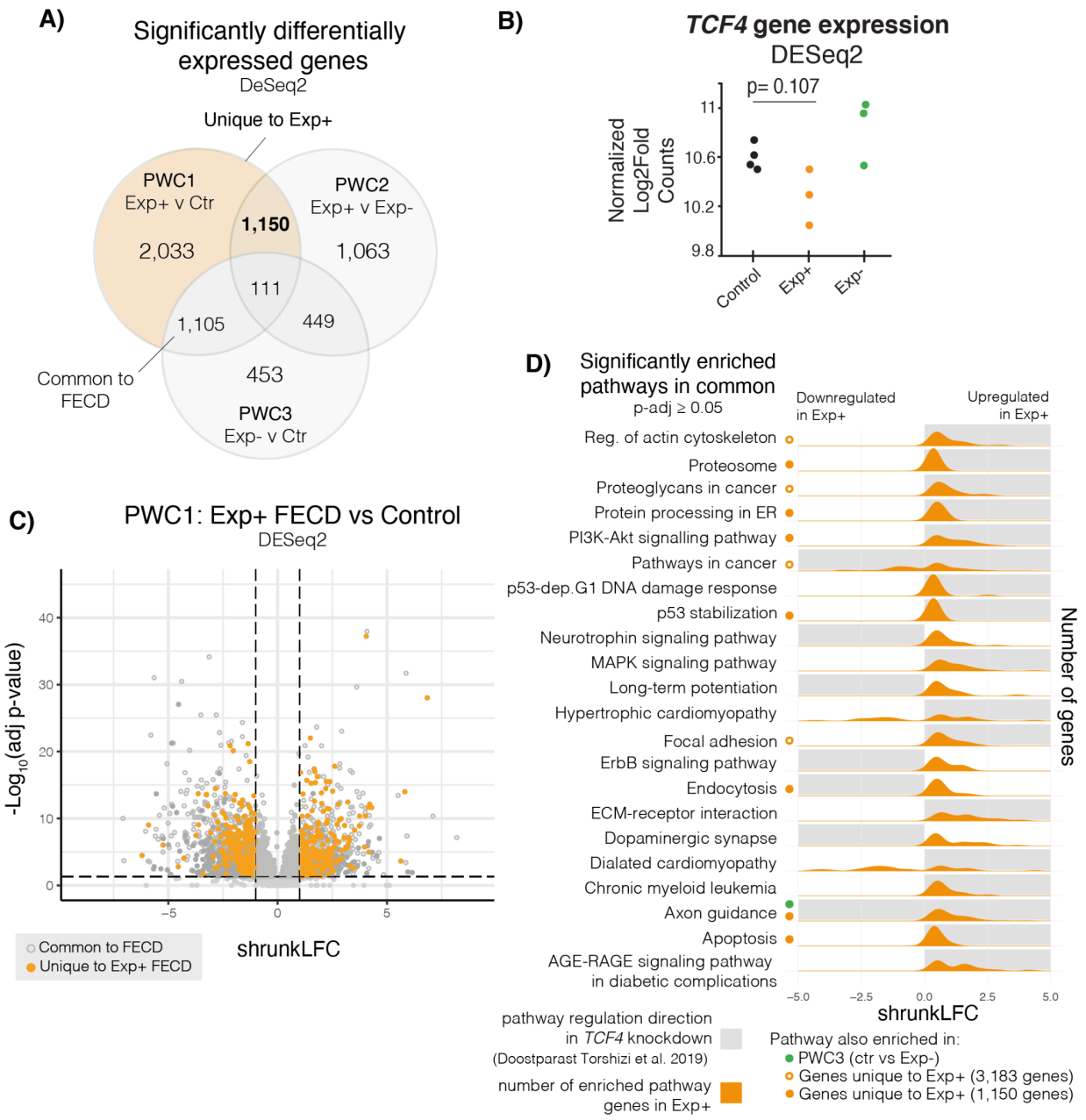


Figure 2: Differential gene expression analysis on Exp+ FECD, Exp- FECD, and control primary CEC transcriptomes. **A)** Differential gene expression was assessed via DESeq2. Three pairwise comparisons were conducted: PWC1 (Exp+ vs control), PWC2 (Exp+ vs Exp-), and PWC3 (Exp- vs control). Venn diagrams show overlap of significantly differentially expressed genes (adjusted p-value < 0.05) within these three comparisons. **B)** Total *TCF4* expression showed no significant dysregulation between the different groups. **C)** 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 grey solid markers show significantly expressed genes with a fold change magnitude greater than 1. Solid orange markers delineate uniquely Exp+ differentially expressed genes. **D)** Pathways enriched in Exp+

vs Control (g:profiler). Shaded grey boxes show direction of pathway enrichment in TCF4 knockdown models (Doostparast Torshizi et al. 2019). Green dots show pathways that are significantly enriched in PWC3 as well. Solid orange dots and open orange circle show that pathway enrichment persists in genes unique to Exp+ (1,150 genes and 3,183 genes, respectively)

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 (**Figure 2D**), 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 2B).** Due to transcriptional diversity and complexity of TCF4 transcripts, DEXSeq2(35) was used to assess differential exon usage utilising our short-read transcriptomic data.

DEXSeq2 utilises generalised 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 S3**), suggesting these patterns of differential exon usage are likely attributed to CTG18.1-related pathology.

Our short-read transcriptomic data reveals that multiple *TCF4* exons (n=15) were significantly dysregulated in Exp+ compared to control and Exp- FECD samples (PWC1 results shown in **Figure 3A, Table S10**), whereas no significant differences were observed between Exp- and control (PWC3, **Figure 3B, Table S10**), demonstrating for the first time that differential *TCF4* exon usage is driven by CTG18.1 expansions and not a downstream feature of FECD unrelated to the primary genetic driver. 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

348 broadly expressed exons are downregulated (appearing in 64 out of 93 transcripts) (**Figure 3A,**
349 **Table S10**). Transcripts containing downregulated exons are primarily included in longer
350 isoforms of the gene which contain 3 AD domains, while the upregulated exons (excluding those
351 surrounding the repeat) appear in shorter *TCF4* isoforms containing only two AD domains. This
352 suggests that Exp+ CECs may be employing certain compensatory measures in response to
353 the transcript-specific dysregulation that is induced by the presence of CTG18.1 repeat
354 expansions. Interestingly, our rMATS analysis independently detects this same phenomenon
355 (**Table S11**), which showed a significant mutually exclusive exon splicing event whereby the last
356 exon in common with most 3 AD domain-containing isoforms (Exon 6 in ENST000000354452.8)
357 was favoured in the control group, while the first exon in common with all isoforms (Exon 7 in
358 ENST000000354452.8) was favoured in the Exp+ group.
359 We also detect significant retention of the intron flanking the CTG18.1 repeat, previously
360 described by Sznajder et al. 2018(36) in Exp+ FECD **(Figure 3A).** The intronic region is
361 significantly upregulated in Exp+ CECs (E173 and E174) in both PWC1 and PWC2 excluding
362 PWC3 **(Figure 3A, Table S10)**. While this intron represents the 5' untranslated region (UTR) of
363 several transcripts of *TCF4*, it is also included within an intronic region in 32 out of 93 currently
364 annotated *TCF4* transcripts.
365

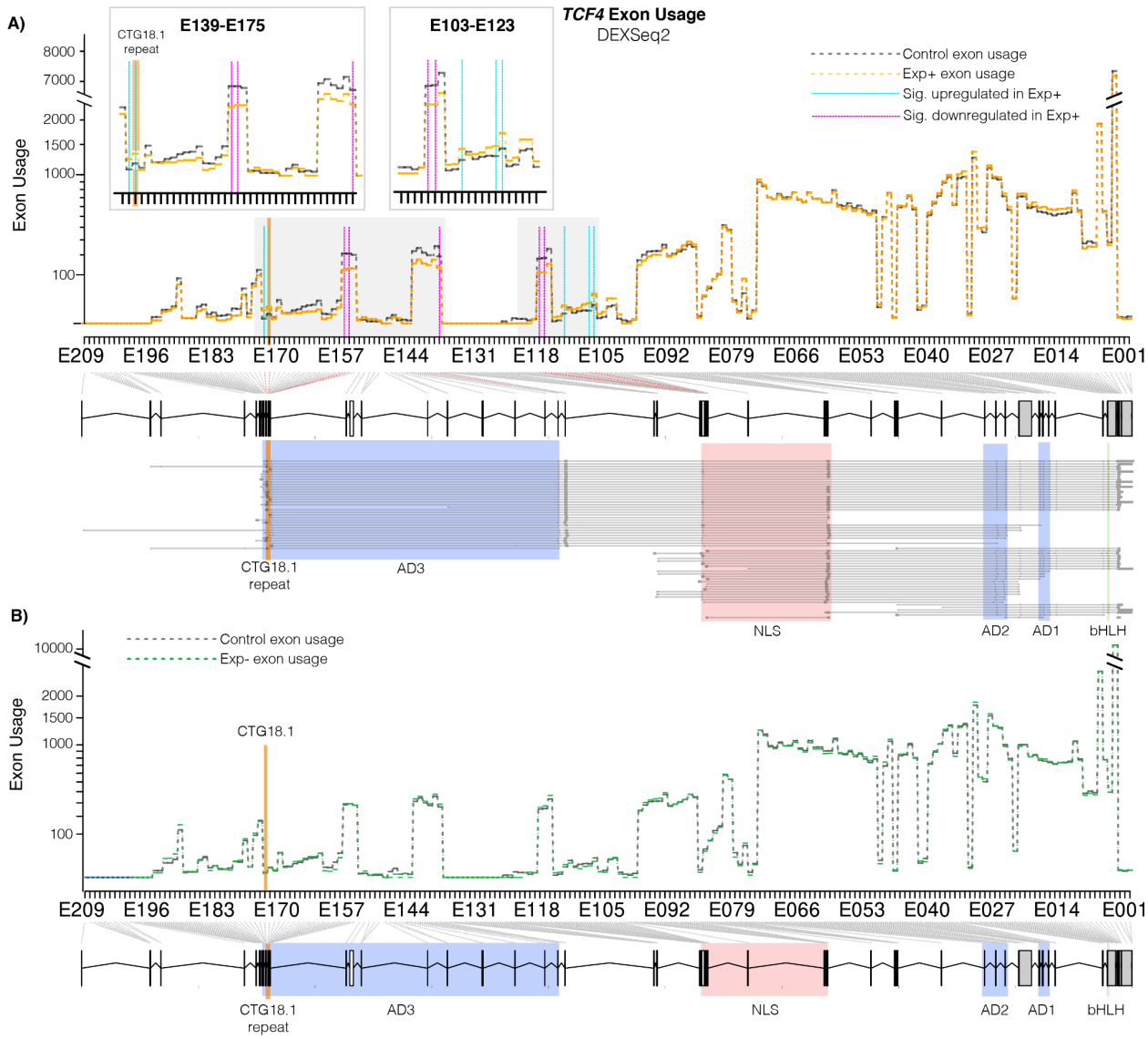


Figure 3: CTG18.1 expansion-mediated FECD results in alternative *TCF4* exon usage within corneal endothelial cells. A) Differences in exon usage between control (grey dotted line) and Exp+ (orange dotted line) CECs. Significantly upregulated and downregulated exons are highlighted with blue and red 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 (grey dotted line) and Exp- (green dotted line) CECs. No exons were statistically dysregulated.

381 ***RNAScope shows robust detection of altered TCF4 isoform expression in CTG18.1-***
382 ***mediated FECD case-derived corneal endothelial cells***
383
384 We performed spatial transcriptomics, via RNAScope, to further validate our hypothesis
385 that a subset of *TCF4* isoforms containing the third AD3 domain are being selectively
386 downregulated due to the presence of an expanded CTG18.1 allele (≥50 repeats). For the first
387 time, we visualised *TCF4* transcripts within Exp+ FECD, Exp- FECD, and control CECs using
388 two distinct sets of probes, termed A and B (**Figure 4A**). Probe set A comprises a set of probes
389 that allow the visualisation of all *TCF4* transcripts in the cell (probe A covers all exons in the
390 transcript Isoform B+ which contains 3 AD domains, TCF4-201, ENST00000354452.8,
391 NM_001083962.2). Probe set B targets exons in the canonical longer TCF4-201 *TCF4*
392 transcript which do not overlap with the canonical shorter *TCF4* transcript (Isoform A+/J which
393 contains 2 AD domains, AD1 and AD2, TCF4-204, ENST00000457482.7, NM_001243234.2).
394 Hence, these probe sets enable the detection, visualisation, and quantification of the ratio of
395 longer AD3-containing transcripts, previously identified to be downregulated in Exp+ CECs via
396 our differential exon usage analysis, compared to total *TCF4* transcripts (**Figure 3A**).
397 The *TCF4* transcripts were visualised in control, Exp+ and Exp- FECD case-derived
398 CECs (n=3 for each group) (**Figure 4B, Table 1 and S12)**. Each independent primary CEC
399 culture was also probed with a negative control (targeting bacterial genes) and a $(CAG)_7$-Cy3
400 probe, using a standard FISH protocol, to validate the presence or absence of CUG-specific
401 RNA foci derived from expanded copies of CTG18.1 (**Table S12**). RNA foci were consistently
402 detected in all CTG18.1 Exp+ CECs, and were absent from control and Exp- CECs. Using
403 CellProfiler (version 4), puncta representing *TCF4* transcripts, detected with both probe set A
404 and B, were quantified and puncta/cell values were established for each image and averaged
405 across images with at least 100 nuclei imaged per condition per line. Finally, a ratio was

406 calculated establishing the proportion of longer AD3 containing transcripts compared to total
407 *TCF4* transcripts for each biological replicate. This showed that, on average, Exp+ CECs had a
408 significantly smaller proportion of the longer AD3-containing transcripts (16.33% ± 4.53)
409 compared to unaffected control CECs (36% ± 6.9, p=0.0145) and non-expanded FECD CECs
410 (30.9% ± 5.17, p=0.0213). This result suggests that the presence of the expanded repeat alone
411 consistently skews *TCF4* transcript ratios to favour shorter isoforms. As this subgroup of
412 isoforms typically contains only two out of three activation domains, a higher prevalence of
413 shorter TCF4 isoforms may affect potential dimerization and TCF4 transcriptional function in
414 Exp+ CECs.
415

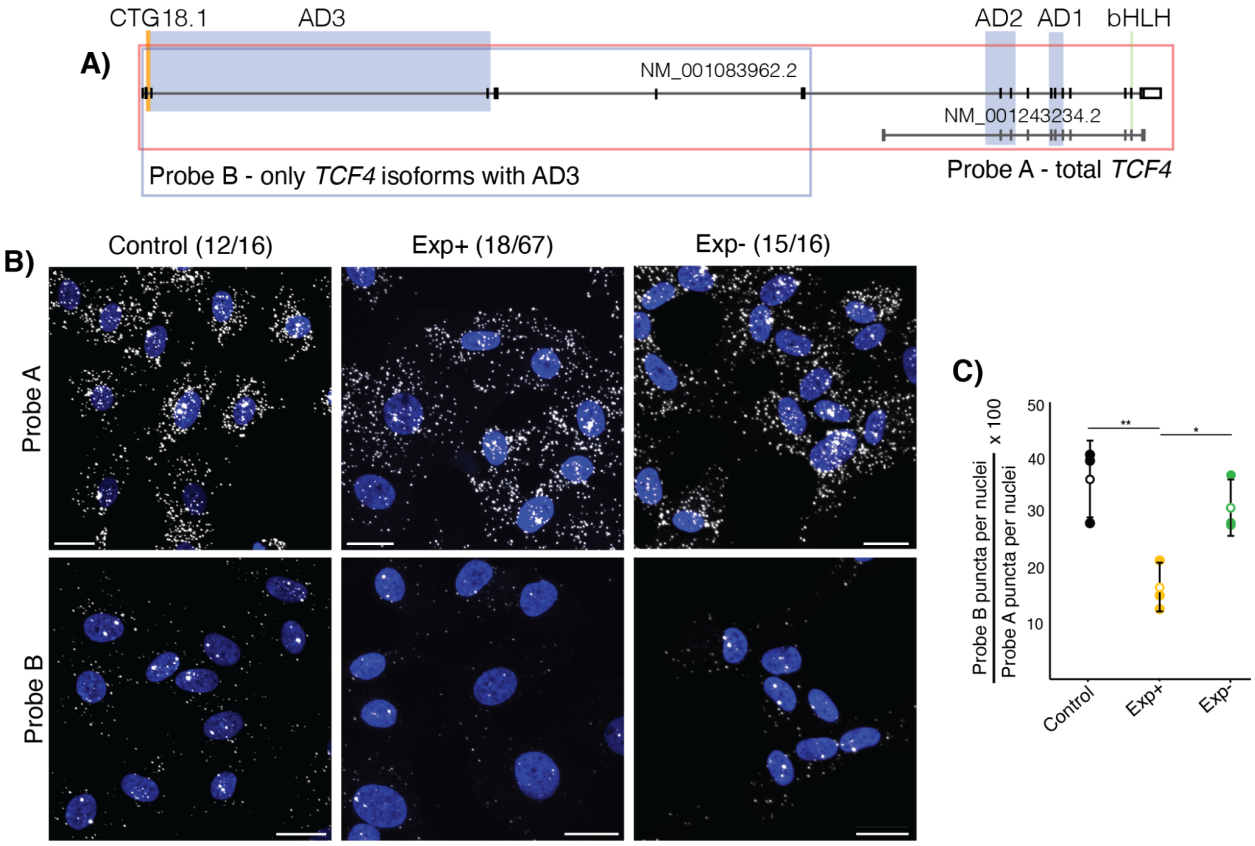


Figure 4: RNAScope probes targeting *TCF4* transcripts demonstrate that longer transcripts containing AD3 account for a smaller proportion of total transcripts in Exp+ FECD CECs. A) Probe targeting schematic for *TCF4*. Probe A targets all exons in NM_001083962.2 and visualises total *TCF4* transcripts, while Probe B targets only the 5' exons, excluding those that are contained within shorter transcripts (NM_001243234.2). Probe B visualises 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 (visualised with Cy3 but represented here with white). Scale bars represent 20 µm. **C)** Plotting the proportion of longer AD3-containing *TCF4* transcripts in Exp+ (orange), control (black), and Exp- (green) shows that Exp+ have significantly lower proportion of AD3-containing transcripts compared to both control ($p = 0.0145$) and Exp- ($p = 0.0213$) CECs, while there is no significant difference between control and Exp- ($p = 0.362$). Open circle denotes mean and error bars show standard deviation. Significance determined with student T-test.

439 ***Rare TCF4 variants identified within a genetically refined CTG18.1 expansion-negative***
440 ***FECD cohort suggest isoform-specific TCF4 dysregulation may be a risk factors for***
441 ***FECD in the absence of CTG18.1 expansions***
442
443 We next wanted to explore if rare and potentially deleterious *TCF4* variants could be
444 contributing to disease in FECD cases without a CTG18.1 expansion, given that Exp+ CECs
445 display alternative *TCF4* exon usage and skewed ratios of *TCF4* isoforms (**Figure 3 and 4**),
446 with detectable downstream features of TCF4-driven dysregulation (**Figure 2D**). To test this
447 hypothesis we interrogated exome data generated from 134 Exp-FECD cases, acquired from
448 our ongoing inherited corneal disease research program, representing approximately 22% of
449 total FECD cases recruited to date. Overall, 8 rare and potentially deleterious (MAF <0.005;
450 CADD score >15) *TCF4* variants were identified in 7/134 probands (**Table 3**). Sanger
451 sequencing confirmed the presence of all 8 variants and exome data from these cases were
452 also interrogated using the same thresholds (MAF <0.005, CADD >15) to determine if any other
453 rare variants in genes previously associated with FECD could also be contributing to disease (3)
454 (**Table S13**). The large number of rare and potentially deleterious *TCF4* variants detected in this
455 cohort seemed intriguing given the high levels of evolutionary constraint on this essential and
456 ubiquitously expressed transcription factor encoding gene (pLI=1, missense Z=4.1 gnomAD).
457 We therefore decided to apply CoCoRV, a rare variant analysis framework to determine if the
458 Exp-FECD cohort was indeed enriched for rare and potentially deleterious *TCF4* variants
459 compared to the gnomAD dataset. CoCoRV utilises the publicly available genotype summary
460 counts to prioritise disease-predisposition genes in case cohorts (37). Under a dominant model,
461 applying the ethnically-stratified analysis pipeline, and filtering for rare (MAF <0.005) and
462 potentially deleterious (CADD score threshold of >15) variants, an inflation factor estimate
463 (λemp) 1.21 was achieved. At the exome wide level *TCF4* was ranked as 34/17,025 genes with
464 a p-value of 0.001 suggesting the Exp-FECD cohort could be enriched for rare and potentially

465 deleterious *TCF4* variants. However, after applying CoCoRV's stringent false discovery rate
466 correction methodology *TCF4* no longer remained significant (corrected p-value = 0.249).
467 Interestingly rare variants identified in Proband A (Variant 1a/1b), B (Variant 2) and D
468 (Variant 4) and E (Variant 5) are all predicted to only affect a small fraction of total *TCF4*
469 transcripts (**Table 3, S14 and Figure 6**). For Proband A, two consecutive *in cis* heterozygous
470 missense and nonsense variants were identified. These occur within the coding region of only
471 one (out of 93) *TCF4* transcripts (ENST00000566286.5; c.[57G>T; 58A>T]; p.[(Arg19Ser;
472 p.Lys20*)]) and in the 5' UTR of a further 5/93 transcripts. The variant identified in Proband B is
473 only encompassed by the same 6 *TCF4* transcripts (ENST00000566286.5; c.66G>A,
474 p.(Glu22=)). Despite being synonymous, Splice AI predicts the variant introduces loss of the
475 splice donor site (SpliceAI Δ score 0.78). Furthermore, SpliceRover also predicts that this could
476 result in the activation of a cryptic splice donor site downstream (from 0.098 to 0.233) of the
477 wildtype donor site, which would introduce a short frameshift insertion followed by a premature
478 termination codon (PTC) c.66_67insGTGCTCGATGAATTTTC, p.(Arg23Valfs*12). The variant
479 identified in Proband D (Variant 4) is within the protein coding region of three (out of 93) *TCF4*
480 transcripts and in the 5' UTR of one additional transcript. Variant 5 is only encompassed by the
481 5' UTR of a single *TCF4* transcript. Collectively, all of these identified variants have the potential
482 to exert a functional impact on TCF4, but importantly only on a small subset of overall
483 transcripts.
484
485

486 **Table 3:** Summary of rare and potentially deleterious *TCF4* variants identified in a FECD
487 CTG18.1 Exp- cohort (n=134)

Variant ID	FECD Exp- Cases (n=134)	Variant (hg38)	HGVSc, HGVSp	CADD, MAF gnomAD	Included within CoCoRV analysis
1a	Proband A	chr18:55588479-C-A	ENST00000566286.1: c.57G>T, p.(Arg19Ser)	18.3, 0	Yes
1b	Proband A	chr18:55588478-T-A	ENST00000566286.1: c.58A>T, p.(Lys20Ter)	16.45, 0	Yes
2	Proband B	chr18:55588470-C-T	ENST00000566286.1: c.66G>A, p.(Glu22=)	18.39, 0.002122	Yes
3	Proband C	chr18:55464063-T-C	ENST00000354452.3: c.207+13A>G	20.6, 0.0002234	Yes
4	Proband D	chr18:55403689-A-G	ENST00000544241.2: c.26T>C, p.(Ile9Thr)	17.29, 0	Yes
5	Proband E	chr18:55321599-C-T	ENST00000354452.3: c.549+28760G>A	16.74, 0	No*
6	Proband F	chr18:55261560-A-C	ENST00000354452.3: c.923-27T>G	22.3, 0.0007644	Yes
7	Proband G	chr18:55261512-G-A	ENST00000354452.3: c.944C>T, p.(Ala315Val)	24.6, 0.0005734	Yes

Rare (MAF < 0.005 in gnomAD genomes) and potentially deleterious (CADD score threshold of >15) *TCF4* variants identified within a FECD CTG18.1 Exp- cohort (n=134). Abbreviations are as follows: gnomAD, The Genome Aggregation Database; 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. *Variant filtered out at Variant Quality Recalibration step.

488

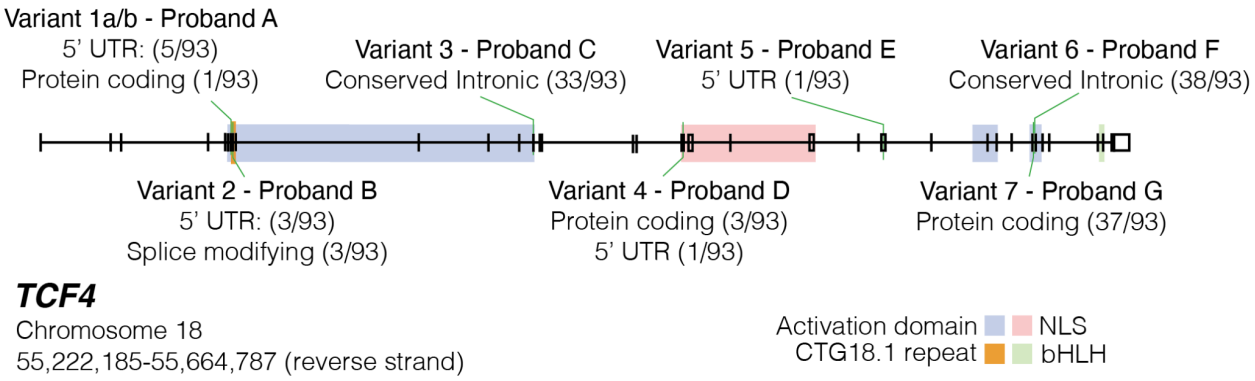


Figure 6: Schematic map of rare and potentially deleterious variants identified in molecularly unsolved FECD cases within *TCF4*, without CTG18.1 expansions. All 8 rare (MAF<0.005) and potentially pathogenic (CADD>15) variants in 7 Exp- Probands are mapped along *TCF4* genomic region and transcriptomic locations contextualised. Proband A variants are found in *cis* within the proband. Conserved intronic transcripts refer to variants located in intronic regions identified as conserved via GERP conservation scores on Ensembl. Splice modifying refers to transcripts where the variant identified would induce alternative splicing. Transcript numbers where the variant occurred within introns were not included. UTR: untranslated region. NLS: nuclear localization signal. bHLH: basic helix loop helix

Discussion:

FECD is now recognized to be the most common repeat expansion-mediated disease in humans, with up to 80% of European FECD cases harbouring CTG18.1 expansions(3). The non-coding nature of the repeat and the transcriptomic complexity of *TCF4* have made it extremely experimentally challenging to decipher the molecular mechanisms underlying this disease. Thus, many aspects of its pathophysiology remain elusive, despite the strong mechanistic parallels with other repeat-mediated diseases, and the urgent clinical need for new treatments for this common and visually disabling disease. To tackle this challenge, we employed a large, biologically independent set of primary FECD case-derived CEC lines and numerous complementary transcriptomic methodologies to gain novel mechanistic insight into both RNA toxicity driven and *TCF4* dysregulation-related molecular features of this repeat-mediated disease. By including both Exp+ and Exp- FECD patient data, we have been able to dissect out more generic downstream biomarkers that occur irrespective of CTG18.1 expansion

514 status. Our extensive long- and short-read RNA-Seq data and spatial transcriptomic analysis of
515 FECD case-derived CEC lines demonstrates that, in addition to being a spliceopathy,
516 expansion-positive FECD is concurrently associated with isoform-specific patterns of *TCF4*
517 dysregulation, hence providing novel mechanistic insights into the pathophysiology of this triplet-
518 repeat mediated disease. Furthermore, we have identified a subset of Exp- FECD cases with
519 rare and potentially deleterious *TCF4* variants, supporting the hypothesis that dysregulation of
520 TCF4 itself irrespective of CTG18.1 expansions may also be a risk factor for FECD.

521 Our robust transcriptomic analysis, presented within a biologically relevant cellular and
522 genomic context, highlights an array of novel molecular signatures driven by CTG18.1
523 expansions, as well as validating a range of previously described Exp+-associated
524 transcriptomic events. We anticipate that many of the novel events identified here could serve
525 as informative biomarkers to aid the development of CTG18.1 targeted therapeutic approaches.
526 To date, CUG-containing RNA foci and MBNL1/2 missplicing have been adopted as the
527 standard for therapeutic outcome measures (6,24,38,39). Data presented here support and
528 expand the hypothesis that accumulation of repeat expansion-derived RNA structures induce
529 toxic gain-of-function effects that disrupt cellular homeostasis and induce aberrant RNA
530 metabolism(6,24). For instance, we have discovered a novel array of RBP encoding transcripts,
531 including *MBNL3*, *CELF1* and a range of *hnRNP* and *DDX* family members, uniquely
532 dysregulated in Exp+ CECs that warrant future follow up to elucidate their respective
533 contributions to aberrant RNA metabolism related mechanisms. However, such biomarkers may
534 only relate to the spliceopathy aspects of the disease and our data go on to also illustrate that
535 translational efforts would also benefit from using more diverse outcome measures. In addition
536 to RNA metabolism biomarkers, global Exp+-mediated differential gene expression profiles,
537 alongside *TCF4*-specific dysregulation itself, should also be included as relevant outcome
538 measures for Exp+ therapies being developed, given we hypothesise these processes are also

539 pivotal to the pathology. Novel cell surface biomarkers reported here that are specifically
540 upregulated in Exp+ (e.g. *ITGA5* and *ITGB1*) could have particular utility in a clinical trial setting
541 due to their accessible locations. Other membrane encoding proteins interfacing with the
542 extracellular matrix or signal transduction could also be a rich source of potential external
543 markers, such as cell surface receptors involved in the P13K-Akt pathway (e.g. *EGFR, MET,*
544 and *LTK*) that we show are specifically upregulated in Exp+ CECs.

545 Genetic stratification of patient-derived CECs in this study has also highlighted that Exp+
546 FECD is transcriptionally distinct from Exp- disease. The data show that 72% of all differentially
547 expressed genes, 81% of skipped exon events, and 72% of all enriched pathways detected in
548 Exp+ are unique to the repeat expansion disease. For example, in accordance with previous
549 studies (11,40), we observe that pronounced upregulation of fibronectin (*FN1*) is a robust
550 transcriptomic feature of FECD, irrespective of genetic underlying cause. However, our
551 comprehensive transcriptomic analysis has enabled us to determine that the EDA/B+ forms of
552 fibronectin (23) are significantly favoured in Exp+ FECD. This illustrates that *FN1* expression
553 and splicing can act as both a general FECD molecular biomarker and an Exp+ specific
554 biomarker. Similarly, *AQP1* previously documented to be downregulated in FECD (41), has here
555 been demonstrated to only display marked downregulation in Exp+ CECs, illustrating the
556 importance of stratifying cases based on CTG18.1 status. Our long-read data revealed that the
557 splicing diversity seen in Exp+ FECD is not predominantly a result of abnormal missplicing,
558 given the proportion of annotated full splice match transcripts remains the same between control
559 and Exp+. This further exemplifies the unique transcriptome of Exp+ FECD endothelial cells.
560 However the increased proportion of non-canonical splice junctions seen in Exp+ also suggests
561 that CTG18.1-specific isoforms may be present within the endothelium and warrants future
562 investigation. These examples, alongside the emerging evidence that the Exp+ FECD displays
563 distinctive pathological features and epidemiological characteristics(4,6), suggests that Exp+

564 FECD should be categorised as a distinct clinical entity. We suggest integrating CTG18.1
565 genotyping into routine FECD clinical care pathways is warranted, especially given the
566 numerous international efforts underway to develop gene-directed therapies for CTG18.1-
567 mediated FECD (24,38,39,42).

568 Our complementary RNA-seq approaches, in combination with the spatial resolution of *TCF4*
569 transcripts subgroups, has highlighted a unique and distinctive pattern of TCF4-specific
570 dysregulation that occurs within Exp+ FECD. Additionally, our pathway analysis has isolated
571 features of dysregulation that we hypothesise represent downstream consequences of this
572 isoform shift and an important pathogenic component of the disease. Notably, we have
573 observed commonality between signalling transduction pathways dysregulated in our Exp+
574 CECs (e.g. P13K-Akt and MAPK) and those dysregulated in neural progenitor cell *TCF4*
575 knockdown transcriptomes(30). We postulate that this convergence is being driven by TCF4
576 functional dysregulation, as the shift in isoform distributions towards shorter (1-2 AD domains)
577 versus longer (2-3 AD domains) isoforms would result in lowered TCF4 activity within the cell.
578 TCF4 isoforms with 3 AD domains have been shown to have an order of magnitude more
579 transcription activation function than the shorter isoforms containing 1 or 2 activation domains
580 (14). Furthermore, given that domains in the AD3-containing isoforms can also dictate
581 intracellular location and dimerization partners of TCF4, we also hypothesise that additional
582 TCF4-driven functional implications of this shift occur within affected CECs that requires future
583 experimental investigation. Interestingly, a recent single-cell (sc)RNA-seq study has identified
584 that varying levels of *TCF4* expression exist within discrete populations of healthy CECs(43).
585 Future application of scRNA-seq to FECD-case derived CEC samples, alongside probing
586 mechanisms of somatic instability within the corneal endothelium will shed light upon the unique
587 vulnerability of this cell layer to CTG18.1 expansions (44,45).
588

589 Additionally, we identify rare and predicted deleterious *TCF4* variants in 7/134 Exp-
590 FECD cases. This was notable to us given the high levels of evolutionary constraint acting on
591 *TCF4* and suggests such rare variants may be acting as risk factors for FECD, independently of
592 CTG18.1 repeat expansion. This could also in-part explain the established missing heritability of
593 the disease(3) and does further reinforcing the hypothesis that dysregulation of *TCF4* itself is
594 pivotal to the pathophysiology of FECD. Despite a rare variant gene-burden style analysis of
595 these data not reaching significance, we observed an emerging trend whereby a subset of
596 qualifying variants identified notably affected regions of the gene not included by the vast
597 majority of *TCF4* isoforms. Hence they can only be predicted to exert a functional impact on a
598 small subset of total TCF4 transcripts. This observation is in keeping with the relatively mild and
599 tissue specific nature of FECD, which is in stark contrast to the severe neurodevelopmental
600 disorder Pitt-Hopkins associated with total *TCF4* haploinsufficiency(46). Hence, we propose that
601 *TCF4* variants which subtly affect TCF4 functionality may induce CEC-specific disease. Further
602 experimental investigation of these variants and the potential tissue specific role they may exert
603 in the corneal endothelium warrants further future investigation.
604
605 Parallels can be drawn to other repeat expansion-mediated disorders, such as fragile X
606 syndrome (*FRM1*)(47), spinocerebellar ataxia type 7 (*ATXN7,* MIM #164500)(48) and *C9orf72*-
607 mediated ALS(49), whereby rare loss-of-function mutations have been proposed to induce
608 comparable phenotypic outcomes to non-coding repeat expansions. However, in these
609 instances total haploinsufficiency is thought to underlie the respective pathogenic mechanisms,
610 in contrast to this scenario whereby we hypothesise only a fraction of total *TCF4* isoforms are
611 susceptible to the effect of either CTG18.1 expansions or *TCF4* variants within the corneal
612 endothelium. In a limited number of DM1 cases, FECD has been recognised to be part of the
613 phenotypic spectrum, suggesting that CEC dysfunction can also arise from CTG repeat
614 expansions in *DMPK* (50,51). Data presented in this study suggests that *TCF4 variants*,

615 independent of the repeat, may also represent a potential path to FECD. We suggest future
616 FECD genetic screening approaches should include sequencing of coding and regulatory
617 regions of *TCF4,* to further elucidate the potential role that rare *TCF4* variants may play in the
618 absence of CTG18.1 expansions.
619
620 In conclusion, this study supports the hypothesis that at least two distinct pathogenic
621 mechanisms, RNA toxicity and *TCF4* isoform-specific dysregulation, underpin the
622 pathophysiology of the transcriptionally distinct CTG18.1-mediated FECD. In our established
623 primary corneal endothelial system, we see evidence of *TCF4* isoform dysregulation at the
624 *TCF4* transcript level and can detect the potential downstream effects of the isoform-specific
625 pattern of TCF4 dysregulation. Our data also suggests that, in rare instances, irrespective of
626 CTG18.1 expansions, FECD cases can harbour *TCF4* variants that may independently induce
627 isoform-specific TCF4 dysregulation and act as risk factors for disease. This supports the
628 hypothesis that *TCF4* dysregulation is a key mechanism underlying this common, age-related
629 and visually disabling disease. We anticipate that concepts and biomarkers identified by this
630 study will inform and guide the development of future translational interventions targeting this
631 triplet-repeat mediated disease, and these in turn may act as exemplars for much rarer and
632 currently untreatable and severe repeat expansion disorders that typically affect less accessible
633 neurological/neuromuscular systems.
634
635

636 **Materials and Methods:**

637

638 ***Subject Recruitment, phenotyping and bio-sample processing***

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

653

654 ***CTG18.1 genotyping***

655 We determined the CTG18.1 repeat length using a short tandem repeat (STR)
656 genotyping assay as previously described in Zarouchlioti et al. 2018(6), originally adapted from
657 Wieben et al. 2012(52). Briefly, STR-PCR was performed using a 6'-FAM-conjugated primer (5'
658 CAGATGAGTTTGGTGTAAGAT 3') upstream of the repeat and an unlabeled primer (5'
659 ACAAGCAGAAAGGGGGCTGCAA 3') downstream. Post-PCR, products were combined with
660 GeneMarker ROX500 ladder and separated on an ABI 3730 Electrophoresis capillary DNA
661 analyzer (Applied Biosystems). Sizing and data analysis was performed on GeneMarker

662 software (SoftGenetics). To distinguish between samples with presumed bi-allelic non-
663 expanded alleles of equal length and samples with a larger repeat length beyond the detection
664 limit of the STR-PCR, we performed a further triplet-primed PCR assay to confirm the presence
665 or absence of a CTG18.1 allele above the detection range of the STR assay (approximately 125
666 repeats), following previously published methods(53).
667
668 ***Primary corneal endothelial cell (CEC) culture***
669 We isolated and cultured FECD and control-derived CECs as described by Zarouchlioti
670 et al. 2019(6), adapted from Peh et al. 2015(54). Briefly, to dissolve the DM and detach the
671 CECs, the tissue sample was incubated for 2 to 4 hours in 0.2% collagenase type I. After
672 incubation, the CECs were resuspended in stabilisation media (containing Human Endothelial-
673 SFM (Life technologies) supplemented with 5% foetal bovine serum (FBS), 1%
674 antibiotic/antimycotic, and 0.1% selective ROCK inhibitor Y-27632 (AdooQ BioSciences)), and
675 then seeded into 6-well plates pre-coated with FNC coating mixture (Stratech). After 24 hours,
676 we replaced the culture media with expansion media (Ham's F-12 Nutrient Mix with GlutaMAX
677 Supplement (Life Technologies)/Medium 199 GlutaMAX Supplement (Life Technologies), 20
678 µg/mL ascorbic acid, 1% insulin-transferrin-selenium (Life Technologies), 5% FBS, 1%
679 antibiotic/antimycotic, 10 ng/mL bFGF (Life Technologies) and 0.1% selective ROCK inhibitor Y-
680 27632 (AdooQ BioSciences)). All cells were cultured at 37ºC and 5% CO_2 with twice-weekly
681 media changes.
682
683 ***RNA extraction***
684 All primary CEC monolayer cultures selected for downstream analysis displayed a
685 consistent morphology characteristic of CECs *in vivo.* Total RNA was extracted from the primary
686 CEC lines using a NucleoSpin RNA XS kit (Macherey-Nagel), in accordance with the
687 manufacturer's protocol. The RNA was eluted and stored at -80°C prior to sequencing.

688

689 ***Short-read RNA-seq***

690 Stranded RNA-seq libraries were prepared by BGI Genomics using the TruSeq RNA
691 Library Prep Kit (Illumina) following the manufacturer's protocol. We sequenced these with
692 either a HiSeq 4000 platform (Illumina) with short-read strand-specific library prep with poly A
693 enrichment (Illumina). Paired-end reads were filtered to remove low-quality reads (Q ≤5) or
694 those with >10% of N base calls. Adaptor sequences were also removed during data quality
695 analysis. We checked the quality of the resulting fastq files for the short-read data using FastQC
696 Quantification(55). Reads were aligned to the human genome (hg38) using STAR(56)
697 (v2.7.10a). After alignment, hierarchical clustering was performed using PoiClaClu(57). We also
698 performed principal component analysis (PCA) using PCAtools(58) (v2.2.0) to determine global
699 sample clustering.

700

701 We analysed differential gene expression using DESeq2(28) (version 1.8.2), with
702 independent hypothesis weighting (IHW) testing of the results to generate FDR values(59,60),
703 and $\log_2$fold change was calculated using the apeglm (v 1.2) package(61). We then determined
704 significance as FDR corrected-p value (padj) <0.05. In addition, we analysed different exon
705 usage amongst the subgroups with DEXSeq2(35) (v 1.17.28). Significant results were filtered
706 using an FDR-corrected P-value of <0.05. Pathway analysis was performed by g:Profiler on
707 significantly dysregulated genes(62).

708

709 We assessed differential splicing at a junction-based level with rMATS-turbo 4.1.2(20)
710 with the Ensembl hg38 release 105 GTF file as reference. Default rMATS settings were used.
711 For this study, we focused on splicing events that are fully/partly annotated by Ensembl.

712 Significant alternative splicing events were identified as an absolute delta percent spliced in

713 (psi) magnitude ≥ 0.1 and FDR ≤ 0.05. rMATs output was subsequently analysed using

714 rMAPS2(27) and maser 1.12.1(63).

715

716 ***Long-read RNA-seq***

717 PacBio Sequel long-read transcriptomic sequencing (Iso-Seq)(64) was applied to

718 generate full-length cDNA sequences from control and FECD Exp+ primary CEC-derived RNA

719 samples. Raw reads were processed and de-multiplexed with Iso-Seq (v3) and merged for

720 transcript collapse using Cupcake(64) (parameters: "-c 0.85 -i 0.95 --dun-merge-5-shorter").

721 High-quality, full-length transcripts from the merged dataset were then aligned to the human

722 reference genome (hg38) using Minimap2(65) (v2.17, parameters: "-ax splice -uf --

723 secondary=no -C5 -O6,24 -B4") and annotated using SQANTI3(66) (v7.4). SQANTI filtering of

724 artefacts (intra-priming and reverse transcriptase template switching) was then applied.

725

726 ***RNAScope detection of TCF4 transcripts***

727 We seeded primary FECD case-derived CECs on chamber slides (Lab-tek) and

728 confluent CEC cultures were stained using standard RNAScope protocols with minor

729 adjustments (ACD Bio). We used the Hs-TCF4-C2 (cat. 557411) probe set to detect all *TCF4*

730 isoforms, as well as a custom set of *TCF4* probes to detect all 5' exons in the *TCF4* transcript

731 NM_001083962.2 excluding all 3' exons contained in the shorter transcript NM_001243234.2.

732 Akoya TSA Cy3 fluorophores were used to visualise probe staining. Negative control probes

733 (cat. 320871) were provided by ACD-Bio to verify specificity. All CEC lines analysed by

734 RNAScope were also tested for RNA foci in parallel via fluorescence *in situ* hybridization (FISH)

735 using a CUG-specific probe (Cy3-$(CAG)_7$)(6). A minimum of 100 cells and four images were
736 acquired for each cell line and each probe.
737
738 ***Exome Sequencing and Rare Variant Analysis***
739 Exome sequencing libraries were generated from 134 Exp- FECD patient whole blood
740 derived gDNA samples, using a SureSelect Human All Exome V6 capture kit (Agilent, USA) or
741 SeqCap EZ MedExome Enrichment Kit (Roche) and sequenced on a HiSeq 4000 or HiSeq
742 2500 platform (Illumina). All raw FECD case sequencing data was aligned and annotated
743 following our previously published methods(67).In brief, sequencing reads were aligned using
744 Novoalign (version 3.02.08) and . Variants and indels were called according to Genome
745 Analysis Toolkit (GATK) (version 4.4.0.0) best practices (joint variant calling followed by variant
746 quality score recalibration) were normalised and quality-controlled, as recommended by
747 CoCoRV(37). CoCoRV, a rare variant analysis framework, was applied to prioritise disease-
748 predisposition genes in the Exp-FECD cohort, utilising publicly available genotype summary
749 counts from gnomAD (genomes version 2.1.1) as a control dataset. For consistency, both case
750 and control variants were annotated using the Variant Effect Predictor (VEP) (release 110) (68).
751 The CoCoRV pipeline was then run in accordance to default setting, applying a dominant model
752 and ethnically stratified in accordance with ethnicity predictions generated from the case cohort
753 using Fraposa(69). Variants of interest were defined as having a CADD score (CADD) >15 (70),
754 and a minor allele frequency (MAF) <0.005 in the Genome Aggregation Database (version
755 2.1.1) (71). FECD cases identified to harbour rare *TCF4* variants were additionally interrogated
756 for rare and potentially disease-associated variants (MAF<0.005 in gnomAD and Kaviar
757 Genomic Variant Database; CADD>15) in previously reported FECD-associated genes,
758 including *COL8A2*, *SLC4A11*, *AGBL1* and *ZEB1*(3). Splice site predictions were subsequently
759 made using SpliceAI(72) and SpliceRover(73) for all qualifying variants in *TCF4* and any other
760 FECD-associated genes.

761

762

763

764 **Author Contributions:** NB and AED conceptualised the study. NB, NJHT, ANS, CZ, JJ, LD,
765 PS, MP, KM, PL, SJT contributed to the generation of the data. NB, NC, NJHT, AS, SKL, TL,
766 IM, ARJ, MEC, AJH, NP and AED all contributed to the analysis and interpretation of the data.
767 AED and NB wrote the manuscript. All authors read and approved the final version of the
768 document.
769
770 **Acknowledgments:** We would like to thank all affected individuals for participating in this
771 research. We would also like to acknowledge Miracles for Sight Eye bank for providing control
772 corneal tissue. This work was funded by a UKRI Future Leader Fellowship MR/S031820/1
773 (AED, CZ, AS), Moorfields Eye Charity GR000060, GR001395, GR001337 (AED, NB, NHT, AS,
774 NP) and Sight Research UK SAC 036 (AED, AJS), the Rosetrees Trust (AED, AJS), and the
775 The National Institute for Health Research Biomedical Research Centre at Moorfields Eye
776 Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology (AED,
777 NC., SJT, KM, AJH, MEC, NP). JJ, LD., PS, and PL were supported by GACR 20-19278S,
778 UNCE/MED/007 and SVV 260631. NP was supported by an NIHR AI Award (AI
779 AWARD02488). This project utilised equipment funded by the UK Medical Research Council
780 (MRC) Clinical Research Infrastructure Initiative (award number MR/M008924/1).
781
782
783
784

References:

1. Lorenzetti DW, Uotila MH, Parikh N, Kaufman HE. Central cornea guttata. Incidence in the general population. Am J Ophthalmol. 1967 Dec;64(6):1155–8.
2. Matthaei M, Hribek A, Clahsen T, Bachmann B, Cursiefen C, Jun AS. Fuchs Endothelial Corneal Dystrophy: Clinical, Genetic, Pathophysiologic, and Therapeutic Aspects. Annu Rev Vis Sci. 2019;5(1):151–75.
3. Fautsch MP, Wieben ED, Baratz KH, Bhattacharyya N, Sadan AN, Hafford-Tear NJ, et al. TCF4-mediated Fuchs endothelial corneal dystrophy: Insights into a common trinucleotide repeat-associated disease. Prog Retin Eye Res. 2021 Mar 1;81:100883.
4. Thaung C, Davidson AE. Fuchs endothelial corneal dystrophy: current perspectives on diagnostic pathology and genetics—Bowman Club Lecture. BMJ Open Ophthalmol. 2022 Jul 1;7(1):e001103.
5. Wieben ED, Baratz KH, Aleff RA, Kalari KR, Tang X, Maguire LJ, 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. 2019 Aug 30;60(10):3636–43.
6. Zarouchlioti C, Sanchez-Pintado B, Hafford Tear NJ, Klein P, Liskova P, Dulla K, et al. Antisense Therapy for a Common Corneal Dystrophy Ameliorates TCF4 Repeat Expansion-Mediated Toxicity. Am J Hum Genet. 2018 Apr 5;102(4):528–39.
7. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am J Hum Genet. 2021 May 6;108(5):764–85.
8. Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. Nat Rev Neurol. 2018 Sep;14(9):544–58.
9. Wieben ED, Aleff RA, Tang X, Butz ML, Kalari KR, Highsmith EW, et al. Trinucleotide Repeat Expansion in the Transcription Factor 4 (TCF4) Gene Leads to Widespread mRNA Splicing Changes in Fuchs' Endothelial Corneal Dystrophy. Invest Ophthalmol Vis Sci. 2017 Jan 24;58(1):343–52.
10. Wieben ED, Aleff RA, Tang X, Kalari KR, Maguire LJ, Patel SV, et al. Gene expression in the corneal endothelium of Fuchs endothelial corneal dystrophy patients with and without expansion of a trinucleotide repeat in TCF4. PLOS ONE. 2018 Jul 2;13(7):e0200005.
11. Chu Y, Hu J, Liang H, Kanchwala M, Xing C, Beebe W, et al. Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. Nucleic Acids Res. 2020 Jul 9;48(12):6740–58.
12. Westin IM, Viberg A, Byström B, Golovleva I. Lower Fractions of TCF4 Transcripts Spanning over the CTG18.1 Trinucleotide Repeat in Human Corneal Endothelium. Genes. 2021 Dec 17;12(12):2006.
13. Sirp A, Leite K, Tuvikene J, Nurm K, Sepp M, Timmusk T. The Fuchs corneal dystrophy-associated CTG repeat expansion in the TCF4 gene affects transcription from its alternative promoters. Sci Rep. 2020 Oct 28;10(1):18424.
14. Sepp M, Kannike K, Eesmaa A, Urb M, Timmusk T. Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. PloS One. 2011;6(7):e22138.
15. Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, et al. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet. 2007 May;80(5):994–1001.
16. Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, et al. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet. 2007 May;80(5):988–93.

17. Hellwig M, Lauffer MC, Bockmayr M, Spohn M, Merk DJ, Harrison L, et al. TCF4 (E2-2) harbors tumor suppressive functions in SHH medulloblastoma. Acta Neuropathol (Berl). 2019 Apr 1;137(4):657–73.
18. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature. 2009 Aug 6;460(7256):744–7.
19. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011 Sep 18;43(10):969–76.
20. Shen S, Park JW, Lu Z xiang, Lin L, Henry MD, Wu YN, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci. 2014 Dec 23;111(51):E5593–601.
21. Shen S, Park JW, Huang J, Dittmar KA, Lu Z xiang, Zhou Q, et al. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. 2012 Apr 1;40(8):e61.
22. Park JW, Tokheim C, Shen S, Xing Y. Identifying Differential Alternative Splicing Events from RNA Sequencing Data Using RNASeq-MATS. In: Shomron N, editor. Deep Sequencing Data Analysis [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2022 Oct 31]. p. 171–9. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-62703-514-9_10
23. Dalton CJ, Lemmon CA. Fibronectin: Molecular Structure, Fibrillar Structure and Mechanochemical Signaling. Cells. 2021 Sep;10(9):2443.
24. Hu J, Rong Z, Gong X, Zhou Z, Sharma VK, Xing C, et al. Oligonucleotides targeting TCF4 triplet repeat expansion inhibit RNA foci and mis-splicing in Fuchs' dystrophy. Hum Mol Genet. 2018 Mar 15;27(6):1015–26.
25. Konieczny P, Stepniak-Konieczna E, Sobczak K. MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res. 2014;42(17):10873–87.
26. Konieczny P, Stepniak-Konieczna E, Sobczak K. MBNL expression in autoregulatory feedback loops. RNA Biol. 2018 Jan 2;15(1):1–8.
27. Hwang JY, Jung S, Kook TL, Rouchka EC, Bok J, Park JW. rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. Nucleic Acids Res. 2020 Jul 2;48(W1):W300–6.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550.
29. Mori K, Lammich S, Mackenzie IRA, Forné I, Zilow S, Kretzschmar H, et al. 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). 2013 Mar;125(3):413–23.
30. Doostparast Torshizi A, Armoskus C, Zhang H, Forrest MP, Zhang S, Souaiaia T, et al. Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. Sci Adv. 2019 Sep 11;5(9):eaau4139.
31. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet. 2014 Dec;15(12):829–45.
32. Li M, Zhuang Y, Batra R, Thomas JD, Li M, Nutter CA, et al. HNRNPA1-induced spliceopathy in a transgenic mouse model of myotonic dystrophy. Proc Natl Acad Sci. 2020 Mar 10;117(10):5472–7.
33. Laurent FX, Sureau A, Klein AF, Trouslard F, Gasnier E, Furling D, et al. New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. Nucleic Acids Res. 2012 Apr;40(7):3159–71.
34. Pettersson OJ, Aagaard L, Andrejeva D, Thomsen R, Jensen TG, Damgaard CK. DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. Nucleic Acids Res. 2014 Jun;42(11):7186–200.

886 35. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-Seq data. Nat
887 Preced. 2012 Apr 4;1–1.
888 36. Sznajder ŁJ, Thomas JD, Carrell EM, Reid T, McFarland KN, Cleary JD, et al. Intron
889 retention induced by microsatellite expansions as a disease biomarker. Proc Natl Acad Sci
890 U S A. 2018 Apr 17;115(16):4234–9.
891 37. Chen W, Wang S, Tithi SS, Ellison DW, Schaid DJ, Wu G. A rare variant analysis
892 framework using public genotype summary counts to prioritize disease-predisposition
893 genes. Nat Commun. 2022 May 11;13(1):2592.
894 38. ProQR Therapeutics. Open-Label, Single-Dose, Exploratory Study With QR-504a to
895 Evaluate Safety, Tolerability, and Corneal Endothelium Molecular Biomarker(s) in Subjects
896 With Fuchs Endothelial Corneal Dystrophy With Trinucleotide Repeat Expansion in the
897 TCF4 Gene (FECD3) [Internet]. clinicaltrials.gov; 2022 May [cited 2023 Mar 6]. Report No.:
898 NCT05052554. Available from: https://clinicaltrials.gov/ct2/show/NCT05052554
899 39. Powers A, Rinkoski TA, Cheung K, Schehr H, Osgood N, Livelo C, et al. GeneTAC^{TM} small
900 molecules reduce toxic nuclear foci and restore normal splicing in corneal endothelial cells
901 derived from patients with Fuchs endothelial corneal dystrophy (FECD) harboring repeat
902 expansions in transcription factor 4 (TCF4). Invest Ophthalmol Vis Sci. 2022 Jun
903 1;63(7):2753-A0242.
904 40. Goyer B, Thériault M, Gendron SP, Brunette I, Rochette PJ, Proulx S. Extracellular Matrix
905 and Integrin Expression Profiles in Fuchs Endothelial Corneal Dystrophy Cells and Tissue
906 Model. Tissue Eng Part A. 2018 Apr;24(7–8):607–15.
907 41. Thériault M, Gendron SP, Brunette I, Rochette PJ, Proulx S. Function-Related Protein
908 Expression in Fuchs Endothelial Corneal Dystrophy Cells and Tissue Models. Am J Pathol.
909 2018 Jul 1;188(7):1703–12.
910 42. Angelbello AJ, Benhamou RI, Rzuczek SG, Choudhary S, Tang Z, Chen JL, et al. A Small
911 Molecule that Binds an RNA Repeat Expansion Stimulates Its Decay via the Exosome
912 Complex. Cell Chem Biol. 2021 Jan 21;28(1):34-45.e6.
913 43. Wang Q, Dou S, Zhang B, Jiang H, Qi X, Duan H, et al. Heterogeneity of human corneal
914 endothelium implicates lncRNA NEAT1 in Fuchs endothelial corneal dystrophy. Mol Ther -
915 Nucleic Acids. 2022 Mar 8;27:880–93.
916 44. Handsaker R, Kashin S, Reed N, Lee WS, McDonald TM, Tan S, et al. Somatic DNA repeat
917 expansion underlies Huntington's Disease neuropathology (Plenary Abstract). In
918 Washington DC; 2023.
919 45. Gall-Duncan T, Sato N, Yuen RKC, Pearson CE. Advancing genomic technologies and
920 clinical awareness accelerates discovery of disease-associated tandem repeat sequences.
921 Genome Res [Internet]. 2021 Dec 29 [cited 2022 Jan 28]; Available from:
922 https://genome.cshlp.org/content/early/2021/12/28/gr.269530.120
923 46. Sirp A, Roots K, Nurm K, Tuvikene J, Sepp M, Timmusk T. Functional consequences of
924 TCF4 missense substitutions associated with Pitt-Hopkins syndrome, mild intellectual
925 disability, and schizophrenia. J Biol Chem. 2021 Dec;297(6):101381.
926 47. Tekendo-Ngongang C, Grochowsky A, Solomon BD, Yano ST. Beyond Trinucleotide
927 Repeat Expansion in Fragile X Syndrome: Rare Coding and Noncoding Variants in FMR1
928 and Associated Phenotypes. Genes. 2021 Oct 22;12(11):1669.
929 48. Nath S, Caron NS, May L, Gluscencova OB, Kolesar J, Brady L, et al. Functional
930 characterization of variants of unknown significance in a spinocerebellar ataxia patient using
931 an unsupervised machine learning pipeline. Hum Genome Var. 2022 Apr 14;9(1):1–12.
932 49. Liu F, Liu Q, Lu CX, Cui B, Guo XN, Wang RR, et al. Identification of a novel loss-of-function
933 C9orf72 splice site mutation in a patient with amyotrophic lateral sclerosis. Neurobiol Aging.
934 2016 Nov 1;47:219.e1-219.e5.
935 50. Winkler NS, Milone M, Martinez-Thompson JM, Raja H, Aleff RA, Patel SV, et al. Fuchs'
936 Endothelial Corneal Dystrophy in Patients With Myotonic Dystrophy, Type 1. Invest

Ophthalmol Vis Sci. 2018 Jun 1;59(7):3053–7.
51. Mootha VV, Hansen B, Rong Z, Mammen PP, Zhou Z, Xing C, et al. Fuchs' Endothelial Corneal Dystrophy and RNA Foci in Patients With Myotonic Dystrophy. Invest Ophthalmol Vis Sci. 2017 Sep 1;58(11):4579–85.
52. Wieben ED, Aleff RA, Tosakulwong N, Butz ML, Highsmith WE, Edwards AO, et al. A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. PLOS ONE. 2012 Nov 21;7(11):e49083.
53. Vasanth S, Eghrari AO, Gapsis BC, Wang J, Haller NF, Stark WJ, et al. Expansion of CTG18.1 Trinucleotide Repeat in TCF4 Is a Potent Driver of Fuchs' Corneal Dystrophy. Invest Ophthalmol Vis Sci. 2015 Jul;56(8):4531–6.
54. Peh GSL, Chng Z, Ang HP, Cheng TYD, Adnan K, Seah XY, et al. Propagation of human corneal endothelial cells: A novel dual media approach. Cell Transplant. 2015;24(2):287–304.
55. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 2022 Oct 31]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.
57. Witten DM. Classification and clustering of sequencing data using a Poisson model. Ann Appl Stat. 2011 Dec;5(4):2493–518.
58. Blighe K. PCAtools: everything Principal Component Analysis [Internet]. 2023 [cited 2023 Feb 20]. Available from: https://github.com/kevinblighe/PCAtools
59. Ignatiadis N, Huber W. Covariate powered cross-weighted multiple testing. J R Stat Soc Ser B Stat Methodol. 2021;83(4):720–51.
60. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat Methods. 2016 Jul;13(7):577–80.
61. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019 Jun;35(12):2084–92.
62. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019 Jul 2;47(W1):W191–8.
63. Veiga DT. maseR: Mapping Alternative Splicing Events to pRoteins [Internet]. 2022 [cited 2023 Feb 20]. Available from: https://github.com/DiogoVeiga/maser
64. Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, et al. Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. PLOS ONE. 2015 Jul 15;10(7):e0132628.
65. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.
66. Tardaguila M, Fuente L de la, Marti C, Pereira C, Pardo-Palacios FJ, Risco H del, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018 Jan 3;28(3):396–411.
67. Pontikos N, Yu J, Moghul I, Withington L, Blanco-Kelly F, Vulliamy T, et al. Phenopolis: an open platform for harmonization and analysis of genetic and phenotypic data. Bioinformatics. 2017 Aug 1;33(15):2421–3.
68. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122.
69. Zhang D, Dey R, Lee S. Fast and robust ancestry prediction using principal component analysis. Bioinforma Oxf Engl. 2020 Jun 1;36(11):3439–46.
70. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for

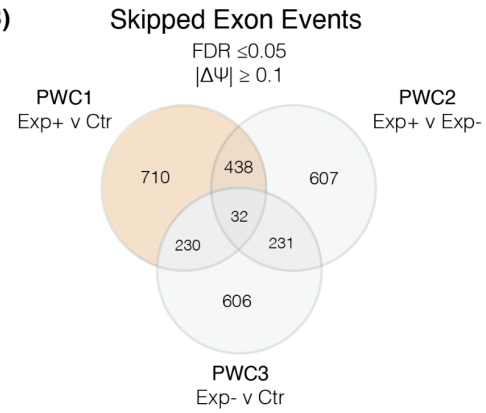
988 estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014
989 Mar;46(3):310–5.
990 71. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for
991 testing SNV novelty. Bioinformatics. 2011 Nov 15;27(22):3216–7.
992 72. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li
993 YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019 Jan
994 24;176(3):535-548.e24.
995 73. Zuallaert J, Godin F, Kim M, Soete A, Saeys Y, De Neve W. SpliceRover: interpretable
996 convolutional neural networks for improved splice site prediction. Bioinformatics. 2018 Dec
997 15;34(24):4180–8.

998 **Supplemental material:**
999

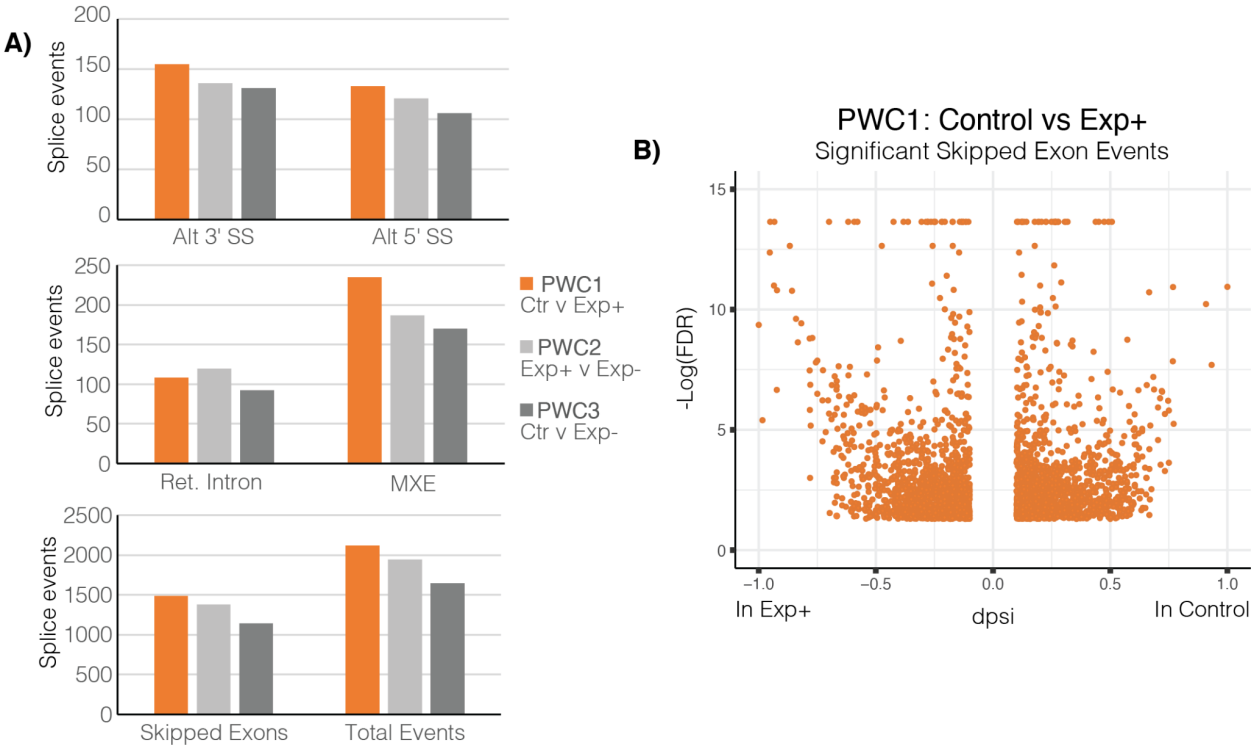
A)

	PWC1:Exp+ v Ctr	PWC2:Exp+ v Exp-	PWC3:Exp- v Ctr
Alternative 3' Splice Site	155 (7.32%)	136 (6.99%)	131 (6.74%)
Alternative 5' Splice Site	133 (6.28%)	121 (6.22%)	106 (5.45%)
Mutually Exclusive Exons	235 (11.10%)	187 (9.61%)	170 (8.74%)
Retained Introns	108 (5.10%)	120 (6.17%)	92 (4.73%)
Skipped Exons	1487 (70.21%)	1381 (71.00%)	1144 (58.82%)
Total Events	2118	1945	1643

B)



1000
1001 **Figure S1: Pairwise comparison of rMATS alternative splicing analysis demonstrates**
1002 **increased levels of alternative splicing in Exp+ FECD compared to both Exp- FECD and**
1003 **control primary corneal endothelial cells.** Short-read RNA-seq data generated from primary
1004 corneal endothelial cells were analysed by rMATS **A)** Table of rMATS events for each pairwise
1005 comparison (PWC). Significance denoted by FDR ≤ 0.05 and deltapsi magnitude larger than
1006 0.1. Values in brackets show percentage of total splice events **B)** Venn diagram of skipped exon
1007 event coordinates between all PWCs. The largest overlap is observed between PWC1 and
1008 PWC2 highlighting an enrichment of skipped exon events in Exp+ FECD compared to Exp-
1009 FECD and controls.
1010



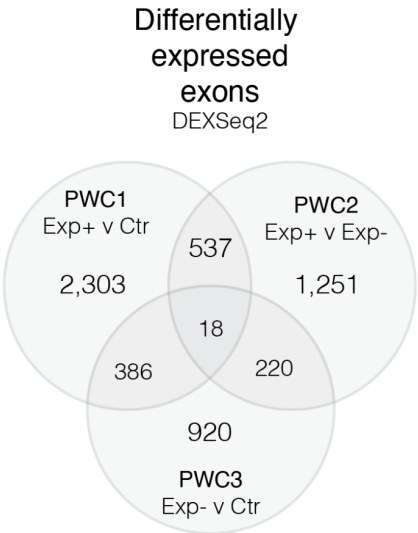
1011

1012 **Figure S2: Alternative splicing analyses of short-read CEC RNA-seq data demonstrate**
1013 **increased levels of alternative splicing in Exp+ FECD. A)** Summary of global differences in
1014 splicing events categorised by rMATS**.** N=4 for controls, N=3 each for FECD Exp+ and FECD
1015 Exp-. The most common splice category of alternative splicing observed between all pairwise
1016 comparisons was skipped exon, representing ~70% of all significant alternative splicing events
1017 detected. Alt 3' SS: alternative 3' splice site, Alt 5' SS: alternative 5' splice site, Ret. Intron:
1018 retained intron, MXE: mutually exclusive exon **B)** Volcano plot of statistically significant skipped
1019 exon events in PWC1 (Control vs Exp+). The dpsi value denotes the magnitude of change for
1020 each dysregulated skipped exon event identified. A positive dspi denotes decreased levels of
1021 exon inclusion in Exp+, whereas a negative dspi denotes increased levels of exon inclusion in
1022 Exp+.
1023

A)

	Number of exons	Number of genes
PWC1: Ctr v Exp+	3245	1560
PWC2: Exp+ v Exp-	2026	1176
PWC3: Exp- v Ctr	1544	893

B)



1024
1025 **Figure S3: Pairwise comparison of differentially expressed exons identified by DEXSeq2**
1026 **demonstrates increased levels of alternative splicing in Exp+ FECD compared to both**
1027 **Exp- FECD and control primary corneal endothelial cells. A)** Table of DEXSeq2 events
1028 and genes for each pairwise comparison (PWC). Significance denoted by p-adj ≤ 0.05. **B)** Venn
1029 diagram of dysregulated exons coordinates between all three PWCs. The largest overlap is
1030 observed between PWC1 and PWC2, highlighting an enrichment of skipped exon events in
1031 Exp+ FECD compared to Exp- FECD and controls.
1032
1033
1034
1035
1036
1037

1038 **Table S1: Summary of SQANTI3 analysis of IsoSeq data.** In depth characterization of
1039 isoforms in control and Exp+ long-read RNA-seq utilising SQANTI3 pipeline following stringent
1040 quality control.
1041

Characterization of transcripts based on splice junctions				
	# isoforms	# genes	# isoforms	# genes
Full Splice Match (FSM)	1,163	1,047	3,130	2,598
Incomplete Splice Match	647	535	2,268	1,377
Novel In Catalog (NIC)	829	685	1,797	1,336
Novel Not In Catalog (NNC)	437	375	778	639
Genic Genomic	3	3	5	5
Antisense	6	5	4	4
Fusion	24	21	35	33
Intergenic	4	4	6	6
Terminology based on Tardaguila et al 2018. n= 4 for controls and n=3 for FECD Exp+ (with one sample duplicated to acquire additional depth). FSM: full splice match. ISM: Incomplete splice match. NIC: Novel in catalog. NNC: Novel not in catalog……				

1042
1043 **Table S2: Summary of rMATs results.** The top five splice events detected for the following
1044 pairwise comparisons (PWC) are presented; **(PWC1)** control versus Exp+ FECD, **(PWC2)** Exp+
1045 FECD versus Exp- FECD, and **(PWC3)** control versus Exp- FECD.

1046 **Table S3. rMATS identified significant differentially spliced events in Exp+ matching**
1047 **published events with strong association to CTG18.1-expansion mediated FECD.**

Gene	Identified by rMATS (yes/no)	rMATS splice type	Differential splicing detected via SQANTI3	Skipped Exon Coordinates (hg38)	dpsi	FDR
ABI1[3,6]	Yes	SE	Yes	chr10:26771075 - 26771089	0.28	1.82E-07
ADD3[3,6]	No	-	Yes	-	-	-
AKAP13[3,6]	Yes	SE	Yes	chr15:85658537 - 85658590	0.203	0.003372
TSPOAP1[3,6]	Yes	SE	Excluded	chr17:58308541 - 58309380	0.222	0.028978
CD46[3,6]	No	-	Yes	-	-	-
CLASP1[3,6]	Yes	SE	Yes	chr2:121445449 - 121445496	0.451	0
COPZ2[3,6]	Yes	SE	No	chr17:48027676 - 48027789	0.118	7.49E-13
EXOC1†[3,6]	Yes	SE	Yes	chr4:55888888 - 55888932	-0.141	0.000691
FGFR1†[3,6]	Yes	SE	Yes	chr8:38429682 - 38429948	-0.248	0
GOLGA2[3,6]	Yes	SE	Yes	chr9:128272785 - 128272865	-0.188	1.10E-07
INF2†[3,6]	Yes	SE	Yes	chr14:104715284 -104715340	0.507	0
ITGA6†[3,6]	Yes	SE	No	chr2:172501772 - 172501901	0.263	0
KIF13A[a3,6]	Yes	SE	Excluded	chr6:17789872 - 17789910	0.17	0.00714
KIF13A[b3,6]	No	-	Excluded	-	-	-
*MBNL1**[3,5,6]	Yes	SE	Yes	chr3:152446704 - 152446757	-0.15	0.000337
*MBNL2**[3,5,6]	Yes	SE	Yes	chr13:97356796 - 97356849	-0.145	3.66E-13
MYO6[3,6]	No	-	Yes	-	-	-
NHSL1[3,6]	Yes	SE	Yes	chr6:138441983 - 138442114	-0.222	0.011795

*NUMA1**[3,5,6]	Yes	SE	Yes	chr11:72012401 - 72012442	0.268	0
PLEKHM2[3,6]	Yes	SE	Yes	chr1:15721329 - 15721388	0.234	1.27E-10
*PPFIBP1**	Yes	SE	Yes	chr12:27677064 - 27677096	0.162	7.22E-05
SCARB1[3,6]	No	-	Yes	-	-	-
*SYNE1**[3,6]	Yes	SE	Yes	chr6:152145487 - 152145555	0.223	4.36E-09
VEGFA[3,6]	No	-	Yes	-	-	-

* Events previously validated by RT-PCR, in addition to Mixture of Isoforms (MISO) and MAP-RSeq.[6–8]

† Events where the dysregulated exons match published work, but the flanking exons are not a perfect match.

SE: skipped exon. Samples labeled as excluded from Iso-seq analysis were excluded due to insufficient coverage during sequencing

1048

1049 **Table S4: Details on the pairwise comparison(s) where rMATS identified new significant**
1050 **differentially spliced events matching published differentially spliced genes with strong**
1051 **association to CTG18.1-expansion mediated FECD.**

Gene	rMATS splice type	Exon Coordinates (hg38)	FDR	dpsi	Verified in Iso-Seq
FGFR1	MXE	chr8:38429681-38429948, 38457355-38457534	0	0.125	Yes
MBNL1	SE	chr3:1522994040-152300367	1.13E-10	-0.101	Yes
KIF13A	SE	chr6:17789871-17789910	1.73E-09	0.339	Excluded
KIF13A	SE	chr6:17794248-17794395	1.63E-07	0.125	Excluded
AKAP13	SE	chr15:85662387-85662453	0.0012	0.126	Yes
MBNL2	SE	chr13:97356795-97356849	0.0492	-0.304	Yes
MBNL2	SE	chr13:97366458-97366553	0.0011	-0.114	Yes
NUMA1	A5SS	chr11:72068078-72068231 (long), 72068104-72068231 (short)	0.0441	0.113	No
NUMA1	SE	chr11:72049418-72049539	0.0175	-0.107	Yes
TSPOAP1	SE	chr17:58320108-58320129	0.0002	-0.362	Excluded

MXE: mutually exclusive exon event. SE: skipped exon event. A5SS: alternative 5' splice site event. Genes excluded from Iso-Seq analysis were done so due to insufficient coverage

1052
1053

1054 **Table S5: Differential gene expression and alternative splicing of fibronectin (*FN1*) in all**
1055 **three pairwise comparisons.**

Differential gene expression (DESeq2)			
	shrunkLFC	*padj*	
PWC1	2.924967657	9.18E-24	
PWC2	n.s	n.s	
PWC3	2.137134317	2.63E-12	
Alternative splicing (rMATS)			
	Dysregulated exons	*dPSI*	*FDR*
PWC1	Exon 25 (EDB) Exon 33 (EDA)	-0.253, -0.305	0, 0
PWC2	Exon 25 (EDB) Exon 33 (EDA)	0.262, 0.179	0, 4.76E-05
PWC3	Exon 25 (EDB)	-0.126	0.007044931
Exon 25 (EDB) is defined by the genomic coordinates chr2:215,392,931-215,393,203 and Exon 33 (EDA) by chr2:215,380,811-215,381,080. Padj = FDR-adjusted p-value, dPSI=delta percent spliced in, FDR = False discovery rate, n.s. = not significant			

1056
1057

1058 **Table S6: rMAPS-identified RNAbinding motif enrichment in PWC1 and PWC2, which was**
1059 **also absent in PWC3.**

GC-containing motifs upregulating skipped exon in Exp+	
PCBP2	CC[CT][CT]CC[ACT}
RBM4	GCGCG[GC][GC]
RBM4	GCGCG[GC]G
RBM45	GACGA[AC][ACG]
TT-containing motifs downregulating skipped exon events in Exp+	
HNRNPC	[ACT]TTTTT[GT]
HNRNPCL1	[ACT]TTTTT[GT]
PCBP1	C[CT]TTCC
ZC3H14	TTT[AGT]TTT

1060
1061 **Table S7: Differentially expressed genes for all three pairwise comparisons including list**
1062 **of dysregulated genes that are not common to FECD in general.**
1063
1064

1065
1066 **Table S8: Dysregulated RNA binding proteins uniquely in PWC1.**

Downregulated uniquely in Exp+ (n=59)
*PPARGC1B, ZNF385A, **ESRP2**, IFIH1, NSUN7, C2orf15, IFIT1, **DDX25**, CELF5, ENDOU, NPM2, AFF3, **DDX60**, IFIT3, OAS3, PAIP2B, OASL, ERN2, PLD6, DDX60L, VARS2, DHX58, CSDC2, CPEB3, IFIT2, NYNRIN, PTGES3L-AARSD1, PIH1D2, DIS3L, ZNF106, MOV10, RAVER2, PNPT1, PDCD4, SCAF11, PAN2, MIF4GD, ZNF473, PARP1, TDRD7, ANGEL1, TRMT44, RRNAD1, PRPF40B, WARS2, TRIM56, TST, MRPL34, MRPS9, ZC3H7A, ENDOV, TRPT1, ADAT1, SKIV2L, HEXIM1, MRPL49, REPIN1, EIF2D, RALYL*
Upregulated uniquely in Exp+ (n=148)
*SRRT, UBAP2L, STRAP, **DDX56**, CCNT2, ZC3H15, TSR3, NOL7, **DDX1**, ZC3H18, SYMPK, PTGES3, NCL, RRP1, PES1, ZC3H14, SCAF4, TNPO3, MTPAP, R3HDM1, SRP54, CNBP, HDLBP, **CELF1**, RUVBL1, SSB, KPNB1, EIF4G2, DHX15, MRPL52, NONO, CMTR2, MTG1, DHX36, NXF1, SMG1, SRP72, **HNRNPDL**, SF3B4, EIF3B, EEF1E1, CAPRIN1, RBMS1, PSMA6, LSM7, FIP1L1, SRSF10, DGCR8, AFF4, LARP1B, IPO4, DENR, LRRFIP1, RP9, WDR3, MVP, DUSP11, NOP16, U2AF2, DYNLL1, **MBNL1**, SRPK2, RRP12, PIN4, UTP6, ABCE1, GRSF1, IMP4, EIF4A1, PNO1, PDCD11, RRBP1, CPSF6, RRS1, RPS24, RBM18, SFPQ, CCAR1, UTP14A, RBM12, DDX52, SUB1, **HNRNPAB**, FARSA, MSL3, TRMT61A, NARS2, EIF4G1, TRA2B, NOP14, RPF1, ZC3H12C, NOL10, RANBP6, EIF6, RTCA, BOP1, SERBP1, WDR4, ZC3H13, NOL6, AIMP2, XPO6, FARSB, **HNRNPM**, DDX21, STAU2, IPO5, RRP9, UTP20, LRRFIP2, RBM14, MAK16, ZNF326, DCAF13, EIF3C, TOP1, EIF5A, SYNCRIP, PELO, AGFG1, CTU2, PRDX1, ISG20L2, TPR, NOC2L, NOP10, CCDC86, RBM3, CALR, POP1, ZNF579, AGO2, PARP4, HINT3, DDX49, BZW1, SPATS2, YRDC, BZW2, ELAVL2, ISG20, DYNC1H1, RBFOX1, RBPMS, PCBP3, ENOX1, IGF2BP3*

1067
1068

Table S9: Pathway enrichment (GO, KEGG, and Reactome) for PWC1, PWC3, and dysregulated genes unique to Exp+.

Table S10: DEXSeq results for *TCF4* in all three pairwise comparisons showing dysregulation of exons within the gene.

Table S11: rMATS *TCF4* mutually exclusive exon events demonstrating a shift between longer isoforms containing 2+ AD domains and shorter isoforms containing 1-2 AD domains detected via alternative splicing pipelines.

	NM_001083962.2/ENST00000354452.8 exon numbering			
Splice type event (PWC1)	**Exon 7**	**Exon 6**	**dpsi**	**FDR**
Mutually exclusive exon	chr18:55350873-55351003	chr18:55403453-55403518	0.103	0.0023
Mutually exclusive exon	chr18:55350873-55351000	chr18:55403453-55403518	0.102	0.0030
Consequence:	**Higher in Exp+**	**Higher in Control**		
	First exon common to both long and short isoforms	**Last exon common to the majority of long isoforms**		

1083 **Table S12**: ***TCF4* isoform RNAScope sample summary with results of FISH with probe**
1084 **targeting repeat and negative RNAScope control experiments (with probes targeting**
1085 **bacterial genes).**

Sample	CTG18.1 repeat genotype	CUG specific foci present	Negative control clear?	Probe B proportion (%)
Unaffected Controls				
Control 8	12/16	No	Yes	29.43
Control 9	15/33	No	Yes	28.08
Control 10	25/26	No	Yes	40.55
			Average	36.02 ± 6.9
CTG18.1 expansion negative FECD				
FECD4exp-	15/15	No	Yes	27.67
FECD5exp-	16/23	No	Yes	28.18
FECD6exp-	15/16	No	Yes	36.86
			Average	30.90 ± 5.17
CTG18.1 expansion positive FECD				
FECD8exp+	63/88	Yes	Yes	15.05
FECD9exp+	12/76	Yes	Yes	12.58
FECD10exp+	18/67	Yes	Yes	21.36
			Average	16.33 ± 4.53

1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097

1098 **Table S13:** Summary of rare, potentially deleterious, variants in FECD-associated genes
1099 identified in Proband B.
1100

Variant (hg38)	Gene	HGVSc, HGVSp	CADD, MAF gnomAD, MAF Kaviar	Proband
chr20-3231199-C-T	*SLC4A11*	Het c.1040G>A, p.(Arg347Gln)	19.24, 1.39×10^{-3}, 5.2×10^{-4}	Proband B 0/1
Exome data from all probands with *TCF4* rare variants were interrogated to determine if rare and potentially deleterious variants in previously reported FECD-associated genes, including *COL8A2*, *SLC4A11*, *AGBL1*, *and ZEB1* could explain their respective disease (MAF < 0.005 in publicly available gnomAD genomes, exomes and Kaviar, CADD score > 15. A single heterozygous *SLC4A11* variant of uncertain significance was identified in Proband B. The annotation from Variant Effect Predictor is filtered to the "most severe" transcripts which is ENST00000380056.3. 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.				

1101
1102
1103

1104 **Table S14: Clinical data of probands with Fuchs endothelial corneal dystrophy identified**
1105 **to have rare and potentially deleterious heterozygous *TCF4* variants.**

Age* Sex Ethnicity	CTG18.1 repeat genotype	Age at keratoplasty (years)***	CCT (µm) before surgery	Pre-op BCVA	CCT (µm) after surgery	Final BCVA**	Associated ocular features	Family history (1º relatives)
57/F/black Proband A	10/12	57 OD 59 OS	495 OD 493 OS	NA	441 OD 431 OS (DMEK)	0.67 OD 0.67 OS	Cataract surgery	Unknown
42/M/black Proband B	17/23	NP	589 OD 577 OS	NP	NP	0.67 OD 0.67 OS	Nil	Unknown
61/F/mixed race (white/black) Proband C	13/18	61 OD 63 OS	NA 691 OS	NA	681 OD (DSAEK) 539 OS (DMEK)	1.0OD 0.67 OS	Cataract surgery	Unknown
74/F/white Proband D	12/15	75 OD 74 OS	NA	NA	442 OD 464 OS (DMEK)	1.0 OD 1.0 OS	Cataract surgery	Yes
60/F/white Proband E	12/16	71 OD 60 OS (PK)	605 OD NA OS	0.2 OD 0.25 OS	NA	0.4 OD 0.6 OS	Cataract surgery, secondary glaucoma	No
64/F/white Proband F	12/18	73 OD 72 OS	640 OD 535 OS	0.5 OD 0.5 OS	572 OD 527 OS (DMEK)	0.67 OD 0.67 OS	Cataract surgery	Yes
56/F/white Proband G	12/18	59 OD (PK) 63 OS (DSAEK) 71 OS	NA	NA	532 OD 539 OS	0.5 OD 0.8 OS	EBMD, cataract surgery, secondary glaucoma	Yes

1106 M male, F female, OD right eye, OS left eye, DMEK Descemet membrane endothelial
1107 keratoplasty, DSAEK Descemet's stripping automated endothelial keratoplasty, PK penetrating
1108 keratoplasty, CCT central corneal thickness, BCVA best corrected visual acuity, NP corneal
1109 surgery not performed, NA not available, EBMD epithelial basement membrane degeneration
1110 (Cogan),*Age at diagnosis, **BCVA values converted to decimal annotation, ***DMEK unless
1111 stated.
1112
1113
1114
1115