Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease

Yongjun Chu[1,2], Jiaxin Hu[1,2], Hanquan Liang[3], Mohammed Kanchwala[3], Chao Xing[3,4,5], Walter Beebe[6], C. Bradley Bowman[6], Xin Gong[7], David R. Corey [1,2,*] and V. Vinod Mootha[3,7,*]

[1]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA, [2]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA, [3]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA, [4]Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA, [5]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA, [6]Cornea Associates of Texas, Dallas, TX, USA and [7]Department of Ophthalmology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9057, USA

ABSTRACT

How genetic defects trigger the molecular changes that cause late-onset disease is important for understanding disease progression and therapeutic development. Fuchs' endothelial corneal dystrophy (FECD) is an RNA-mediated disease caused by a trinucleotide CTG expansion in an intron within the *TCF4* gene. The mutant intronic CUG RNA is present at one–two copies per cell, posing a challenge to understand how a rare RNA can cause disease. Late-onset FECD is a uniquely advantageous model for studying how RNA triggers disease because: (i) Affected tissue is routinely removed during surgery; (ii) The expanded CTG mutation is one of the most prevalent disease-causing mutations, making it possible to obtain pre-symptomatic tissue from eye bank donors to probe how gene expression changes precede disease; and (iii) The affected tissue is a homogeneous single cell monolayer, facilitating accurate transcriptome analysis. Here, we use RNA sequencing (RNAseq) to compare tissue from individuals who are pre-symptomatic (Pre_S) to tissue from patients with late stage FECD (FECD_REP). The abundance of mutant repeat intronic RNA in Pre_S and FECD_REP tissue is elevated due to increased half-life in a corneal cells. In Pre_S tissue, changes in splicing and extracellular matrix gene expression foreshadow the changes observed in advanced disease and predict the activation of the fibrosis pathway and immune system seen in late-stage patients. The absolute magnitude of splicing changes is similar in pre-symptomatic and late stage tissue. Our data identify gene candidates for early drivers of disease and biomarkers that may represent diagnostic and therapeutic targets for FECD. We conclude that changes in alternative splicing and gene expression are observable decades prior to the diagnosis of late-onset trinucleotide repeat disease.

INTRODUCTION

An inherent paradox of inherited late-onset degenerative disease is the fact that the genetic basis for disease exists years before symptoms arise. At some point the genetic defect begins to trigger a cascade of cellular changes that result in disease, but how is the cascade initiated? Unfortunately, the search for early drivers of many genetic diseases that involve mutant RNA is hindered by lack of accessible pre-symptomatic tissue to identify early changes in gene expression that occur years prior to diagnosis. Corneal tissue provides an accessible source of both pre- and post-symptomatic tissue to explore the molecular origins of inherited late-onset disease.

Corneal diseases represent one of the leading causes of vision loss and blindness globally ([1]). Inherited corneal dystrophies can compromise the structure and transparency of the cornea. Late-onset Fuchs' endothelial corneal dystrophy (FECD) is one of the most common genetic disorders,

---

*To whom correspondence should be addressed. Tel: +1 214 645 6155; Email: david.corey@utsouthwestern.edu
Correspondence may also be addressed to V. Vinod Mootha. Tel: +1 214 648 2367; Fax: +1 214 648 2469; Email: vinod.mootha@utsouthwestern.edu

affecting four percent of the population in the US over the age of 40 (2–4). The corneal endothelium is the inner hexagonal monolayer responsible for maintenance of stromal dehydration and corneal clarity.

In FECD, the post-mitotic endothelium undergoes premature senescence and apoptosis (5–11). Descemet's membrane, the basement membrane of the endothelium, becomes diffusely thickened and develops focal excrescences called guttae. Guttae are clinically diagnostic of FECD by slit-lamp biomicroscopy (12). Confluence of central guttae and concomitant loss of endothelial cell density results in corneal edema, scarring, and loss of vision, making FECD the leading indication for corneal transplantation in the US and developed world (13,14).

Two-thirds of FECD cases are caused by an expansion of the trinucleotide CTG within the *TCF4* gene (15–18) making the corneal dystrophy the most common human disorder mediated by simple DNA repeats. FECD can also be caused by a CTG expansion within the 3′-untranslated region (3′-UTR) of the *DMPK* gene (19–21), implicating mutant expanded CUG RNA as the root cause for repeat-associated FECD (FECD_REP). The remaining one third of patients lack the expanded repeat (FECD_NR) but the two types of FECD are indistinguishable during normal clinical observation.

FECD can be treated with corneal transplantation and surgical outcomes have improved following the development of endothelial keratoplasty in which a single-cell endothelium monolayer is transplanted using donor corneal tissue (22). While effective, this approach is limited by the availability of donor corneas and access to suitable treatment centers. A deeper understanding of FECD might permit the development of pharmacological approaches that might reach a larger patient population and reduce or delay the need for surgical intervention. Successful drug development would benefit from a more detailed understanding of disease progression and the fundamental drivers of early disease.

Trinucleotide or hexanucleotide repeats are the cause for many delayed-onset neurodegenerative disorders such as myotonic dystrophy type 1 and *C9orf72* amyotrophic lateral sclerosis/frontal temporal dementia (ALS/FTD) (23). FECD has unique advantages as a model for gaining insight into the molecular mechanisms of delayed-onset degenerative disease: (i) The tissue is more readily available for analysis because it is routinely removed during surgery (Figure 1A–C); (ii) Samples are a single near homogeneous layer of cells, facilitating RNAseq and other methods for analyzing gene expression (Figure 1D); (iii) The corneal endothelium and disease progression can be evaluated visually during a standard clinical examination; (iv) Like peripheral neurons, the corneal endothelium originates from neural crest and may serve as a model post-mitotic tissue to examine the effects of toxic repeat RNA over time; (v) Donor corneal tissues that are not used for surgery are available from eye bank samples for comparison; (vi) Because the prevalence of the triplet repeat mutation within the *TCF4* gene is relatively high, donor eyes provide a significant number of pre-symptomatic samples (Pre_S) that possess the CTG expanded repeat; (vii) The availability of tissue from four different cohorts (Figure 2), control, Pre_S, FECD_NR,

FECD_REP allows multiple cross-comparisons into the different stages and types of FECD. Analysis of the Pre_S cohort has the potential to gain insights into early drivers of disease.

Here, we use the advantages gained from analyzing the corneal endothelial monolayer to evaluate transcriptomic data from four cohorts of corneal endothelial tissue: control, Pre_S, FECD_REP and FECD_NR. We identify extensive and large magnitude changes in RNA splicing that are shared between Pre_S and FECD-REP cohorts but are not observed in FECD_NR samples. Levels of gene expression are changed in Pre_S samples relative to controls and suggest early triggering of the fibrosis pathway prior to clinical observation of disease symptoms. In late stage disease, pathways related to fibrosis and activation of immune system are shared by FECD_REP and FECD_NR but mitochondrial dysfunction is more pronounced in the FECD_REP cohort. These results lay a basis for understanding the onset of FECD and other trinucleotide repeat diseases and provide potential targets for therapeutic design.

MATERIALS AND METHODS

Isolation of corneal tissue

The study was conducted in compliance with the tenets of the Declaration of Helsinki and with the approval of the institutional review board of the University of Texas Southwestern Medical Center (UTSW). Subjects underwent a complete eye examination including slit lamp biomicroscopy by a cornea fellowship-trained ophthalmologist. Patients underwent endothelial keratoplasty for FECD severity Krachmer grade 5 ($\geq$5 mm central confluent guttae without stromal edema) or 6 ($\geq$5 mm central confluent guttae with stromal edema) assessed by slit lamp microscopy (24). After surgery, surgically explanted endothelium-Descemet's membrane monolayers were preserved in Optisol GS corneal storage media (Bausch & Lomb, Rochester, NY, USA) prior to storage at -80 Celsius. Genomic DNA was extracted from peripheral blood leukocytes of each study subject using Autogen Flexigene (Qiagen, Valencia, CA, USA).

Corneal endothelial samples from post-mortem donor corneas preserved in Optisol GS corneal storage media (Bausch & Lomb, Rochester, NY, USA) were obtained from the eye bank of Transplant Services at UT Southwestern. Certified eye bank technicians screened the donor corneal endothelium with slit lamp biomicroscopy and Cellchek EB-10 specular microscopy (Konan Medical). Endothelium-Descemet's membrane monolayers from donor corneas were micro-dissected and stored as previously described (25). DNA from the remaining corneal tissue of each sample was extracted with TRIzol reagent (ThermoScientific).

TCF4 CTG18.1 polymorphism genotyping

Genomic DNA from subjects' peripheral leukocytes or corneal tissue was used for genotyping. The CTG18.1 trinucleotide repeat polymorphism in the *TCF4* gene was genotyped using a combination of short tandem repeat (STR)
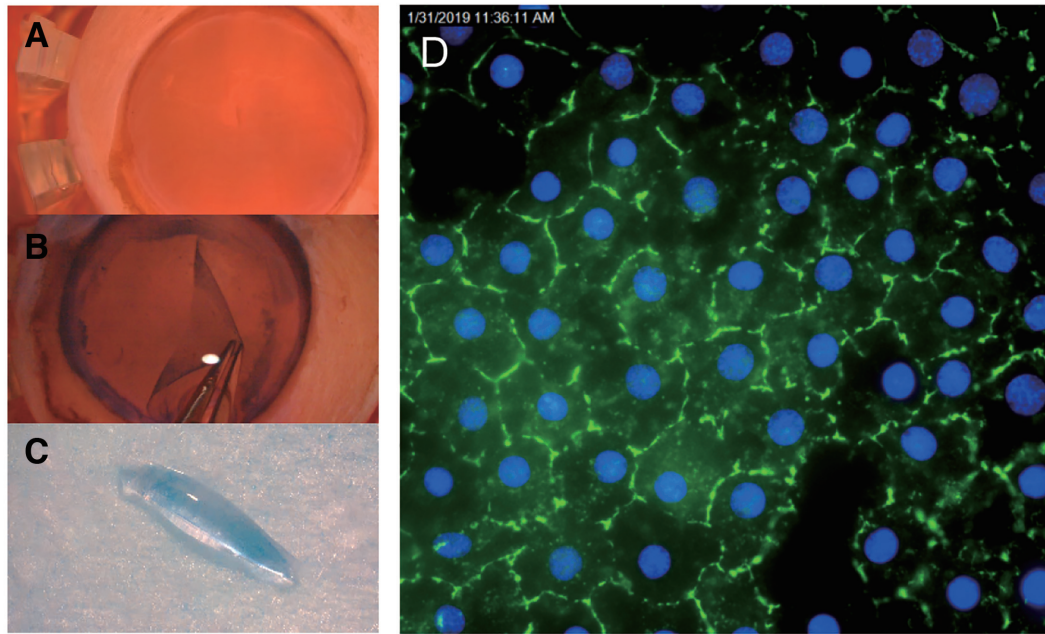
Figure 1. Preparation of homogeneous tissue monolayers for analysis. (**A**) Human donor cornea in corneal viewing chamber with Optisol corneal storage media (Bausch & Lomb). (**B**) Corneal endothelium/Descemet's membrane monolayer being dissected from underlying stromal tissue. (**C**) Monolayer of corneal endothelial cells assumes a 'scroll' shape. This single cell monolayer will be used for sequencing and other analyses. (**D**) Immunostaining of monolayer of cells with corneal endothelial-specific marker, zonula occudens-1 (ZO-1) (Blue-DAPI).

and triplet repeat primed polymerase chain reaction (TP-PCR) assays as we have previously described (20). For the STR assay, a pair of primers flanking the CTG18.1 locus was utilized for PCR amplification with one primer labeled with FAM on 5′ end. The TP-PCR assay was performed using the 5′ FAM-labeled primer specific for the repeat locus paired with repeat sequence targeted primers for PCR amplification. PCR amplicons were loaded on an ABI 3730XL DNA analyzer (Applied Biosystems, Foster City, CA, USA) and the results analyzed with ABI GeneMapper 4.0 (Applied Biosystems). Large triplet repeat expansions were sized by Southern blot analysis using digoxigenin-labeled probes.

RNA isolation and sequencing

Total RNA was isolated from each of 25 tissue samples (6 FECD_REP, 4 FECD_NR, 6 Pre_S and 9 Controls) by homogenization in QIAzol lysis reagent, chloroform extraction and isolation with NucleoSpin RNA XS (Macherey-Nagel GmbH & Co., Germany). RNA quantity and quality were determined by Bioanalyzer 2100. RNA libraries were prepared for each tissue sample with high RIN (> 5.0), using the TruSeq RNA sample Prep kit version 2 (Illumina, San Diego, CA, USA). For TruSeq stranded total RNAseq, ribosomal transcripts were depleted from total RNA, using Ribo-Zero Gold RNA removal kit followed by replacement of deoxythymidine triphosphate (dTTP) with deoxyuridine triphosphate (dUTP) during reverse transcription in the second strand synthesis, using TruSeq stranded total library preparation kit. The resulting libraries were minimally amplified to enrich for fragments using adapters on both ends and then quantified for sequencing at eight samples/flow
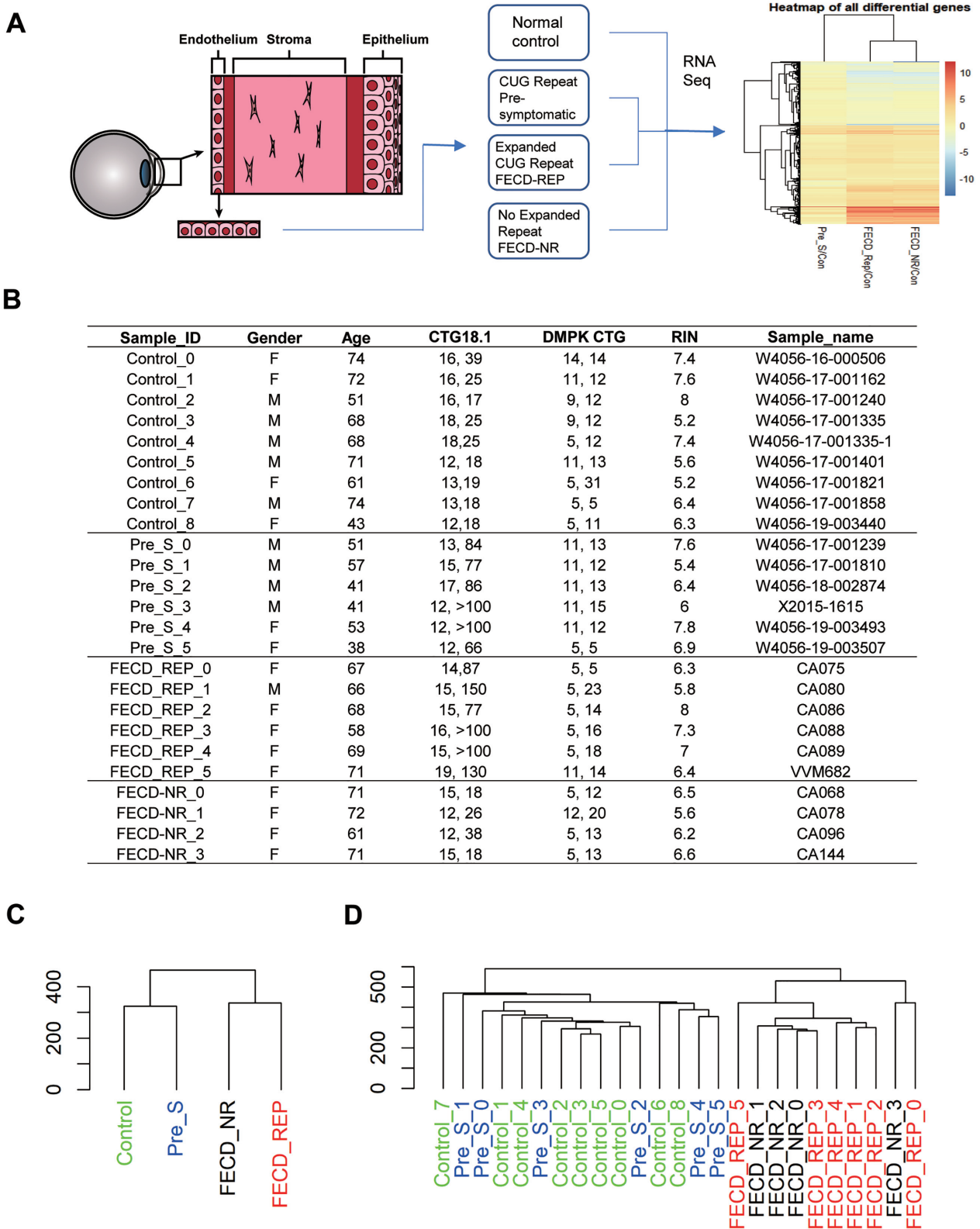
cell by using a NextSeq 500/550 (Illumina) sequencer (PE 150).

Analysis of differentially expressed or spliced genes

Whole transcriptomic sequencing data from each tissue sample was analyzed using an analysis pipeline that includes STAR for initial mapping and Cufflinks (v2.21) for gene and isoform differential analysis, among other publicly available programs. For gene/isoform differential analysis, the minimum expression level of 1.5 FPKM and a false discovery rate (FDR) < 0.05 were chosen as the threshold. The meta gene pathway analysis was carried out with IPA (Qiagen). The binary alignment map files from STAR were analyzed using rMATS (v.4.0) that quantitates the expression level of alternatively spliced genes between groups. To find the most significant events, we used stringent filtering criteria within rMATS to perform pairwise comparisons among four groups: percentage of spliced in (PSI) changes > 0.15; FDR < 0.001. For PSI, rMATS calculates a value for every differential splicing event, providing a range from 0 to 1, with 0 being completely excluded and 1 being uniformly included in the splicing products. Alternative splicing events were also compared to those obtained in tibialis anterior muscle of myotonic dystrophy type one patients (DM1). DM1 raw data were obtained from Gene Expression Omnibus (GSE86356, six Control and six DM1 tissue samples were used) and analyzed similarly as the FECD data (visit DMseq.org for more information).

Measurement of TCF4 RNA half-life

FECD corneal endothelial cell lines (F35T and F45SV) and a control corneal endothelial cells (W4056–17-001579) were

Sample_ID	Gender	Age	CTG18.1	DMPK CTG	RIN	Sample_name
Control_0	F	74	16, 39	14, 14	7.4	W4056-16-000506
Control_1	F	72	16, 25	11, 12	7.6	W4056-17-001162
Control_2	M	51	16, 17	9, 12	8	W4056-17-001240
Control_3	M	68	18, 25	9, 12	5.2	W4056-17-001335
Control_4	M	68	18,25	5, 12	7.4	W4056-17-001335-1
Control_5	M	71	12, 18	11, 13	5.6	W4056-17-001401
Control_6	F	61	13,19	5, 31	5.2	W4056-17-001821
Control_7	M	74	13,18	5, 5	6.4	W4056-17-001858
Control_8	F	43	12,18	5, 11	6.3	W4056-19-003440
Pre_S_0	M	51	13, 84	11, 13	7.6	W4056-17-001239
Pre_S_1	M	57	15, 77	11, 12	5.4	W4056-17-001810
Pre_S_2	M	41	17, 86	11, 13	6.4	W4056-18-002874
Pre_S_3	M	41	12, >100	11, 15	6	X2015-1615
Pre_S_4	F	53	12, >100	11, 12	7.8	W4056-19-003493
Pre_S_5	F	38	12, 66	5, 5	6.9	W4056-19-003507
FECD_REP_0	F	67	14,87	5, 5	6.3	CA075
FECD_REP_1	M	66	15, 150	5, 23	5.8	CA080
FECD_REP_2	F	68	15, 77	5, 14	8	CA086
FECD_REP_3	F	58	16, >100	5, 16	7.3	CA088
FECD_REP_4	F	69	15, >100	5, 18	7	CA089
FECD_REP_5	F	71	19, 130	11, 14	6.4	VVM682
FECD-NR_0	F	71	15, 18	5, 12	6.5	CA068
FECD-NR_1	F	72	12, 26	12, 20	5.6	CA078
FECD-NR_2	F	61	12, 38	5, 13	6.2	CA096
FECD-NR_3	F	71	15, 18	5, 13	6.6	CA144

Figure 2. Experimental scheme and sample cohorts. (**A**) Homogenous samples of corneal endothelium are surgically removed from normal, presymptomatic and, affected (FECD or FECD_NR) tissue bank donors or patients and used for RNAseq followed by data analysis. (**B**) Detailed description of tissue samples. (**C**) Dendrogram for four groups of samples and (**D**) all replicates. RIN:RNA integrity number. CTG18.1: the number of repeats on each chromosome.

cultured as described (26) The C9 and VVM84 skin fibroblasts were maintained at 37°C and 5% CO$_2$ in Minimal Essential Media Eagle (MEM) (Sigma, M4655) supplemented with 15% heat inactivated fetal bovine serum (Sigma) and 0.5% MEM nonessential amino acids (Sigma).

Endothelial or fibroblast cells were seeded in 6-well plate at 90% confluence. At the next day, actinomycin D was added into the wells at 5 μg/ml final concentrations. Cells were harvest using TRIzol agent (Sigma) at different time point. The *TCF4* RNA levels were analyzed by qPCR.

Validation of differential gene expression and alternative splicing patterns

Total RNA was extracted from control, Pre_S or FECD corneal endothelial tissues (Supplementary Table S1) by NucleoSpin RNA XS kit (Macherey-Nagel). cDNAs were prepared by reverse transcription. RT-PCR was performed using ChoiceTaq Blue Mastermix (Denville). PCR amplification was as follow: 94°C for 3 min (1 cycle), 94°C denaturation for 30 s, 60°C annealing for 30 s and 72°C extension for 1 min (38 cycles), and a 7 min 72°C extension. The PCR primers were used as reported (27) (Supplementary Table S2). The amplification products were separated by 1.5% agarose gel electrophoresis. qPCR experiments were performed on a 7500 real-time PCR system (Applied Biosystems) using iTaq SYBR Green Supermix (Bio-Rad). Data was normalized relative to levels of *RPL19* mRNA.

RESULTS

Experimental design

Samples were homogeneous endothelial cell monolayers surgically removed from patients or dissected from donor tissue (Figures 1 and 2A). The homogenous nature of the tissue samples, in contrast to the more complex mixtures of cells often contained in tissues related to other disease, facilitates subsequent analysis and data interpretation.

We compared four groups of human tissue by RNA sequencing (RNAseq) with attention to changes in alternative splicing, differential gene expression and pathway analysis (Figure 2A). Tissue that was mutant for the CTG expanded repeat within intron 2 of the *TCF4* gene was obtained from pre-symptomatic eye bank donors (Pre_S) and from FECD patients after transplant surgery (FECD_REP). Tissue that was negative for the expanded CTG mutation within *TCF4* was obtained from eye bank donors (Control) and from patients with non-CTG related FECD (FECD_NR). Genotype data indicated that all samples had normal numbers of CTG repeats within the *DMPK* gene (Figure 2B).

Control tissues from the eye bank were chosen for analysis that possessed normal endothelial morphology by specular microscopy, were negative for the expanded CTG mutation and were from donors with ages comparable to our FECD patient cohort (Figure 2B). Obtaining tissue from Pre_S individuals was possible because of the relatively high prevalence, 3%, of the expanded triplet repeat mutation in *TCF4* gene within the general Caucasian population (2,3). Pre_S tissue was identified by the presence of the CTG repeat expansion by genotyping. Normal corneal endothelial morphology (absence of central corneal guttae) in Pre_S tissues was confirmed by specular microscopy.

RNA sequencing (RNAseq) was performed on tissue samples from nine non-FECD/non-expanded repeat donors (Control), six pre-symptomatic donors with the expanded repeat (Pre_S), six patients with the expanded repeat and FECD (FECD_REP) and four patients with FECD who did not have the expanded CTG repeat mutation (FECD_NR) (Figure 2B) (Supplementary Figures S1–3). Only samples with RNA integrity numbers greater than five were used (Supplementary Figure S3). Gene body coverage analysis suggests that all RNA samples are largely intact (Supplementary Figure S2). We carried out a paired-end 150 nt RNA-Seq on Illumina NextSeq sequencer. We regularly obtained an output of 50–60 million raw paired reads per sample. Obtaining adequate sequencing depth required using all or most of each sample. Because of the rarity of Pre_S samples, a fraction of tissue was withheld from sequence for eventual experimental validation of gene expression changes suggested by RNAseq.

Clustering analysis of overall gene expression patterns revealed that samples from control and Pre_S donors were closer to one another than to samples from patients with FECD_REP or FECD_NR late stage disease (Figure 2C and D; Supplementary Figure S1E and F). This result is consistent with the clinical observation of large visible differences between diseased and non-diseased tissue and that the clinical manifestations of FECD_REP and FECD_NR are almost identical.

Stability of TCF4 intron 2 in FECD mutant and non-mutant tissue

Analysis of RNAseq data from intron 2 (the intron that contains the CTG repeat) within the *TCF4* gene showed a striking difference between representative samples from individuals with the expanded CTG repeat mutation (FECD_REP and Pre_S) and individuals who lack the mutation (control and FECD_NR) (Figure 3A). Samples from the cohort of control individuals showed similar low numbers of reads upstream or downstream relative to the trinucleotide repeat region. RNA obtained from FECD_NR patients' tissue also showed the same similarity between upstream and downstream reads.

By contrast, both Pre_S and FECD_REP tissue showed more reads for intronic RNA upstream of the trinucleotide repeat relative to downstream (Figure 3A). The overall expression of mature *TCF4* mRNA from the four cohorts was not significantly different making haploinsufficiency of gene product less likely as the mechanism of disease (Supplementary Figure S4). These results suggest that an early molecular disease trigger—increased stability the mutant *TCF4* intron 2 upstream of the expanded repeat—occurs at the presymptomatic stage and distinguishes FECD_REP from FECD_NR in late stage disease.

To investigate factors that might contribute to the prevalence of upstream intronic reads, we examined cultured cells derived from FECD_REP patient corneal endothelium (F35T), FECD_REP patient skin fibroblasts (VVM84), control (without CUG expanded repeat) corneal endothelium (W4056) or control skin fibroblasts (C9) (Figure 3BC).
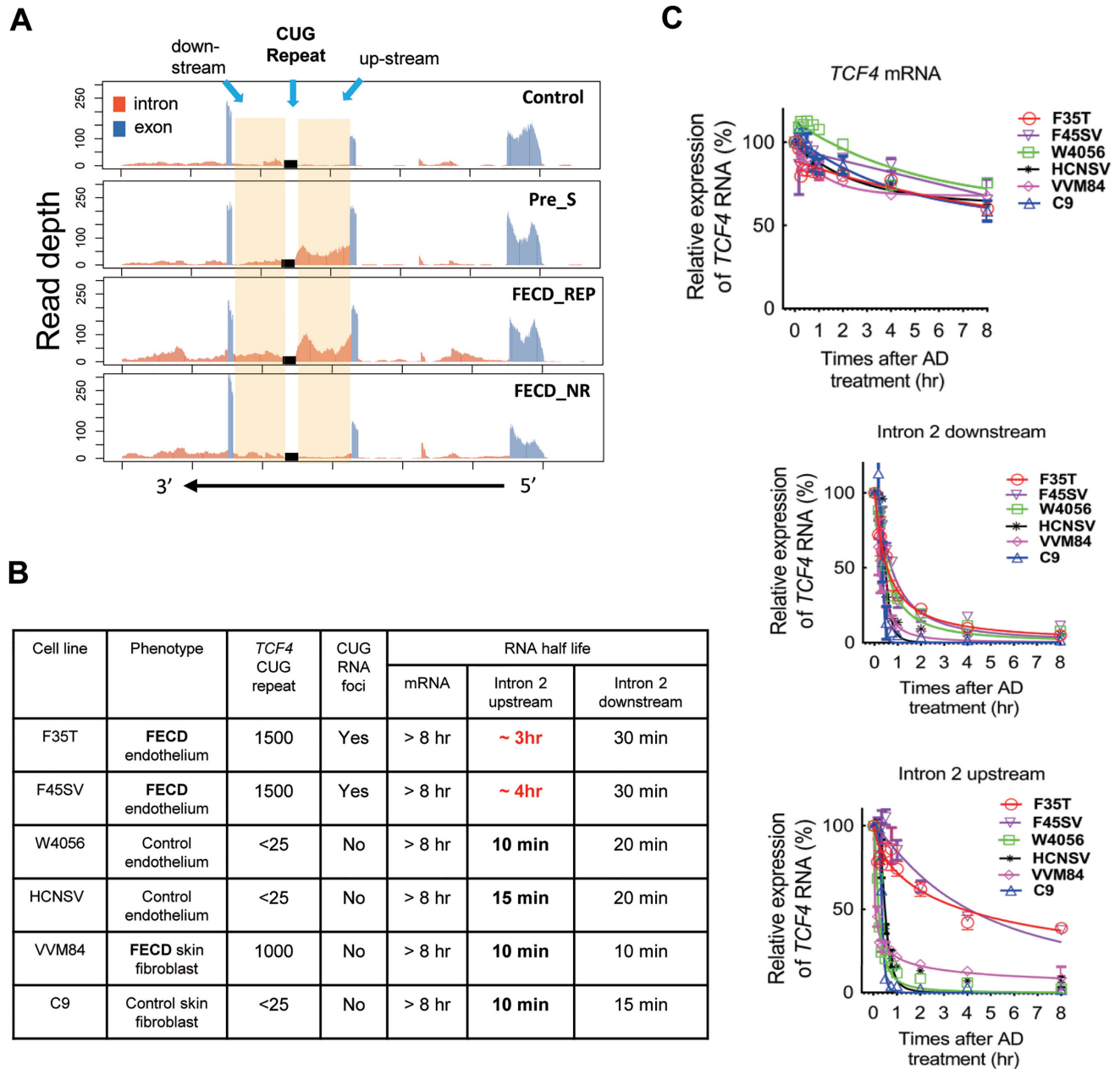
Figure 3. Intronic RNA is stabilized in Pre_S and FECD_REP tissue and FECD_REP patient-derived F35T & F45SV corneal endothelial cells. (**A**) Representative RNAseq data showing relative read number for *TCF4* intron 2 RNA from each sample cohort. (**B**) Summary of RNA half-life of different cell lines after treatment with actinomycin D. The much longer half life of intron 2 up-stream RNA in F35T & F45SV cells is highlighted in red. (**C**) Graphs of time-dependent RNA decay following treatment with actinomycin D.

VVM84 FECD skin fibroblasts and F35T FECD corneal endothelial cells both have the expanded CTG repeat, but the F35T corneal cells have nuclear CUG foci that can be detected by RNA-FISH while VVM84 skin cells lack detectable foci indicating cell-specificity for expanded CUG repeat RNA accumulation. Expression of *TCF4* is similar in F35T, F45SV and VVM84 cells (Supplementary Figure S5).

We treated cells with actinomycin D to arrest transcription and examine the half-life of mature *TCF4* mRNA and sequences either upstream our downstream relative to the

intronic *TCF4* CTG repeat using quantitative PCR (qPCR) (Figure 3B and C). Regardless of whether the expanded repeat mutation was present, the half-life of the mature *TCF4* mRNA was similar, >8 h (Figure 3C top). Likewise, the half-life of the intron 2 downstream region was also similar in each cell line, varying from 10 to 30 min (Figure 3C, middle). These data suggest that the mutation does not affect stability of the parent mRNA and has only a modest effect on the downstream region of intron 2.

By contrast, we observe a striking ∼20-fold increase in the half-life of the upstream region of intron 2 in FECD

corneal cells (F35T and F45SV) that possess the expanded CTG mutation and have detectable foci (Figure 3C, bottom). For cell lines that lacked expanded CUG nuclear foci, this region had a half-life of only 10 min. For FECD corneal cells, the half-life was 3–4 h.

This increased half-life of upstream intronic RNA is consistent with the observation from RNAseq of many more reads covering the upstream region of intron 2 *TCF4* from Pre_S or FECD_REP tissue relative to samples from individuals in the control or FECD_NR cohorts who lack the expanded CTG repeat (Figure 3A). Stabilization of the upstream portion of *TCF4* intron 2 in corneal endothelial tissue and cultured corneal endothelial cells makes more mutant repeat RNA available to perturb gene expression.

To address the possibility that the 5′ half of intron 2 might be retained in the mature mRNA, we performed PCR using primers complementary to exon 2 and exon 3 in both FECD and control cells. We observed a single PCR product of the predicted length in both FECD and control cells (Supplementary Figure S6). This data supports the conclusion that the intron is not detectably retained.

Changes in alternative splicing triggered by the expanded repeat within TCF4

CUG repeat RNA is known to bind the splicing factors muscleblind-like 1 and 2 (MBNL1 and MNBL2) (28–31). Previous studies have proposed that sequestration of MBNL1 and MBNL2 reduces levels of available MBNL protein, causing the changes of splicing observed in tissue from FECD_REP patients (27,32–33) and in patients with myotonic dystrophy who possess expanded CUG repeats within 3′-untranslated region of the *DMPK* gene (28–29,34).

We used RNAseq to evaluate splicing changes between control tissue and the Pre_S, FECD_REP and FECD_NR tissue cohorts (Figure 4). To classify changes, we used the FDR and delta PSI (the net change of inclusion percentage) as the determinant metrics. Any changes that are <0.001 on FDR and ≥0.15 on PSI were determined to be significant.

Regardless of which tissue was analyzed, the primary changes in alternative splicing relative to control tissue were increases in the number of skipped exons (SE) events (Figure 4A). FECD_REP or FECD_NR tissues showed more splicing changes than Pre_S tissue (Figure 4B). The greater number of splicing changes is consistent with the extensive cellular degeneration observed in late stage disease. However, while not as many as in tissue from advanced disease, ∼450 changes in alternative splicing were observed in Pre_S tissue.

Three hundred and thirteen of the alternative splicing events in Pre_S tissue involved exon skipping (Supplementary Table S3). Heatmap evaluation of 313 skipped exons in Pre_S tissue revealed that the genes hosting the skipped exon events clustered most closely with FECD_REP tissue (Figure 4C). One hundred and thirty-two skipped exon events were shared between Pre_S and FECD_REP tissue, compared to only 28 were shared between Pre_S and FECD_NR tissue (Figure 4B), consistent with the hypothesis that the changes in Pre_S tissue foreshadow the alterations observed in FECD-REP.

While not nearly as frequent, other forms of alternative splicing events, alternative 5′ splice site (A5SS), alternative 3′-splice site (A3SS), mutually exclusive exon (MXE) and retained intron (RI), also showed that the Pre_S cohort was more similar to the FECD_REP group than the FECD_NR cohort (Supplementary Figure S7A–D). Taken together, these data demonstrate that splicing changes in Pre_S tissue are precursors to the changes in FECD_REP—but not FECD_NR—late stage disease.

We then separated the top 25 skipped exon events in Pre_S tissue, which are shared with FECD_REP for evaluation. As with the overall group of 313 events, the splicing of these genes was much more like FECD_REP tissue than to control or FECD_NR tissue (Figure 4D). Of the top 25 skipped exon events shared by FECD_REP and Pre_S tissue, approximately half are also seen in tibialis anterior muscle of myotonic dystrophy type 1 (DM1) subjects (35) (Figure 4 D and E; Supplementary Figure S7E–G). This similarity, in spite of the comparison being made between samples from different tissues, suggests that the expanded CTG repeats in DM1 and FECD_REP have a common mechanism for producing splicing changes and that this mechanism is activated in Pre_S tissue.

Notable genes that show changes in splicing include the splicing factors *MBNL1* and *MBNL2*. Splicing of these genes was also changed in tissue from a myotonic dystrophy mouse model (35) and in a comparison of FECD_REP and FECD_NR tissue by Fautsch and colleagues (36,37). Changes in *MBNL1* and *MBNL2* splicing in Pre_S (Supplementary Table S3) cells may trigger other splicing changes and eventually lead to the larger scale change that characterizes late stage disease.

Changes in alternative splicing are quantitatively similar in Pre_S and FECD_REP tissue

We experimentally validated changes in splicing for Pre_S relative to control tissues (Figure 5 and Supplementary Figure S8). Genes were chosen for validation based on the gene expression level and the potential biological role of the gene suggested by pathway analysis (Figure 8).

We used reverse-transcription PCR to evaluate changes of splicing for six genes, *INF2*, *NUMA1*, *SORBS1*, *SYNE1*, *MBNL1* and *MBNL2* (Figure 5). INF2 protein associates with microtubules, and may affect cell shape. NUMA1 protein is component of the nuclear matrix and may affect mitotic spindle organization and proper cell division. *SORBS1* was chosen for its protein product's involvement in cell adhesion and extracellular matrix (ECM). *SYNE1* encodes nesprin-1 which links the nuclear membrane to the actin cytoskeleton and may also affect cell morphology. *MBNL1* and *MBNL2* were chosen because potential alterations in their expression might feed back into even greater changes in splicing.

The changes in splicing predicted by global RNAseq analysis were confirmed by visual inspection of genes using Sashimi plots (Figure 5A) and validated by semiquantitative analysis using reverse-transcriptase PCR (RT-PCR) (Figure 5B) using monolayer human corneal endothelial tissue. Pre_S tissue showed the changes in splicing predicted by our RNAseq data (Figure 5). Both Sashimi
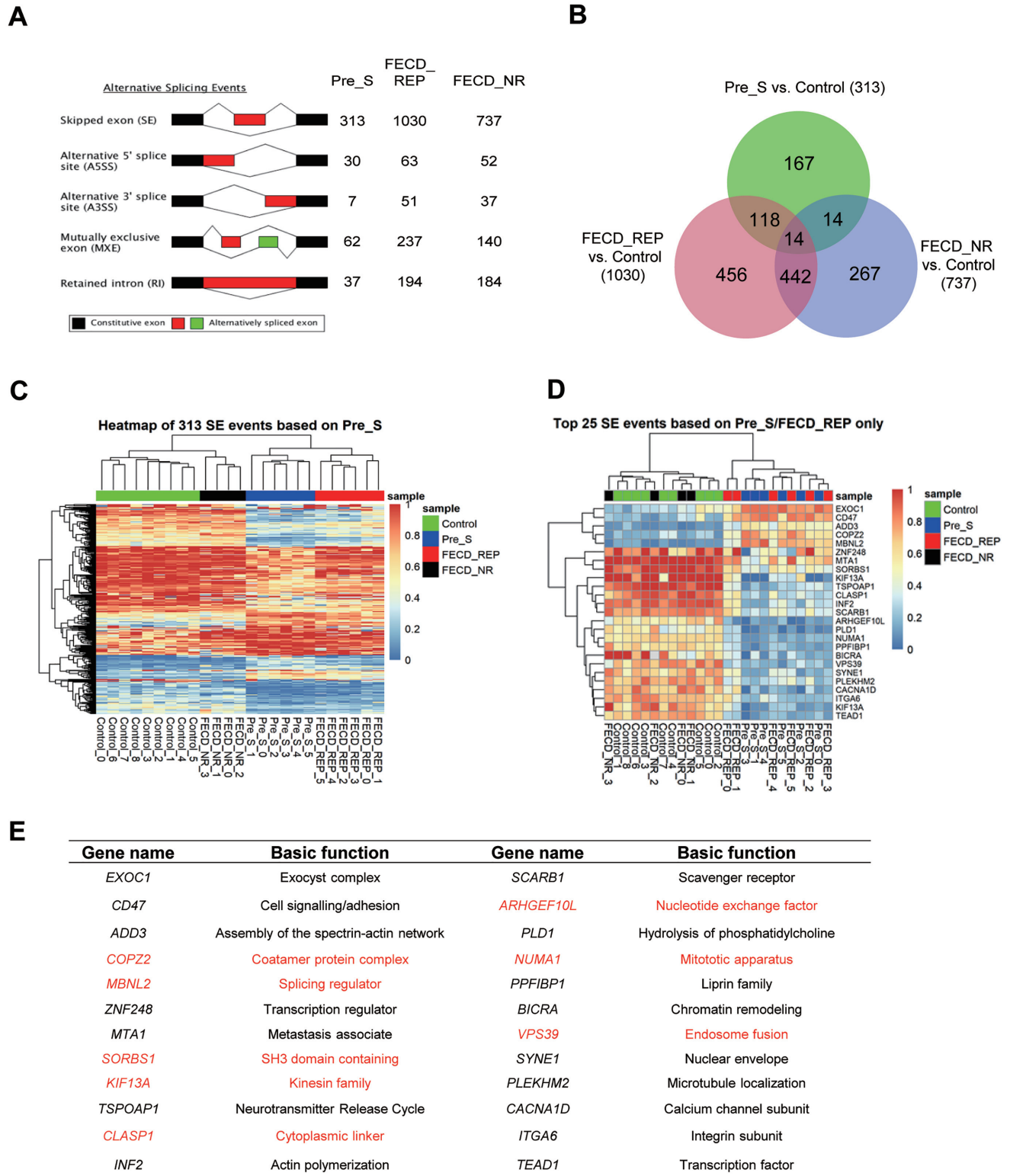
Figure 4. RNAseq data demonstrates changes in splicing in pre-symptomatic, FECD_REP and FECD_NR tissue. (**A**) Alternative splicing events in Pre_S, FECD_REP and FECD_NR cohorts. (**B**) Overlap between sample cohorts for skipped exon (SE) events showing that late stage disease FECD_REP and Pre_S cohorts cluster differently from control or FECD_NR cohorts. (**C**) Heat map comparing inclusion levels of exons among four cohorts. The chosen SE events were based on the 313 significant skipped exon events identified in Pre_S tissues. The Pre_S cohort is more similar to FECD_REP than it is to the other two cohorts. (**D**) The similarity of Pre_S and FECD_REP is emphasized by a heat map comparing top 25 SE events in common between FECD_REP and Pre_S tissues and corresponding changes in FECD_NR and control tissues. (**E**) Identity of top 24 SE events in common between FECD_REP and Pre_S. Genes in red are also observed in RNAseq data from DM1 tissue, demonstrating substantial overlap despite different tissue origins. The threshold used in identifying the significant events: FDR < 0.001, |IncLevel Difference| ≥0.15.
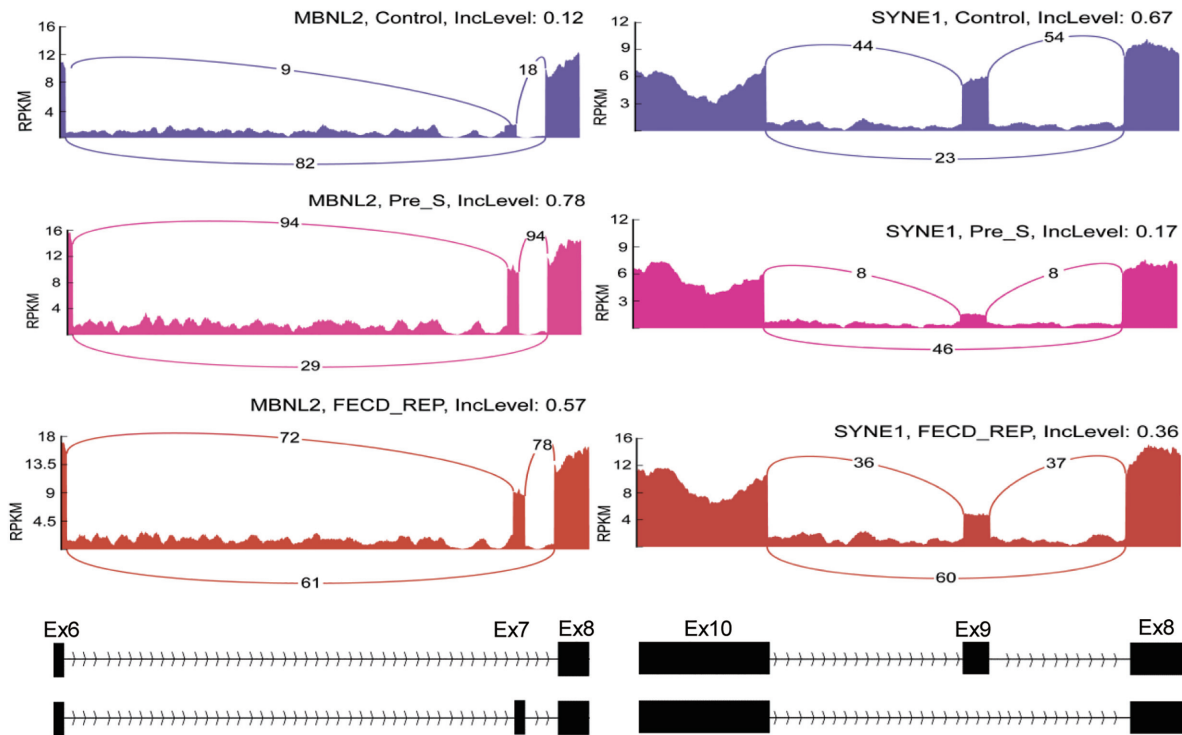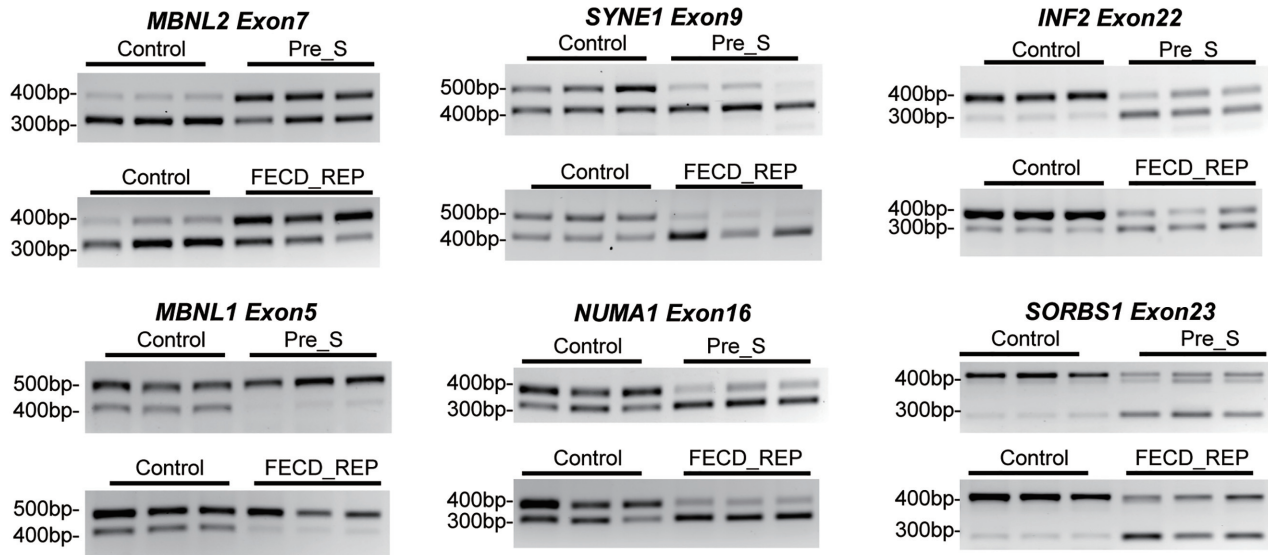
Figure 5. Differential gene splicing is similar in Pre_S and FECD_REP tissue. (**A**) Representative Sashimi plots showing changes in alternative splicing for MBNL2 and SYNE1 transcripts in Control vers Pre_S and FECD_REP tissues. (**B**) Reverse-transcriptase PCR evaluation of RNA splicing changes in Control, Pre_S and FECD_REP tissue. All measurements used surgically prepared corneal endothelium from disease patients or donors detailed in Supplementary Table S1A. The genomic locations of spliced exons are listed in Supplementary Table S1B.

plots and semi-quantitative RT-PCR analysis reveal that the absolute magnitude of splicing changes are similar in FECD_REP late stage disease tissues. The amount of change in splicing in Pre_S tissue is substantial, similar to that observed in late stage tissue.

We also carried out a differential alternative splicing analysis between Pre_S and FECD_REP tissues directly, rather than using the Control as the reference as described above. Although there are hundreds of significant skipped exon events identified between Pre_S and FECD_REP, out of 132 skipped exon events shared between Pre_S/Control and FECD_REP/Control comparison, only five of them are significantly different between Pre_S and FECD_REP (Supplementary Table S4). This means that majority of shared skipped exon events identified between Pre_S/Control and FECD_REP/Control have similar magnitude in inclusion of exon level changes, suggesting that missplicing of key genes does not gradually increase as disease symptoms progress, rather substantial missplicing is a leading indicator of disease.

Differential gene expression

We compared gene expression level changes in Pre_S, FECD_REP and FECD_NR tissue cohorts relative to control tissue. To classify changes as significant, we used the adjusted *P*-value generated by Cuffdiff, one of the programs within Cufflinks suite, as the determining metric. Changes with a *P*-value < 0.05 were deemed significant and included in our analysis. Analysis by DeSeq2, an alternate program for evaluating RNAseq data, produced similar results when compared with the output of CuffDiff (Supplementary Figure S9 and Supplementary Table S5).

All sample cohorts showed expression changes relative to control tissue (Pre_S: 215, FECD_REP: 1330; FECD_NR: 696) (Figure 6A and B). The greater number of gene expression changes in the FECD_REP and FECD_NR tissues is consistent with the severe cellular phenotype observed in late stage disease. A total of 602 out of the 696 genes differentially expressed in the FECD_NR tissues were also found in the FECD_REP group suggesting significant overlap of the common final molecular genetic mechanisms in the two forms of late-stage disease.

Pre_S tissue had 215 genes with significantly altered expression levels relative to control tissue. Only five changes in gene expression were uniquely shared between Pre_S and FECD_NR tissue, compared to 73 shared changes with FECD_REP (Figure 6A). The closer relationship between Pre-S and FECD_REP is consistent with our splicing data (Figures 4 and 5) and supports the conclusion that patterns of gene expression in mutant expanded repeat cells are established long before symptoms or disease findings are observable.

Volcano plots allow a global overview of individual gene expression changes. They are useful for visualizing patterns of changes and identifying 'outlier' genes that combine highly significant changes in gene expression with higher fold changes. The fold change among top genes are less in Pre_S than in FECD_REP or FECD_NR and the identity of top genes differ (Figure 6B and C; Supplementary Figure S10 and Tables S6–8). This finding is consistent with

the severity of late stage disease and the disruption of many gene expression programs.

Evaluation of Pre_S tissue may provide a window to identify early gene drivers before symptom-driven secondary changes in gene expression overwhelm analysis. The top 20 differentially over-expressed genes identified in Pre_S tissue (Figure 6C) include genes involved in the ECM and its assembly, cochlin (*COCH*) (38,39), fibronectin (*FN1*) (40) and thrombospondin (*THBS2*) (41).

Quantitative analysis of gene expression changes

We used quantitative PCR (qPCR) to confirm the changes in the level of gene expression detected by RNAseq in both Pre_S and FECD tissue (Figure 7). qPCR measurements of corneal endothelial tissues were challenging because of the limited amount of material available, but we could compare expression of eight genes identified in our RNAseq data, *FN1*, *COL4A2*, *COCH*, *CTGF*, *MSI1*, *LUM*, *KDR* and *SOD3*. Four of these genes, *FN1*, *COL4A2*, *CTGF*, *KDR*, were within the fibrosis pathway. *COCH* and *LUM* encode ECM proteins, *MSI1* encodes a RNA binding protein/splicing factor and SOD3 protein is related to oxidative stress. The observed changes in gene expression confirm our RNAseq results (Figure 6B and Supplementary Table S9).

Pathway analysis

To elucidate the potential impact of changes in the expression of individual genes on physiologic processes, we applied Ingenuity Pathway Analysis (IPA) to our RNAseq data. Overwhelmingly, the top common canonical pathway was hepatic fibrosis/hepatic stellate cell activation (42) (Figure 8A and Supplementary Table S9). Involvement of the hepatic fibrosis pathway genes in FECD_REP and FECD_NR is consistent with the observed accumulation of ECM in advanced FECD with thickening of Descemet's membrane with focal excrescences (guttae) and with previously reported RNA measurements using tissue from late stage disease (43). Our FECD transcriptome data indicates robust ECM production in late-stage disease possibly regulated by transforming growth factor-β (TGF-β), the most potent fibrogenic cytokine released by a number of cell populations in the body including the liver (42) (Supplementary Figure S11A and B).

Activation of the fibrosis pathway was also observed in Pre_S tissue (Figures 8A and B and 9). Genes that showed statistically significant increases in expression include fibronectin *FN1*, one of the highest differentially expressed genes in Pre_S tissue (Figure 7). Other genes include connective tissue growth factor (*CTGF*) and four members of the collagen alpha chain family including *COL1A2* which is also abundant in liver fibrosis (44). Kinase insert domain receptor (KDR, also known as vascular endothelial growth factor receptor-2) showed decreased expression. Relevant in the fibrosis pathway, KDR protein interacts with VEGF to mediate vascular endothelial cell proliferation. These genes are also disturbed in advanced disease - another indication that gene expression programs associated with fibrosis are activated in pre-symptomatic carriers prior to observable symptoms of disease.
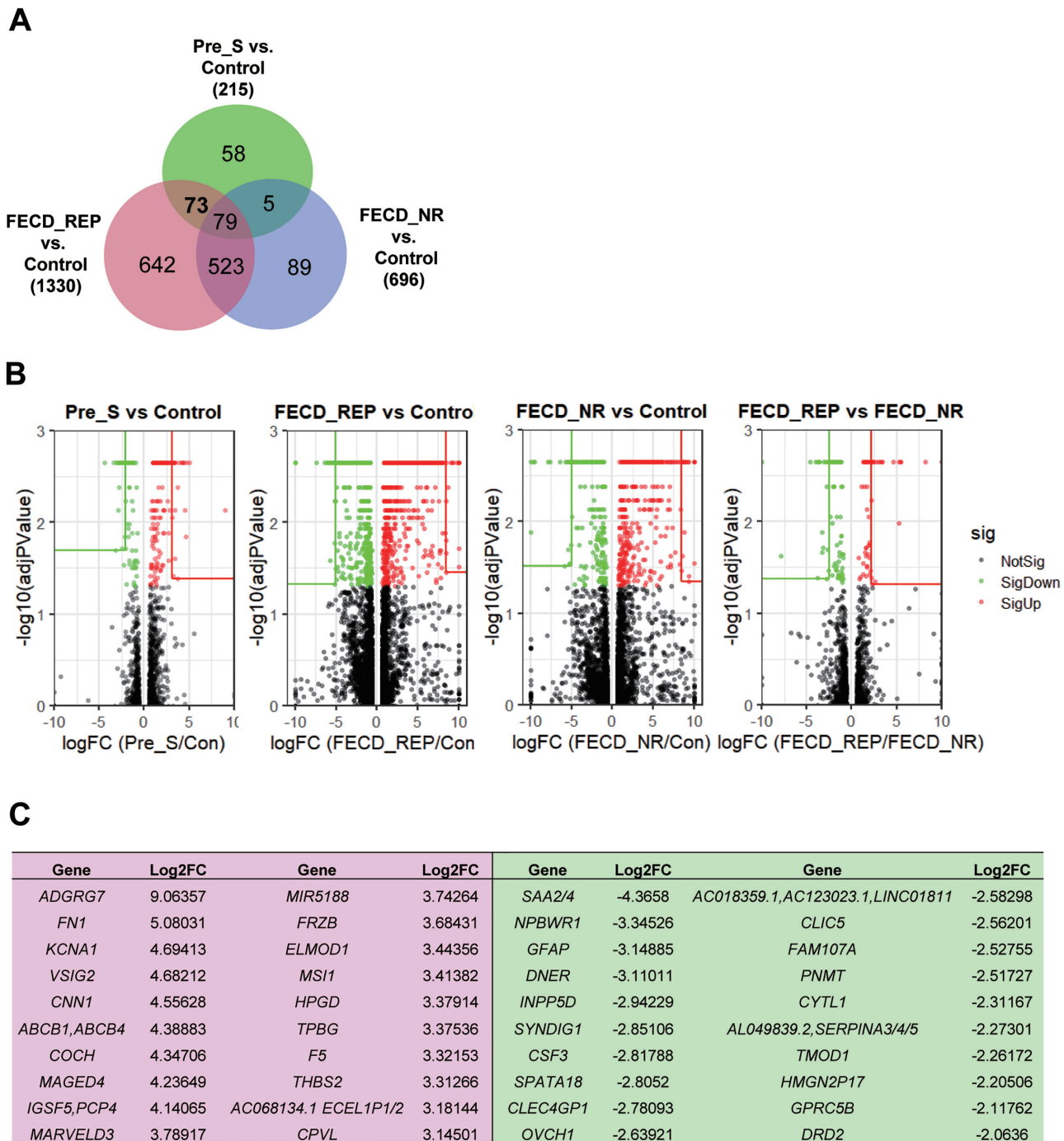
Figure 6. RNAseq analysis of gene expression. (**A**) Overlap for gene expression of Pre_S, FECD_REP and FECD_NR cohorts. Numbers indicate the number of differentially expressed genes identified in each individual region. (**B**) Volcano plots for Pre_S versus Control, FECD_Rep versus Control, FECD_NR versus Control and FECD_REP versus FECD_NR and top genes in each comparison. (**C**) Top 20 upregulated and downregulated genes in Pre_S versus control comparison.

The pathways underlying FECD_REP and FECD_NR advanced stage disease had significant overlap. There were numerous shared pathways implicating the immune system related to helper T-cell activation, signaling and neuroinflammation (Figure 8A and C). Marked overexpression of genes encoding proteins on the surface of antigen-presenting cells including the B7 protein, *CD86* (>2000-fold

increase) and class II major histocompatibility (MHC) proteins (>250-fold) both required for these cells to activate helper T cells implicate the immune system in both FECD groups in late-stage disease (45) (Figure 8C and Supplementary Figure S11C–E).

A few molecular pathways were changed in FECD_REP but not FECD_NR (Figure 8D). The canonical pathway
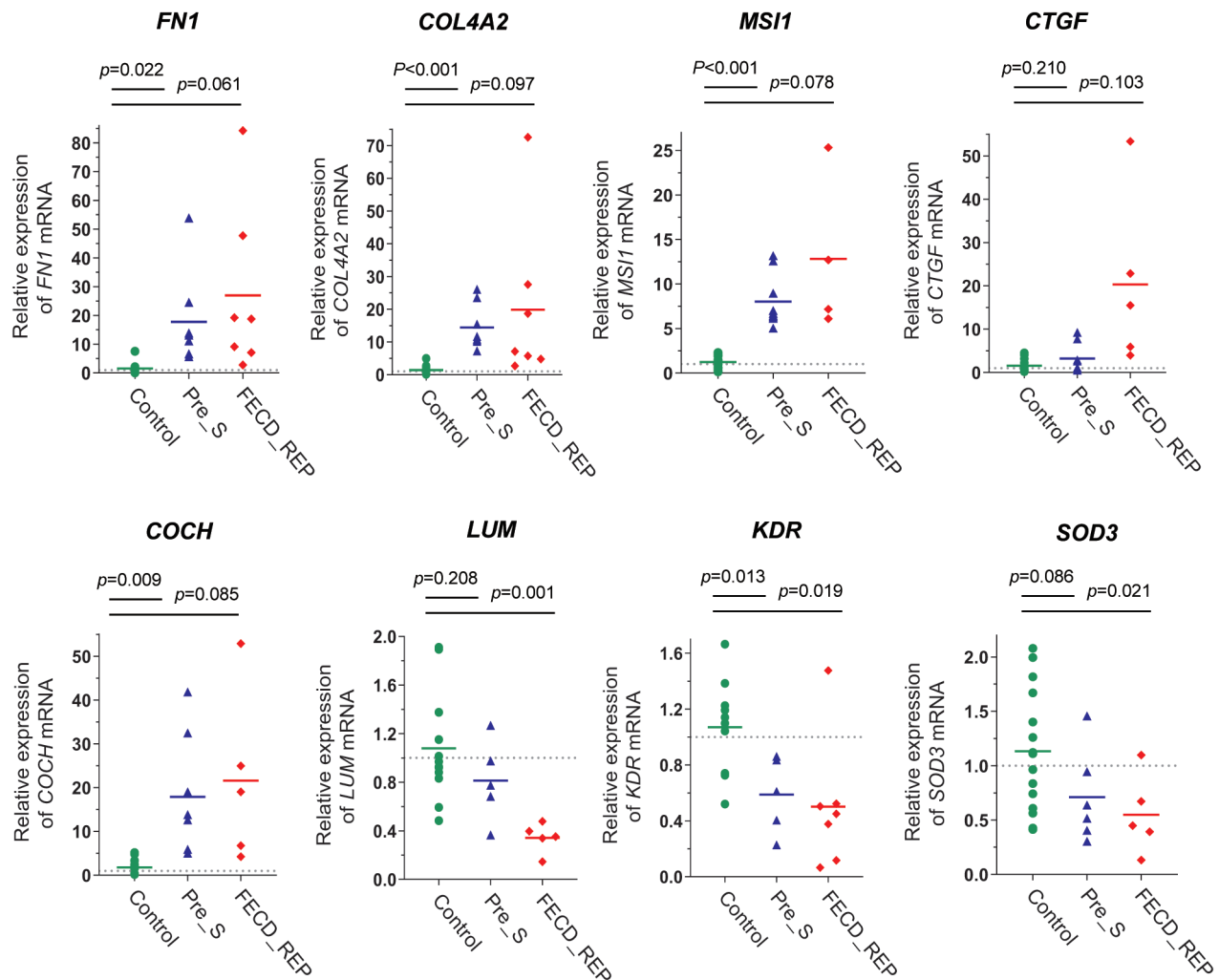
Figure 7. Changes in gene expression in Pre_S and FECD_REP tissue. qPCR evaluation of gene expression levels in Control, Pre_S and FECD_REP tissue. All measurements used surgically prepared single cell monolayer corneal endothelium from disease patients or donors detailed in Supplementary Table S1.

related to mitochondrial dysfunction showed the large difference with *P*-values of 10^{-5} and zero respectively for FECD_REP and FECD_NR (Figure 8A and Supplementary Figure S11F–H). Decreased expression of oxidative phosphorylation genes was more pronounced in FECD_REP compared to the FECD_NR (Figure 8D).

DISCUSSION

Identifying the early drivers of late-onset disease is important for understanding disease progression and developing therapeutics. Studying early drivers, however, is often not practical because pre-symptomatic tissue is difficult to obtain. Because the expanded CTG repeat mutation within *TCF4* intron 2 that causes FECD_REP is so prevalent (3% of the Caucasian population), significant numbers of pre-symptomatic samples can be obtained from individual donors positive for the CTG expansion. These tissues, together with FECD_REP, FECD_NR and control tissues (Figures 1 and 2) provide an advantageous model for better investigating the early links between expanded trinucleotide repeat mutations and disease.

The goal of this study was to understand whether the expanded CUG RNA repeat was changing gene expression in Pre_S tissue and how such changes might relate to the gene expression and phenotypic changes known to occur in late-stage FECD.

FECD is a disease of mutant RNA

FECD_REP is caused by an expanded CTG trinucleotide repeat within mutant *TCF4* intronic RNA (Figure 10) (15–18). Remarkably, FECD is also caused by the mutant CTG expanded repeat within the 3′-untranslated region of the *DMPK* gene that is also associated with myotonic dystrophy (19–21). Unlike other trinucleotide repeat diseases where the contribution of mutant RNA is debated, these data demonstrating that CUG RNAs expressed from two different genes are both responsible for FECD offers strong support for RNA playing a central role in the molecular origins of the disease.

The expanded CUG RNA can be detected by fluorescent *in situ* hybridization (FISH) as RNA foci. These RNA foci are a hallmark of both Pre_S and FECD_REP corneal en-

A

| Canonical Pathways	FECD_NR	FECD_REP	Pre_S
Hepatic Fibrosis	20.81	16.30	3.03
Neuroinflammation Signaling	5.72	7.05	0.00
GP6 Signaling	5.71	6.78	2.27
IL-8 Signaling	4.39	6.62	0.98
Th1 and Th2 Activation	7.83	6.33	0.00
Osteoblasts/Osteoclasts/Chondrocytes in Rheumatoid Arthritis	2.94	6.23	0.88
Axonal Guidance Signaling	3.62	5.80	1.24
Macrophages/Fibroblasts/Endothelial Cells in Rheumatoid Arthritis	3.95	5.69	0.87
Apelin Liver Signaling	4.40	5.64	1.71
Th1	5.63	5.48	0.00
NFAT Regulation/Immune Response	7.47	5.41	0.00
Th2	6.43	5.29	0.00
Osteoarthritis	6.24	5.26	2.70
CD28 Signaling in T Helper Cells	5.06	5.07	0.00
STAT3 Pathway	3.15	5.02	0.00
Mitochondrial Dysfunction	0.00	4.96	0.38
Dendritic Cell Maturation	3.79	4.93	1.09
PTEN Signaling	1.81	4.91	1.06
Inhibition of Angiogenesis by TSP1	2.79	4.56	0.61
Inflammasome	2.83	4.53	0.00
PKCθ Signaling in T Lymphocytes	4.49	4.51	0.00
Molecular Mechanisms of Cancer	4.80	4.39	0.37
IGF-1 Signaling	5.80	4.30	1.80
Phagosome Formation	2.00	4.26	1.53
T Cell Exhaustion Signaling	2.91	4.22	0.00

B
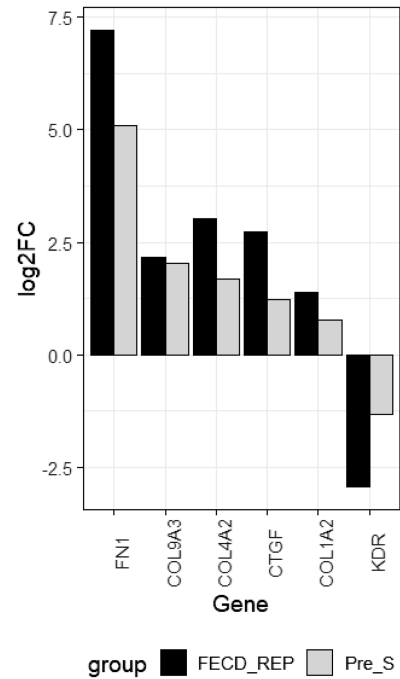
C
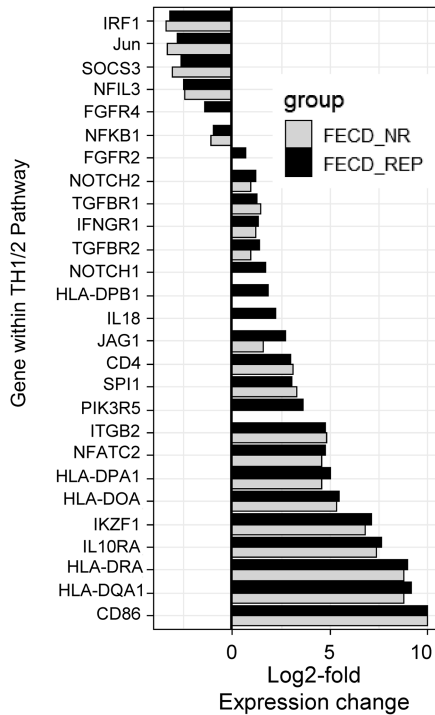
D
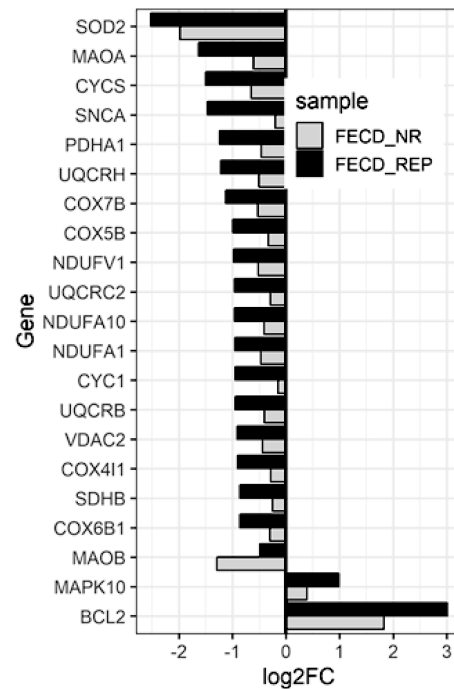
Figure 8. Pathway analysis. (**A**) Top 25 pathways based on FECD_REP. The *P*-values for three cohorts for each pathway are shown in the last three columns. (**B**) Shared differentially expressed genes in Fibrosis pathway between Pre-S and FECD_REP. (**C**) Changes in Th1/2 pathway genes for FECD_NR and FECD_REP. (**D**) Significant expression level changes in genes involved in mitochondrial dysfunction seen in FECD_REP but not in FECD_NR. FPKM > 1.5, fold change > 1.5, FDR ≤ 0.05.
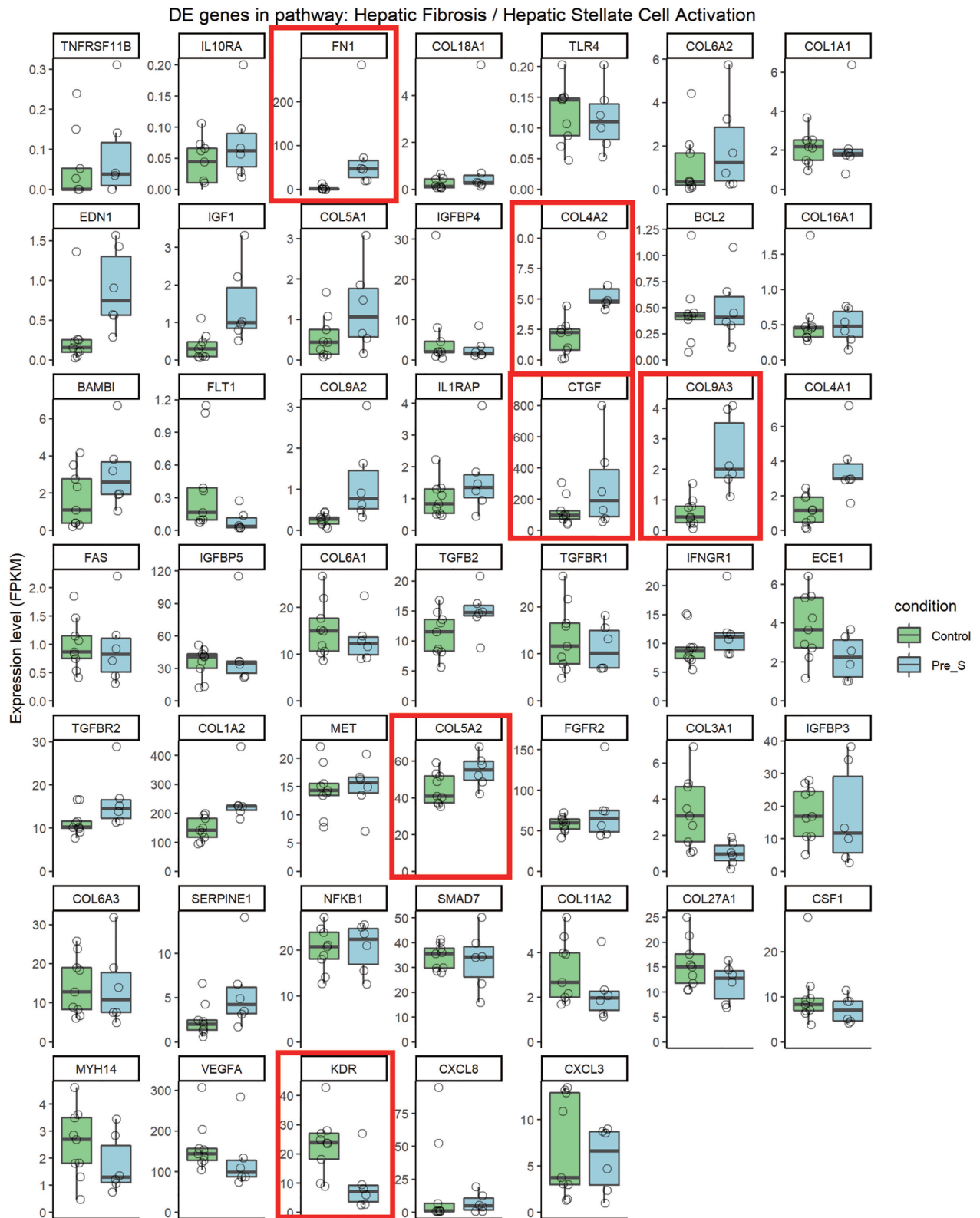
Figure 9. Expression profile of genes in Hepatic fibrosis pathway for Pre_S cohort. FPKM > 1.5, fold change > 1.5, FDR ≤ 0.05. The six largest fold changes in gene expression (FN1, COL4A2,CTGF, COL9A3, COL5A, KDR) (Figure 8) are boxed.
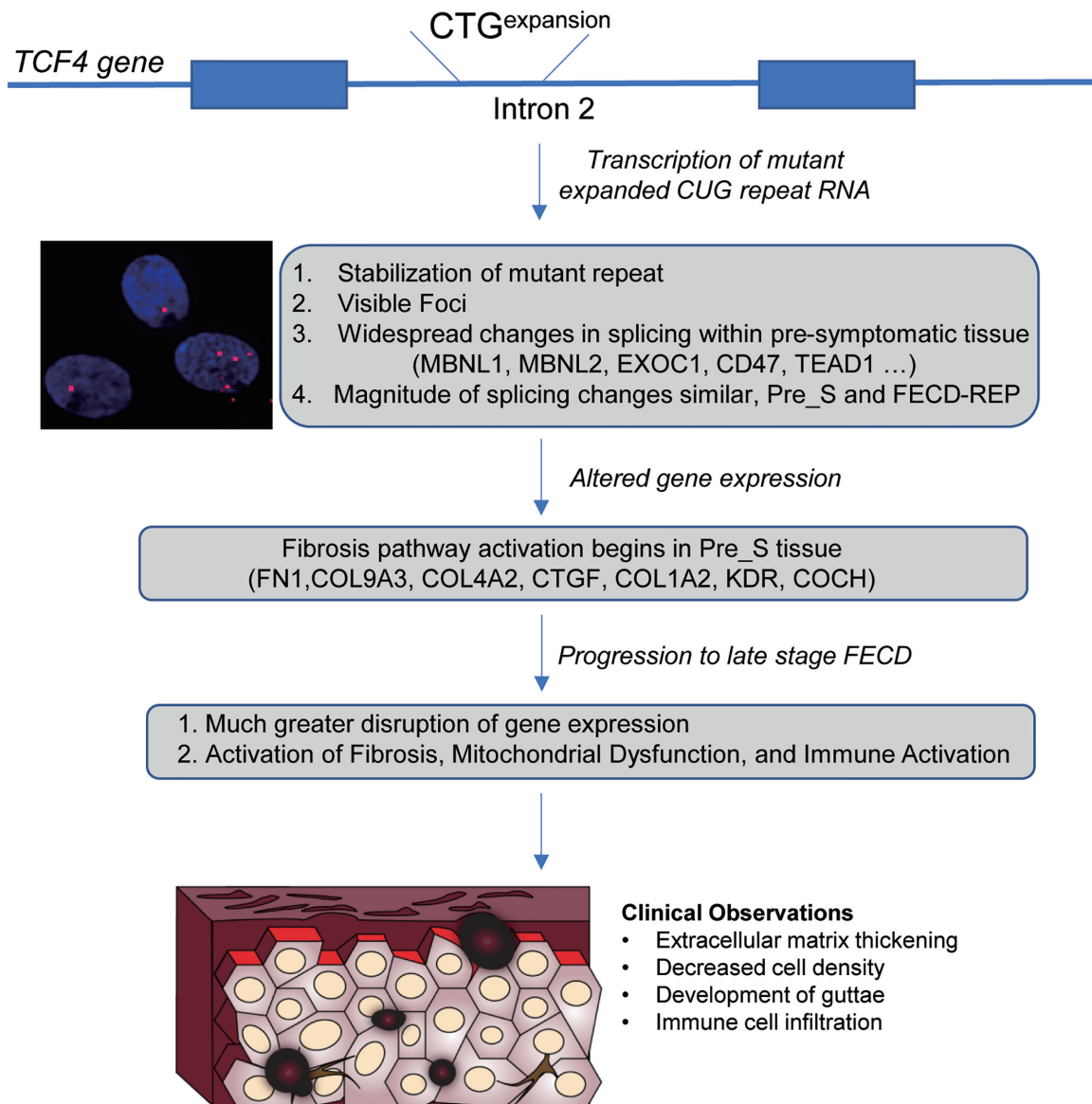
Figure 10. Schematic diagram of FECD molecular and disease progression from CTG expanded repeat mutation at the *TCF4* locus to advanced FECD. Initially, the GTC expansion expresses the CUG repeat RNA. The expanded repeat mutation that causes FECD-REP is a relatively common genetic mutation, making Pre_S tissue available for analysis. While Pre_S tissue appears normal upon clinical observation, foci can be detected, we observe changes in splicing and gene expression, the mutant intronic repeat is stabilized and early signs of fibrosis pathway activation are apparent. In late stage disease, more pronounced changes in splicing and gene expression accompany clinically observable systems and loss of vision.

dothelium (Figure 10) (25). While Pre_S and FECD_REP tissue both possess the RNA trigger for FECD, Pre_S tissue is visually indistinguishable from control tissue upon specular imaging. By contrast, FECD_REP tissue is dramatically different from control tissue, with reduced cell density and the formation of focal collagen accumulations known as guttae.

Consistent with our published clinical FECD cohort at our institution (18), the mean ages of the individuals in this study with late-stage disease in FECD_REP and FECD_NR groups were comparable at 66.5 and 68.8 years, respectively. The Pre_S group had a mean age of 46.8 years suggesting that at least two decades from onset of mutant CUG repeat RNA mediated mis-splicing and al-

tered expression of ECM genes are required to develop findings of late-stage disease which is clinically compatible with the slow natural disease progression of this age-related disorder.

What does it take for an individual with a mutant *TCF4* repeat expansion to become a FECD patient? We can only speculate at this stage. It is possible that the gene expression changes observed in pre-symptomatic individuals increase the likelihood for dysfunction in individual cells. Eventually, this dysfunction crosses a molecular threshold and leads triggers the fibrosis pathway in individual cells. As patients age, more cells are affected, producing the findings of FECD visible to the ophthalmologist and the symptoms related to loss of vision noticed by patients.

The FECD CUG repeat RNA is present at only a few (<10) copies per cell in disease tissue (26). It is likely that each 'focus' detected by FISH is a single RNA molecule. A low copy number for a disease-causing RNA has also been observed for myotonic dystrophy type 1(DM1) (46) and *C9orf72* ALS/FTD (47). We find that the CUG repeat expansion stabilizes *TCF4* intronic RNA in corneal cells (Figure 3). This enhanced stability may contribute to an ability of a small number of RNA molecules to bind protein sufficiently to affect overall function in cells and eventually produce observable symptoms that characterize a delayed onset disease like FECD.

Although we observed stabilization of the mutant intronic RNA in two patient-derived corneal cell lines and not in skin fibroblasts harboring the expansion, additional studies in other cell types are warranted to test the hypothesis that there is a corneal cell-specific mechanism that stabilizes the expanded CUG repeat.

Previous work has suggested that the CTG repeat within intronic *TCF4* and other microsatellite expansions are associated with intron retention (48). Our detection of nuclear CUG RNA foci by FISH and upstream intronic RNA due to its increased half-life may be compatible with a disease model where the *TCF4* intron with the expanded repeat is spliced out, forms a linear stable intronic sequence RNA, and undergoes preferential 3′ to 5′ exonuclease degradation in the nucleus.

There is no definitive mechanistic insight into how the relatively rare expanded CUG repeat RNA can cause FECD. How one or a few copies of RNA triggers widespread changes in gene expression and late onset disease remains a major unanswered question. However, studies of myotonic dystrophy have suggested that the CUG repeat binds members of the muscleblind-like (MBNL) protein family. MBNL proteins are splicing factors and their sequestration reduces the concentration of free MBNL in cell nuclei and affects splicing. Overexpression of MBNL1 can reverse RNA missplicing and myotonia in a DM1 mouse model (34). MBNL is also associated with mutant CUG RNA in FECD cells and tissue (27,32–33). Blocking the CUG repeat region using antisense oligonucleotides can reverse missplicing in DM1 (caused by a CTG repeat within the *DMPK* gene) (49–52) and FECD (26,53) tissues. These reports support the hypothesis that the mutant RNA may act by binding to MBNL and affecting splicing (Figure 10).

We have recently reported that relatively little MBNL protein is in the nuclei of FECD_REP cells and human tissue (54). Using quantitative protein titrations against a known standard, we calculated that there were 65,000 copies of MBNL1 and MBNL2 per cell and less than 2000 copies were present in cell nuclei. Low copy numbers for MNBL in the nuclei of affected tissue are consistent with the hypothesis that even a small amount of mutant expanded CUG repeat RNA may be sufficient to affect the available pool of MNBL protein. A reduction in available MBNL protein would produce the alterations of splicing that are a hallmark of FECD_REP disease (Figures 4 and 5). It is also possible that the MNBL:mutant RNA interaction may nucleate additional protein or RNA interactions to amplify disruptive effects on gene expression and changes in alternative splicing.

Expanded CUG mutant RNA causes splicing changes in presymptomatic tissue

Many of the alterations in splicing observed in FECD_REP tissue also define gene expression in presymptomatic tissue, Pre_S (Figures 4 and 5). The splicing factors MBNL1 and MBNL2 are among the genes showing altered splicing in Pre_S samples. Altered splicing of MBNL1 and MBNL2 may decrease the reservoir of functional MNBL protein. Less functional MBNL protein, in addition to sequestration of MNBL protein by the expanded CUG repeat, may help push corneal endothelial cells toward full blown FECD.

The similarity of alternative splicing changes between symptomatic FECD_REP and presymptomatic Pre_S tissue is much greater than that between Pre_S and FECD_NR samples. The data suggest that there are fundamental differences in the origins of the two forms of FECD. While their origins differ, late-stage FECD_REP and FECD_NR converge at a common set of clinical findings.

We note that the FECD_NR cohort shares gene splicing changes with both the Pre_S (27) and FECD_REP (456) cohorts. The 28 splicing events shared by Pre-S and FECD_NR samples are less than the 132 splicing changes shared by the FECD_REP and FECD_NR samples. Nevertheless, the commonality of splicing changes between FECD_NR and the other cohorts may indicate the presence of an undiscovered trinucleotide repeat expansion that might trigger changes in gene splicing.

The data also suggest that splicing changes and perturbation of ECM genes seen in FECD_REP late stage disease tissue begin to be observed long before symptoms are observed during standard clinical evaluation. The molecular trigger, mutant RNA and early molecular changes, altered splicing and observable RNA foci, co-exist in cells that appear to have a normal phenotype. The changes in the magnitude of RNA splicing between control tissue and either Pre_S or late stage FECD_REP samples are similar. This observation suggests that the mutant RNA triggers the splicing changes in key genes independent of the progression of disease.

The finding that splicing is an early trigger has important implications for the development of agents to treat FECD. It is reasonable to expect that such agents would be most effective when administered early in disease progression prior to production of ECM with degeneration of the corneal endothelium and activation of the immune system. During drug development it should be possible to monitor the changes in alternative splicing and expression of ECM biomarkers caused by expression of the mutant trinucleotide repeat and rank drug candidates by their ability to return splicing to a more normal state—agents that reverse the splicing defect would be promising development candidates. Monitoring splicing of key genes would offer a rapid and definitive assay for screening compounds. We have previously shown that synthetic oligonucleotides complementary to the CUG repeat can reverse the splicing defect (25).

Recently, Fautsch and colleagues reported that some individuals without previously noted guttae who possessed the expanded repeat but developed 'non-FECD corneal edema' do not show changes in alternative splicing (36). These results might appear to be in conflict with our observation

that all individuals with expanded CUG repeat RNAs exhibited substantial changes in alternative splicing, including many shared between Pre_S and FECD_REP individuals (Figure 4). We note that the expanded repeat positive individuals with corneal decompensation without findings of FECD were between 67 and 83 years old, much older than any individual in our Pre_S cohort. As postulated by Fautsch and colleagues, it is possible that these individuals possessed protective mutations that might prevent splicing changes and block the molecular events leading to symptomatic FECD.

As seen in many other unstable microsatellite diseases, we reported a significant positive correlation between the clinical severity of FECD and the triplet repeat length in peripheral leukocytes (18). Previously, we observed a trend of toward a positive correlation between the number of CUG repeat RNA foci in endothelial cells of surgical samples and triplet repeat allele but it did not reach statistical significance (26). However, our RNA-seq tissue cohort in this study was not adequately powered to examine for a possible correlation of differential splicing and expression of genes to repeat length.

The fibrosis pathway is activated in Pre_S tissue

In the clinic, late stage FECD is a disease of cellular degeneration and aberrant ECM deposition (Figure 10). Previous studies of FECD_REP tissue have supported activation of the fibrosis pathway as a primary cause for late stage disease pathology (33). We confirmed fibrosis as the highest ranked canonical pathway in both late stage FECD_REP and FECD_NR (Figure 8A). The fibrosis pathway is also activated in Pre_S tissue (Figure 8A and B). These results demonstrate that changes in splicing, changes in gene expression and changes in a key disease pathway begin years before symptoms are observed.

Among the top ten genes in pre-symptomatic tissue versus control tissue are cochlin (*COCH*) and fibronectin (*FN1*) with >16- and >32-fold change increases respectively. Cochlin is a secretory ECM protein originally identified in the cochlear cells of the inner ear (38). *COCH* is also expressed by the endothelial cells within the trabecular meshwork of subjects with primary open-angle glaucoma (POAG), another common age-related degenerative disorder (39). There appears to be a convergence of two age-related disorders of the anterior segment of the eye mediated by *COCH*. Primary open-angle glaucoma may be more prevalent in patients with advanced FECD (55). Primary open angle glaucoma is also a disease mediated by transforming growth factor-β which increases aqueous humor outflow resistance by dysregulation of ECM genes in the endothelial cells lining the trabecular meshwork (56). One area for future research will be to determine whether *COCH* or other ECM genes with altered expression in Pre_S tissue can serve as biomarkers for early detection of FECD or to facilitate the monitoring of clinical trials.

All late onset diseases share a common question 'What triggers disease onset and what factors influence when disease onset occurs? Our data suggest that the CUG repeat expansion triggers changes in gene splicing. These changes are associated with gene expression changes, including changes in genes associated with fibrosis. We speculate that redundancy within cells and throughout the corneal endothelium delay the appearance of visible changes. Eventually, however, disruption of cellular processes becomes substantial enough to trigger outright fibrosis, which in turn leads to great disruption of corneal cells and tissues and visible symptoms of FECD.

Late stage FECD tissue is characterized by changes in immune cell-related and mitochondrial dysfunction pathways

Both FECD_REP and FECD_NR tissues show activation of genes related to immune system required for helper T-cell activation, signaling and neuroinflammation (Figures 8A and C and Figure 10). In both groups, we detected a 2000-fold increase in *CD86* and marked upregulation of MHC genes required by antigen presenting cells to activate helper T cells. This gene expression data along with the recent observation of cells with a dendritic morphology and positive for the hematopoietic marker CD45 in the endothelial tissue keratoplasty specimens of patients (42) suggest an important role for antigen presenting cells in late-stage disease in both forms of FECD.

The mitochondrial dysfunction pathway is activated in FECD_REP tissue with the expression of over twenty genes changed, little change was seen in FECD_NR tissue (Figure 8D). Further research will be needed to determine whether this difference plays a role in disease progression or response to treatment.

CONCLUSION

FECD has many advantages as a model for understanding the origins of trinucleotide repeat disease because presymptomatic tissue is relatively accessible. Examination of Pre_S tissue reveals changes in gene expression that preview the more extensive changes in late stage disease. In particular, there is early activation of key genes associated with the fibrosis pathway, the pathway that defines the primary phenotype observe during advanced disease. Splicing patterns and levels of expression for key genes change decades prior to observation of the clinical manifestations of FECD_REP. Surprisingly, many changes in alternative splicing are similar in magnitude in Pre_S and advanced stage FECD_REP tissue. Many altered alternative splicing changes are shared with myotonic dystrophy, another disease caused by expanded CTG trinucleotide repeats, and it is possible that our findings will also be applicable to the genesis and temporal progression of other trinucleotide repeat diseases.

DATA AVAILABILITY

The Gene Expression Omnibus (GEO) accession number is GSE142538.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142538.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Barrientz,B., Nicholas,S.E., Whelchel,A., Sharif,R., Hjortdal,J. and Karamichos,D. (2019) Corneal injury: clinical and molecular aspects. *Exp. Eye Res.*, **186**, 107709.
2. Eghrari,A.O., Riazuddin,S.A. and Gottsch,J.D. (2015) Fuchs corneal dystrophy. *Prog. Mol. Biol. Trans. Sci.*, **134**, 79–97.
3. Vedana,G., Villarreal,G. and Jun,A.S. (2016) Fuchs endothelial corneal dystrophy: current perspectives. *Clin. Opthamol.*, **10**, 321–330.
4. Matthaei,M., Hribek,A., Clahsen,T., Bachmann,B., Cursiefen,C. and Jun,A.S. (2019) Fuchs endothelial corneal dystrophy: clinical, genetic, pathopysiologic, and therapeutic aspects. *Ann. Rev. Vis. Sci.*, **5**. 151–175.
5. Chi,H.H., Teng,C.C. and Katzin,H.M. (1958) Histopathology of primary endothelial-epithelial dystrophy of the cornea. *Am. J. Ophthalmol.*, **45**, 518–535.
6. Laing,R.A., Leibowitz,H.M., Oak,S.S., Chang,R., Berrospi,A.R. and Theodore,J. (1981) Endothelial mosaic in Fuchs' dystrophy. A qualitative evaluation with the specular microscope. *Arch. Ophthalmol.*, **99**, 80–83.
7. Bigar,F. (1982) Specular microscopy of the corneal endothelium. Optical solutions and clinical results. *Dev. Ophthalmol.*, **6**, 1–94.
8. Borderie,V.M., Baudrimont,M., Vallee,A., Ereau,T.L., Gray,F. and Laroche,L. (2000) Corneal endothelial cell apoptosis in patients with Fuchs' dystrophy. *Invest. Ophthalmol. Visual Sci.*, **41**, 2501–2505.
9. Li,Q.J., Ashraf,M.F., Shen,D.F., Green,W.R., Stark,W.J., Chan,C.C. and O'Brien,T.P. (2001) The role of apoptosis in the pathogenesis of Fuchs endothelial dystrophy of the cornea. *Arch. Ophthalmol.*, **119**, 1597–1604.
10. Matthaei,M., Zhu,A.Y., Kallay,L., Eberhart,C.G., Cursiefen,C. and Jun,A.S. (2014) Transcript profile of cellular senescence-related genes in fuchs endothelial corneal dystrophy. *Exp. Eye Res.*, **129**, 13–17.
11. Jurkunas,U.V., Bitar,M.S., Funaki,T. and Azizi,B. (2010) Evidence of oxidative stress in the pathogenesis of fuchs endothelial corneal dystrophy. *Am. J. Pathol.*, **177**, 2278–2289.
12. Lorenzetti,D.W., Uotila,M.H., Parikh,N. and Kaufman,H.E. (1967) Central cornea guttata. Incidence in the general population. *Am. J. Ophthalmol.*, **64**, 1155–1158.
13. Sarnicola,C., Farooq,A.V. and Colby,K. (2018) Fuchs Endothelial Corneal Dystrophy: update on pathogenesis and future directions. *Contact Lens*, **45**, 1–10.
14. Gain,P., Jullienne,R., He,Z., Aldossary,M., Acquart,S., Cognasse,F. and Thuret,G. (2016) Global survey of corneal transplantation and eye banking. *JAMA Opthamol.*, **134**, 167–173.
15. Wieben,E.D., Aleff,R.A., Tosakulwong,N., Butz,M.L., Highsmith,W.E., Edwards,A.O. and Baratz,K.H. (2012) A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One*, **7**, e49083.
16. Mootha,V.V., Gong,X., Ku,H.C. and Xing,C. (2014) Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. *Invest. Ophthalmol. Vis. Sci.*, **55**, 33–42.
17. Xing,C., Gong,X., Hussain,I., Khor,C.C., Tan,D.T., Aung,T., Mehta,J.S., Vithana,E.N. and Mootha,V.V. (2014) Transethnic replication of association of CTG18.1 repeat expansion of TCF4 gene with Fuchs' corneal dystrophy in Chinese implies common causal variant. *Invest. Ophthalmol. Vis. Sci.*, **55**, 7073–7078.
18. Soliman,A.Z., Xing,C., Radwan,S.H., Gong,X. and Mootha,V.V. (2015) Correlation of severity of fuchs endothelial corneal dystrophy with triplet repeat expansion in TCF4. *JAMA Ophthalmol.*, **133**, 1386–1391.
19. Gattey,D., Zhu,A.Y., Stagner,A., Terry,M.A. and Jun,A.S. (2014) Fuchs endothelial corneal dystrophy in patients with myotonic dystrophy: a case series. *Cornea*, **33**, 96–98.
20. Mootha,V.V., Hansen,B., Rong,Z., Mammen,PP., Zhou,Z., Xing,C. and Gong,X. (2017) Fuchs' endothelial corneal dystrophy and RNA foci in patients with myotonic dystrophy. *Invest. Ophthalmol. Vis. Sci.*, **58**, 4579–4585.
21. Winkler,N.S., Milone,M., Martinez-Thompson,J.M., Raja,H., Aleff,R.A., Patel,S.V., Wieben,E.D. and Baratz,K.H. (2018) Fuch's corneal endothelial dystrophy in patients with myotonic dystrophy, type 1. *Invest. Opthamol. Vis. Sci.*, **59**, 3053–3057.
22. Deng,S.X., Lee,W.B., Hammersmith,K.M., Kuo,A.N., Li,J.Y., Shen,J.F., Weikert,M.P. and Shtein,R.M. (2018). Descemet membrane endothelial Keratoplasty: Safety and Outcomes: A report by the American academy of ophthalmology. *Ophthalmol.*, **125**, 295–310.
23. Rohilla,K.J. and Gagnon,K.T. (2017). RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropthologica Comm.*, **5**, 63.
24. Krachmer,J.H., Purcell,J.J., Young,C.W. and Bucher,K.D. (1978) Corneal endothelial dystrophy. A study of 64 families. *Arch. Ophthalmol.*, **96**, 2036–2039.
25. Mootha,V.V., Hussain,I., Cunnusamy,K., Graham,E., Gong,X., Neelam,S., Xing,C., Kittler,R. and Petroll,W.M. (2015) TCF4 triplet repeat expansion and nuclear RNA foci in Fuchs' endothelial corneal dystrophy. *Invest. Opthamol. Vis. Sci.*, **56**, 2003–2011.
26. Hu,J., Rong,Z., Gong,X., Zhou,Z., Sharma,V.K., Xing,C., Watts,J.K., Corey,D.R. and Mootha,V.V. (2018) Oligonucleotides targeting TCF4 triplet repeat expansion inhibit RNA foci and mis-splicing in Fuchs' dystrophy. *Hum. Mol. Genet.*, **27**, 1015–1026.
27. Wieben,E.D., Aleff,R.A., Tang,X., Butz,M.L., Kalari,K.R., Highsmith,E.W., Jen,J., Vasmatzis,G., Patel,S.V., Maguire,L.J. *et al.* (2017) 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.*, **58**, 343–352.
28. Kino,Y., Mori,D., Oma,Y., Taeshita,Y., Sasagawa,N. and Ishiura,S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495–507.
29. Dansithong,W., Paul,S., Comai,L. and Reddy,S. (2015) MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J. Biol. Chem.*, **280**, 5773–5780.
30. Teplova,M. and Patel,D. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struc. Mol. Biol.*, **15**, 1343–1351.

31. Konieczny,P., Stepniak-Konieszna,E. and Sobczak,K. (2014) MNBL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.*, **42**, 10873–10877.

32. Du,J., Aleff,R.A., Soragni,E., Kalari,K., Nie,J., Tang,X., Davila,J., Kocher,J-P., Patel,S.V., Gottesfeld,J.M. *et al.* (2015) RNA toxicity and missplicing in the common eye disease Fuchs' Corneal endothelial dystrophy. *J. Biol. Chem.*, **290**, 5079–5990.

33. Cui,Z., Zeng,Q., Guo,Y., Liu,S., Wang,P., Xie,M. and Chen,J. (2018) Pathological molecular mechanism of symptomatic late-onset Fuchs endothelial corneal dystrophy by bioinformatic analysis. *PLoS One*, **13**, e0197750.

34. Kanadia,R.N., Shin,J., Yuan,Y., Beattie,S.G., Wheeler,T.M., Thornton,C.A. and Swanson,M.S. (2006) Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 11748–11753.

35. Wang,E.T., Cody,N.A., Jog,S., Biancolella,M., Wang,T.T., Treacy,D.J., Luo,S., Schroth,G.P., Housman,D.E., Reddy,S. *et al.* (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, **150**, 170–724.

36. Wieben,E.D., Baratz,K.H., Aleff,R.A., Kalari,K.R., Tang,X., Maguire,L.J., Patel,S.V. and Fautsch,M.P. (2019) Gene expression and missplicing in the corneal endothelium of patients with TCF4 trinucleotide repeat expansion without Fuchs' dystrophy. *Invest. Opthamol. Vis. Sci.*, **60**, 3636–3643.

37. Wieben,E.D., Aleff,R.A., Tang,X., Kalari,K.R., Maguire,L.J., Patel,S.V., Baratz,K.H. and Fautsch,M.P. (2018) Gene expression in the corneal endothelium of Fuchs endothelial corneal dystrophy patients with and without expansion of the trinucleotide repeat in TCF4. *PLoS One*, **13**, e021000005.

38. Robertson,N.G., Resendes,B.L., Lin,J.S., Aster,J.C., Adams,J.C. and Morton,C.C. (2001) Inner ear localization of the mRNA and protein products of COCH, mutated in the sensorineural deafness and vestibular disorder, DFNA9. *Hum. Mol. Genet.*, **10**, 2493–2500.

39. Bhattacharya,S.K., Rockwood,E.J., Smith,S.D., Bonilha,V.L., Crabb,J.S., Kuchtey,R.W., Roberston,N.G., Peachery,N.S., Morton,C.C. and Crabb,J.W. (2005) Proteomics reveal cochlin deposits associated with glaucomatous travecular meshwork. *J. Biol. Chem.*, **280**, 6080–6084.

40. Singh,P., Carraher,C. and Schwarzbauer,J.E. (2010) Assembly of fibronectin extracellullar matrix. *Annu. Rev. Cell Dev. Biol.*, **26**, 397–419.

41. Ullrich-Murphy,J.E. and Suto,M.J. (2018) Thrombospondin-1 regulation of latent TGF-b activation: a therapeutic target for fibrotic disease. *Matrix Biol.*, **68–69**, 28–43.

42. Leask,A. and Abraham,D.J. (2004) TGF-b signaling and the fibrotic response. *FASEB J.*, **18**, 816–827.

43. Weller,J.M., Zenkel,M., Schlotzer-Schrehardt,U., Bachmann,B.O., Tourtas,T. and Kruse,F.E. (2014) Extracellular matrix alterations in late-onset Fuchs' corneal dystrophy. *Invest. Opthamo.l Vis. Sci.*, **55**, 3700–3708.

44. Karin,D., Koyama,Y., Brenner,D. and Kisseleva,T. (2016) The characteristics of activated fibrobasts/myofibroblasts in liver fibrosis. *Differentiation*, **92**, 84–92.

45. De Roo,A.K., Wouters,J., Govaere,O., Foets,B and van den Oord,J.J. (2017) Identification of circulating fibrocytes and dendritic derivatives in corneal endothelium of patients with Fuchs' dystrophy. *Invest. Opthamol. Vis. Sci.*, **58**, 670–681.

46. Gudde,E.E.G., Gonzalez-Barriga,A., van der Broek,W.J.A.A, Wieringa,B. and Wansink,D.G. (2016) A low absolute number of expanded transcripts is involved in mytonic dystrophy type 1 manifestation in muscle. *Hum. Mol. Genet.*, **25**, 1648–1662.

47. Hu,J., Liu,J., Li,L., Gagnon,K.T. and Corey,D.R. (2015) Targeted recognition of GGGGCC/CCCCGG repeats at the C9orf72 locus by duplex RNA. *Chem. Biol.*, **22**, 1505–1511.

48. Sznajder,L.J., Thomas,J.D., Carrell,E.M., Reid,T., McFarland,K.N., Cleary,J.D., Oliveira,R., Nutter,C.A., Bhatt,K., Sobczak,K. *et al.* (2018) Intron retention induced by microsatellite expansions as a disease biomarker. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 4234–4239.

49. Sobczak,K., Wheeler,T.M., Wang,W. and Thornton,C.A. (2013) RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy. *Mol. Ther.*, **21**, 380–387.

50. Lee,J.E., Bennett,C.F. and Cooper,T.A. (2012) RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 4221–4226

51. Wojtkowiak-Szlachcic,A., Taylor,K., Stepniak-Konieczna,E., Sznajder,L.J., Mykowska,A., Sroka,J., Thornton,C.A. and Sobczak,K. (2015) Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy. *Nucleic Acids Res.*, **43**, 3318–3331.

52. Klein,A.F., Varela,M.A., Arandel,L., Holland,A., Naouar,N., Arzumanov,A., Seolane,D., Revillod,L., Bassez,G., Ferry,A. *et al.* (2019) Peptide-conjugated oligonucleotides evoke long-lasting myotonic dystrophy correction in patient-derived cells and mice. *J. Clin. Invest.*, **129**, 4739–4744.

53. Zarouchlioti,C., Sanchez-Pintado,B., Hafford Tear,N.J., Klein,P., Liskova,P., Dulla,K., Semo,M, Vugler,A.A., Muthusamy,K., Dudakova,L. *et al.* (2018) Antisense therapy for a common corneal dystrophy ameliorates TCF4 repeat Expansion-Mediated toxicity. *Am. J. Hum. Genet.*, **102**, 528–539.

54. Rong,Z., Hu,J., Chu,Y., Corey,D.R. and Mootha,V.V. (2018) Depletion of muscleblind-like protein by expanded CUG repeat RNA is unlikely to contribute to disease pathogenesis in Fuchs Dystrophy. *Invest. Opthamol. Vis. Sci.*, **60**, 3980–3991.

55. Wang,J., Harris,A., Prendes,M.A., Alshawa,L., Gross,J.C., Wentz,S.M., Rao,A.B., Kim,N.J., Snyder,A. and Siesky,B. (2017) Targeting transforming growth factor-β signaling in primary open-angle glaucoma. *J. Glaucoma*, **26**, 390–395.

56. Nagarsheth,M., Singh,A., Schotzer,B., Binbineau,D.C., Sugar,J., Lee,W.B. and Iyengar,S.K. (2012) Fuchs' Multi-Center Study Group Relationship between Fuchs endothelial corneal dystrophy severity and glaucoma and/or ocular hypertension. *Arch. Opthamol.*, **130**, 1384–1388.