

Tissue-specific *TCF4* triplet repeat instability revealed by optical genome mapping



Christina Zarouchlioti,^a Stephanie Efthymiou,^b Stefano Facchini,^b Natalia Dominik,^b Nihar Bhattacharyya,^a Siyin Liu,^{a,c} Marcos Abreu Costa,^a Anita Szabo,^a Amanda N. Sadan,^a Albert S. Jun,^d Enrico Bugiardini,^b Henry Houlden,^b Andrea Cortese,^b Pavlina Skalicka,^{e,f} Lubica Dudakova,^f Kirithika Muthusamy,^c Michael E. Cheetham,^a Alison J. Hardcastle,^a Petra Liskova,^{e,f} Stephen J. Tuft,^{a,c} and Alice E. Davidson^{a,c,*}



^aUCL Institute of Ophthalmology, London, UK

^bUCL Queen Square Institute of Neurology, Department of Neuromuscular Diseases, London, UK

^cMoorfields Eye Hospital, London, UK

^dCornea, Cataract, and External Disease Division, Wilmer Eye Institute, Johns Hopkins Medicine, Baltimore, USA

^eDepartment of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

^fDepartment of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

Summary

Background Fuchs endothelial corneal dystrophy (FECD) is the most common repeat-mediated disease in humans. It exclusively affects corneal endothelial cells (CECs), with $\leq 81\%$ of cases associated with an intronic *TCF4* triplet repeat (CTG18.1). Here, we utilise optical genome mapping (OGM) to investigate CTG18.1 tissue-specific instability to gain mechanistic insights.

Methods We applied OGM to a diverse range of genomic DNAs (gDNAs) from patients with FECD and controls ($n = 43$); CECs, leukocytes and fibroblasts. A bioinformatics pipeline was developed to robustly interrogate CTG18.1-spanning DNA molecules. All results were compared with conventional polymerase chain reaction-based fragment analysis.

Findings Analysis of bio-samples revealed that expanded CTG18.1 alleles behave dynamically, regardless of cell-type origin. However, clusters of CTG18.1 molecules, encompassing ~ 1800 – $11,900$ repeats, were exclusively detected in diseased CECs from expansion-positive cases. Additionally, both progenitor allele size and age were found to influence the level of leukocyte-specific CTG18.1 instability.

Interpretation OGM is a powerful tool for analysing somatic instability of repeat loci and reveals here the extreme levels of CTG18.1 instability occurring within diseased CECs underpinning FECD pathophysiology, opening up new therapeutic avenues for FECD. Furthermore, these findings highlight the broader translational utility of FECD as a model for developing therapeutic strategies for rarer diseases similarly attributed to somatically unstable repeats.

Funding UK Research and Innovation, Moorfields Eye Charity, Fight for Sight, Medical Research Council, NIHR BRC at Moorfields Eye Hospital and UCL Institute of Ophthalmology, Grantová Agentura České Republiky, Univerzita Karlova v Praze, the National Brain Appeal's Innovation Fund and Rosetrees Trust.

Copyright © 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Fuchs endothelial corneal dystrophy; Trinucleotide repeat expansion disease; Triplet repeat expansion-mediated disease; Somatic mosaicism; Tissue-specific repeat instability; Optical genome mapping

Introduction

Fuchs endothelial corneal dystrophy (FECD) is a common age-related cause of heritable visual loss primarily affecting the corneal endothelium and is the leading indication for corneal transplantation in high-income

countries.^{1–3} The post-mitotic monolayer of corneal endothelial cells (CECs) maintains a leaky barrier that regulates corneal hydration.⁴ In FECD, accelerated CEC loss occurs, compromising barrier function and progressively resulting in corneal swelling, clouding and

*Corresponding author. UCL Institute of Ophthalmology, 11-43 Bath Street, London, UK.
E-mail address: alice.davidson@ucl.ac.uk (A.E. Davidson).

Research in context

Evidence before this study

Somatic instability of short tandem repeats (STRs) is recognised to be a key mechanism contributing to the tissue-specific, age-dependent and progressive nature of several neurological and neuromuscular repeat expansion diseases (Huntington's disease and myotonic dystrophies). Fuchs endothelial corneal dystrophy (FECD) is the most common repeat expansion-mediated disease reported to date in humans, predominantly attributed to expansion of a non-coding CTG element (CTG18.1) in the ubiquitously expressed gene, *TCF4*. At the onset of this study, the contribution of somatic instability to this common, tissue-specific and sight threatening disease was unknown.

Added value of this study

Our study provides insights into the tissue-specific nature of FECD. Utilising optical genome mapping (OGM) to obtain single-molecule resolution of CTG18.1, we detected extreme levels of somatic repeat instability exclusively in affected corneal endothelial cells (CECs) from expansion positive individuals. Furthermore, analysis of leukocyte gDNA revealed

that both progenitor allele size and age influence somatic instability rates at this locus. More generally, these data also exemplify OGM as a powerful tool to explore somatic instability of STR loci across the human genome, especially in instances whereby patient-derived gDNA availability from affected tissues is limited, and/or repeats are large and unstable, given OGM imposes no size detection limit.

Implications of all the available evidence

These findings reveal that CEC-specific somatic instability is a pivotal prerequisite of downstream pathogenic events elicited by CTG18.1 expansions. Thus, our data shed insights into the molecular mechanisms of this disease and, in-part, explain its tissue-specific nature. This knowledge is anticipated to inform the development of new therapies for this common cause of age-related visual loss. Furthermore, these findings now highlight the broader relevance of FECD as a translationally relevant model system to the growing number of rarer and currently incurable repeat-mediated human diseases similarly underpinned by somatic instability mechanisms.

reduced visual acuity.⁵ To date, expansion of an intronic triplet repeat within the *TCF4* gene (termed CTG18.1; OMIM #613267) has been identified as the most common risk factor for FECD in all ethnic groups studied.⁶ Remarkably, such expansions are detected in up to 81% of European patients with FECD, making it by far the most common short tandem repeat (STR) expansion disease in humans.^{6,7} Thus, as well as being a common cause of age-related vision loss, FECD represents an important paradigm for much rarer, currently incurable and devastating STR diseases, such as Huntington's disease (OMIM #143100) and myotonic dystrophy (OMIM #160900; 602668).

STRs are somatically unstable elements that can contract and expand in an age-, repeat length- and tissue-specific manner.^{8–10} We have previously utilised a long-read amplification-free sequencing method to determine that expanded copies of CTG18.1 (defined as ≥ 50 repeats) are somatically unstable in peripheral blood leukocytes, with progenitor repeat lengths positively correlating with increased CTG18.1 instability.¹¹ Southern blot data of patient-derived corneal endothelial cell cultures have also suggested that CTG18.1 repeat length is greater in affected cells compared to leukocytes, in keeping with the dogma that STRs typically display high levels of instability within affected cell types.^{12,13} For example, in myotonic dystrophy type 1, the disease-associated repeat has been reported to expand up to 4000 repeats in skeletal muscle, representing up to a 25-fold increase in length compared to blood.¹⁴ Multiomic data studies of Huntington's disease have recently brought disease-associated STR instability mechanisms

to the forefront of translational genomic medicine, revealing that repeat-length mosaicism within affected cell populations is a fundamental driver of disease.¹⁵

Disease-associated STR instability is hypothesised to be a unifying mechanism and translationally relevant pathway for many STR-associated diseases.¹⁶ However, it can be extremely challenging to measure repeat length mosaicism in affected cell types, given the rarity of relevant material and the technical difficulties associated with accurately sizing or sequencing large repetitive and unstable genomic elements. Conventionally, STR length is assessed by gel or capillary electrophoresis of PCR amplicons and/or Southern blot.¹⁷ More recently, short-read next-generation sequencing (NGS) approaches have been developed to quantify levels of repeat length instability.¹⁸ Amplification-free long-read sequencing methods have also yielded novel insights but require high levels of input DNA.^{19,20} Thus, collectively, these approaches are either restricted by repeat size detection thresholds, resolution or the large inputs of DNA required, which are scarcely available from affected patient-derived tissues.

Optical Genome Mapping (OGM) offers a powerful method to interrogate native ultra-long molecules of DNA (>150 kb) to reveal large and/or complex structural variants (>500 nucleotides) across the genome.²¹ To date, it has been suggested as a viable way to detect STR expansions and contractions associated with a subset of neurological and neuromuscular conditions.^{22–25} Despite *TCF4* being ubiquitously expressed, CTG18.1 expansions have only been robustly associated with FECD, a corneal endothelial cell-specific disease. This

knowledge, in combination with increased levels of somatic instability reported in other disease-associated STR loci within affected cell populations, prompted us to explore CTG18.1 length and instability in patient-derived CECs.

Here, we apply OGM to enable single molecule resolution of CTG18.1 in a series of genomic DNA samples from expansion-positive individuals affected by FECD. Utilising gDNA samples isolated from unaffected peripheral blood leukocytes and affected corneal endothelial cells, we explore the utility of this method to size the repeat and, importantly, to gain insights into the tissue-specific nature of this common and sight-threatening disease.

Methods

Subject recruitment and bio-sample processing

All FECD cases recruited to this study had clinical signs of disease at the time of bio-sampling and displayed similarly advanced disease (Krachmer Grade 5). The diseased corneal endothelium attached to the Descemet membrane was removed and collected as part of posterior corneal transplantation surgery to alleviate symptoms in individuals with FECD. Control corneal endothelium was similarly isolated from donor corneoscleral discs rejected for clinical transplantation (Miracles in Sight Eye Bank, Texas, USA). Excised tissues were stored in Leibovitz L-15 media (Life Technologies) or Eusol-C (Alchimia) until processing. Karyotypic sex of all participants in the study was determined based on the OGM data.

Ethics

This study was approved by the Research Ethics Committees of University College London (UCL) (22/EE/0090) or the General University Hospital (GUH) Prague (2/19 GACR) and was conducted in accordance/compliance with the Declaration of Helsinki. Written informed consent form was obtained and archived from all participants who provided peripheral blood, skin biopsies and/or corneal endothelial samples.

Primary corneal endothelial cell (CEC) culture

CECs were isolated and cultured as described by Zarouchlioti et al., 2018, originally adapted from Peh et al., 2015.²⁶ After excision of the Descemet membrane with the attached endothelium, the tissue was incubated in 0.2% collagenase type I for 2–4 h to digest the Descemet membrane and release the CECs. Following digestion, the CECs were pelleted and resuspended in stabilisation media (Human Endothelial-SFM (Life Technologies) supplemented with 5% foetal bovine serum (FBS), 1% antibiotic/antimycotic, and 10 μ M ROCK inhibitor Y-27632 (AdooQ BioSciences)). Cells were seeded onto FNC-coated tissue culture plates (Strattech) with stabilisation media. The following day, cell cultures were

switched to expansion medium (Ham's F-12 Nutrient Mix with GlutaMAX Supplement (Life Technologies)/Medium 199 GlutaMAX Supplement (Life Technologies) (1:1), 20 μ g/mL ascorbic acid, 1% insulin-transferrin-selenium (Life Technologies), 5% FBS, 1% antibiotic/antimycotic, 10 ng/mL bFGF (Life Technologies) and 10 μ M ROCK inhibitor Y-27632 (AdooQ BioSciences)). Cells were cultured at 37 °C, 5% CO₂ with media changes twice weekly. Cells were passaged once to expand the cell lines and were collected when they reached maximum confluency.

F35T cell culture

F35T is an immortalised human corneal endothelial line (RRID: CVCL_E2X2), originally isolated from a 62-year-old female with FECD.^{27,28} F35Ts were cultured on FNC-coated tissue culture vessels with expansion media (recipe described in the previous section). F35T cells were passaged using 1X TrypLE Express (Life Technologies) and the media was changed three times per week. F35Ts were cultured in a controlled environment of 37 °C and 5% CO₂. Cells were authenticated by GlobalFiler™ IQC PCR Amplification kit (Thermo Fisher Scientific) and were routinely tested for mycoplasma using MycoStrip™ (InvivoGen).

Fibroblast cell culture

Fibroblast cell lines isolated from dermal skin biopsies were cultured in DMEM/F-12, GlutaMAX (Life Technologies) supplemented with 10% FBS and 1% penicillin/streptomycin. Fibroblast cell cultures were maintained at 37 °C, 5% CO₂. Fresh medium was added every 2–3 days and cells were passaged using 1X TrypLE Express (Life Technologies).

Dual CTG18.1 targeted genotyping

To stratify participants in the study, a dual PCR-based approach was applied to genotype CTG18.1 and determine progenitor allele size. Genomic DNA was extracted from peripheral blood samples from individuals with FECD (QIAGEN Gentra Puregene Blood Kit) or from the scleras of the control corneoscleral discs (QIAGEN Blood and Tissue kit) in accordance with the manufacturer's guidelines.

The CTG18.1 repeat length was determined using a dual PCR-based approach adapted from Weiben E et al. 2012.²⁹ In brief, to detect and size CTG18.1 repeat length STR PCR using a 6'-FAM-conjugated primer (5'-CAGATGAGTTTGGTGTAAGAT-3') upstream of the repeat and an unlabeled primer (5'-ACAAGCA-GAAAGGGGGCTGCAA-3') downstream was performed. A further triplet-primed (TP) PCR assay to confirm the presence or absence of a CTG18.1 allele above the detection limit of the STR assay (approximately 125 repeats), was also performed utilising a 5'-6-FAM-conjugated primer (5'-AATCCAAACCGCCTTCCAAGT-3') upstream of the repeat in combination with

a reverse primer complementary to the repeat sequence and encompassing a common 5' sequence (tail) that shares no homology to human genomic sequence (5'-TACGCATCCCAGTTTGA-GACGCAGCAGCAGCAGCAG-3') to serve as an anchor for a second reverse primer (5'-TACGCATCC-CAGTTTGA-GACGCAG-3'). All PCR amplicons were mixed with the GeneMarker ROX500 ladder (Thermo Fisher Scientific). Capillary electrophoresis was performed on ABI 3730 (Applied Biosystems) and CTG18.1 alleles were sized using GeneMarker (SoftGenetics).

OGM: sample preparation, cryopreservation and workflow

Fresh blood was collected from all participants before being stored at -80°C . CEC cultures were collected and cryopreserved in 10% DMSO in liquid nitrogen until gDNA extraction. UHMW gDNA was extracted using the "SP or SP-G2 Blood and Cell Culture DNA Isolation" kits (Bionano Genomics) along with the recommended extraction protocols for either frozen blood or cryopreserved cell samples. After homogenisation, the gDNA was labelled with fluorophores at genome-wide DLE-1 recognition motifs, while the DNA backbone was also stained using the "Direct Label and Stain (DLS or DLS-G2)" kits (Bionano Genomics). Stained UHMW DNA molecules were loaded into Saphyr G2.3 or G3.3 chips and single linearised molecules running through the nanochannels were imaged using the Saphyr instrument (Bionano Genomics). High-resolution images were acquired for a throughput of 1.5 TB per sample and a minimum expected effective coverage of 400X per sample. As per Bionano's recommendations, all samples had 14-17 labels/100 kb, >85 map rates, $N50 \geq 150$ kb ≥ 230 kb, $N50 \geq 20$ kb ≥ 150 kb.

Customised OGM analysis pipeline to estimate CTG18.1 repeat length

Molecules were aligned to the hg38 reference using the `align_mol_to_ref.py` script available in Bionano Solve 3.7.1 software package (<https://bionano.com/software-downloads/>). CTG18.1 is located between markers 10,414 and 10,415 of chromosome 18 (hg38; chr18:55,584,360–55,594,648). From the alignment files, all molecules overlapping both markers were selected for the downstream analysis pipeline. For each molecule, the expansion size is estimated by calculating the distance difference between the two markers of interest in relation to the reference. The distance between markers 10,414 and 10,415 is 10,288bp. However, a correction was to be applied in this case: when two theoretical binding sites for the labelling enzyme are too close (around 1 kb or less), only one of the two is detected (randomly chosen), meaning that when comparing the molecule with the reference, the average position of these two reference markers must be used. The labelling pattern of our region of interest was also

checked in the telomere-to-telomere hg38, but since it was identical, we proceeded with the correction as shown in [Figure S1](#). We applied the same correction scheme used by the standard Bionano DeNovo pipeline: keeping 10,414 (chr18:55,584,360) as the upstream marker and averaging markers 10,415 (chr18:55,594,648) and 10,416 (chr18:55,595,438) for the downstream marker, which gives a corrected reference distance of 10,683bp. Finally, to correct for the presence of 24 CTG repeats in hg38, we subtracted 72 bp more resulting in a final reference distance of 10,611bp. For the leukocytes and CEC data series, molecules were plotted based on their size in histograms using 200 bp bin widths. The code to reproduce the analyses described here is available from GitHub (https://github.com/stfacc/extract_gaussian_alleles/blob/main/aln.py).

Statistical analysis

Significant deviation from normality was observed for mean OGM molecule values (Anderson-Darling test $p < 0.001$) and for that reason only non-parametric tests were used in the study. Spearman correlation analysis was used to investigate the relationship between the largest progenitor allele size and the measured molecule size by OGM. Simple and multiple linear regression models using the log-transformed outcome were employed to model the effect of mean progenitor allele size per patient on the mean measured molecule size per patient. The model optimization steps are provided in the results. The final model included an interaction term, and the independent variables were adjusted using mean centring, which eliminated the multicollinearity. We used the Mann-Whitney test to compare the mean measured molecule size per patient between the expansion-positive and expansion-negative subgroups. Ancestry was not considered in the models due to the overall sample size limitation and the risk of overfitting.

Role of funders

None of the funders had a role in the study design, data collection or data analyses.

Results

Application of OGM to interrogate CTG18.1 repeat expansions

CTG18.1 is typically sized from peripheral blood leukocyte gDNA samples using a dual PCR-based approach that detects and sizes PCR amplicons by capillary electrophoresis.^{12,30} This protocol consists of an STR PCR-based assay to size CTG18.1 alleles of ≤ 125 CTG repeats ([Fig. 1A i,iii,v](#)) and a triplet-primed PCR assay (TP-PCR) to detect the presence or absence of expanded alleles, including those beyond the detection limit of the STR-PCR assay (>125 repeats) ([Fig. 1A](#)

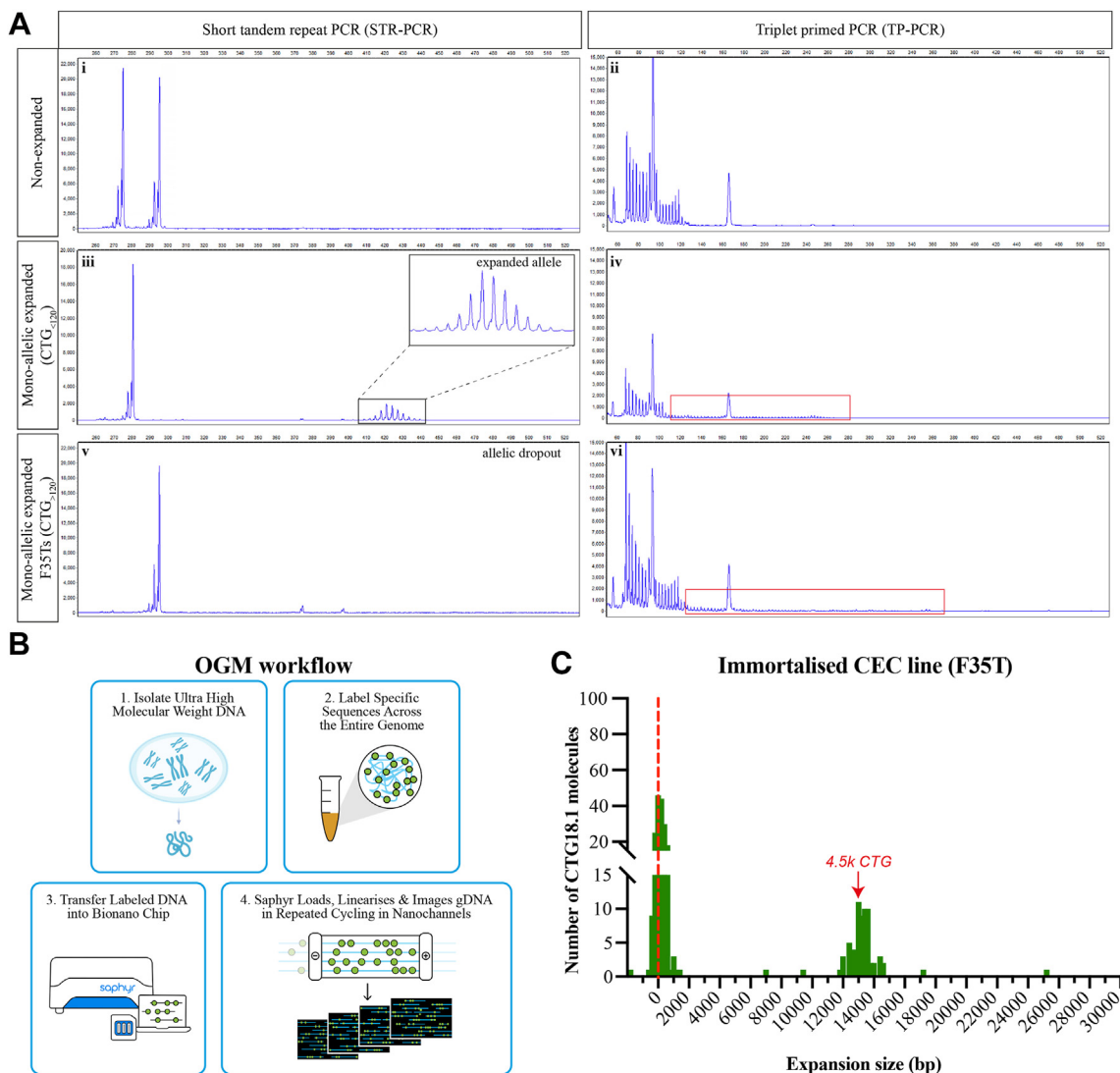


Fig. 1: Optical genome mapping (OGM) effectively detects large expanded CTG18.1 alleles. (A) Detected traces after capillary electrophoresis of STR-PCR (i, iii, v) and TP-PCR (ii, iv, vi) products amplified from non-expanded whole-blood derived gDNA samples (i-ii), mono-allelic expanded whole-blood derived (iii-iv) and F35T cell-derived (v-vi) gDNA samples. Red boxes highlight the presence of expanded alleles as indicated by TP-PCR traces. (B) Schematic summary of OGM methodology; (1) extraction of ultra-high molecular weight (UHMW) gDNA that is (2) subsequently labelled via covalent modification at genome-wide CTTAAG hexamer motifs before (3) linearising and imaging the decorated molecules on nanochannels (image adapted from: <https://bionanogenomics.com>). (C) Histogram of OGM CTG18.1 molecule sizes (bp) observed in immortalised CEC line F35T. The red dotted line indicates alleles around the lowest detection threshold of the method, likely representing the non-expanded allele.

ii,iv,vi). Despite proving an efficient and precise way to genotype leukocyte-derived CTG18.1 alleles, the upper detection limit of the STR assay precluded sizing of 3.3% (14/427) expanded CTG18.1 alleles reported in our DNA bio-resource of patients with FECD ($n = 450$).⁷

Given that OGM can detect very large structural variants, we aimed to investigate the utility of this method to interrogate CTG18.1 and explore the capabilities of this assay to reveal ultra-long CTG18.1

expanded alleles. As a positive control, we selected the F35T immortalised corneal endothelial cell line as it has previously been reported by Southern blot to have a monoallelic expansion of ~4500 CTG repeats.²⁷ When analysed with the STR assay, only one CTG18.1 allele harbouring 21 CTG repeats was detected (i.e., allelic dropout is observed). In contrast, TP-PCR confirmed the presence of a longer expanded allele beyond the detection threshold of the STR assay (Fig. 1A v-vi).

Ultra-high molecular weight (UHMW) gDNA was extracted from F35T cells and analysed by OGM (Fig. 1B). After correcting for the flanking region between the labels of interest and the 24 CTG repeats in hg38, the size of the CTG18.1 interval was determined for each individual DNA molecule (Fig. 1C). An accumulation of molecules was detected around 0 bp, likely representing the shorter non-expanded allele, as this is below the reported lower detection limit of this assay (± 500 bp). Importantly, a cluster of 13,500 bp molecules was also detected, corresponding to approximately 4500 CTG repeats, replicating previously reported Southern blot data.²⁷ A small number of additional DNA molecules was detected above (up to 25.2 kb long) and below (single molecule at -1.8 kb) the expected size range, based on the dual PCR assay and Southern blot data. These single non-clustering molecules were deemed likely to be artefacts arising from the misalignment of long repetitive regions with relatively sparse fluorescent labels.³¹ Nonetheless, the detection of molecules clustering at approximately 13,600 bp (representing 4500 repeats) indicated the presence of a large CTG18.1 allele that evaded sizing by the dual PCR-based genotyping assay and was concordant with Southern blot data, thus supporting the utility of OGM to detect the presence of large expanded CTG18.1 alleles (Table 1).

CTG18.1 displays tissue-specific somatic instability

In the knowledge that OGM can effectively detect and size a large expanded CTG18.1 allele in the immortalised F35T cell line, we next wanted to explore CTG18.1 length and instability levels in the primarily affected cell population. Given that patients with FECD typically only undergo corneal transplantation when there is advanced disease, when the CEC density has substantially declined, total cell number and gDNA yields from these specimens are extremely low. To overcome this limitation and acquire enough cells for the OGM protocol (approximately 1 million cells per sample), we used primary CEC cultures generated from endothelial keratoplasty specimens removed following planned corneal surgery to analyse CEC-specific gDNA. We have previously demonstrated that these primary CEC cultures robustly maintain the biomarkers of CTG18.1-mediated disease and generate a pure corneal endothelial cell model that enables the investigation of FECD in a disease-relevant cell context.^{7,32}

We acquired both corneal endothelial specimens and peripheral blood leukocytes from 9 patients with FECD to directly compare instability levels between the affected and unaffected cell types within patients. Samples were selected to reflect the overall range of CTG18.1 expanded alleles observed in our FECD patient cohort.⁷ UHMW genomic DNA from leukocytes and CEC samples was isolated using the recommended Bionano protocols. Initially, all samples were analysed using the dual PCR-based genotyping method (Table 1).

For all leukocyte-derived samples, two alleles were detected, one non-expanded and one expanded (<125 CTG repeats) allele. However, similar to the F35Ts, only one non-expanded allele was detected by STR-PCR in all CEC samples for subjects 1–9. Thus, allelic dropout of larger alleles beyond the sizing threshold of the STR assay consistently occurred, while TP-PCR indicated the presence of an expanded allele in all expansion-positive samples (Table 1), as previously observed with the F35T cell line (Fig. 1A).

Next, gDNA samples were analysed by OGM (Fig. 1B; Fig. 2). The mean detectable repeat size in the leukocyte-derived gDNA sample series was approximately 97 CTG repeats (range 54–160) (Fig. 2). These figures reflect the number of molecules detected from the expanded and non-expanded alleles (unphased molecules) within each of the nine samples. As most of the molecules from the respective samples are below the lower detection threshold of the assay (500 bp) they could not be sized with confidence. However, when comparing the mean molecule size of each of the nine leukocyte samples to their respective affected CEC sample from the same individual, significantly higher levels of CTG18.1 repeat instability were observed in all CEC samples (Wilcoxon matched-pairs signed rank test, $p = 0.004$). Furthermore, molecules ranging from 5439–35,561 bp were only detected in the CEC-derived samples, equating to approximately 1800–11,900 repeats, with peaks of molecules ranging from 3900–6400 repeats (Fig. 2). In addition, we also investigated three FECD dermal fibroblast lines, generated from CTG18.1 expansion-positive cases, to further explore repeat instability within another unaffected cell type (Table S1). When analysed by OGM, CTG18.1 molecule sizes were comparable to leukocytes (mean repeat length ~ 97 CTG repeats) (Figure S2).

The vast majority of FECD cases harbour only one expanded allele (*e.g.*, all nine cases presented in Fig. 2). However, we acquired corneal endothelial tissue and subsequently established a primary CEC culture from one additional case with bi-allelic CTG18.1 expansions (baseline CTG18.1 expansion status of 57/93 by STR-analysis of leukocyte-derived DNA). OGM molecule distribution from this sample, alongside representative bi-allelic non-expanded and mono-allelic expanded CEC-derived gDNA FECD samples, are shown in Fig. 3. Although we could not phase the OGM data, these data illustrate how the zygosity status of the expansion influences molecule distributions. The mean molecule length for the bi-allelic expanded CEC sample (CEC-10) was 16,803bp, equating to approximately 5600 repeats (Table 1). This increased mean molecule length is driven by the shift in the ratio of total molecules ≤ 2000 bp versus >2000 bp compared to the mono-allelic expanded and expansion-negative FECD CECs. Molecules >2000 bp were exclusively detected in expansion-positive CECs, while these data suggest that the

Subject	Sex/ ancestry [#]	FECD diagnosis	Age of 1st keratoplasty	STR genotype	TP- PCR	Subject sampling age	Sample type	OGM Mean allele length (bp)	OGM Max allele length (bp)	Number of CTG18.1 molecules
Immortalised corneal endothelial cell (CEC) line										
F35T	F/EUR	✓		21/	+	62		3832	25,159	250
FECD subjects biopsied for both peripheral blood leukocytes (BL) and primary CEC-derived gDNA samples										
1	a	M/EUR	✓	61	16/63	+	62	BL-1	196	438
	b				16/	+	62	CEC-1	2771	254
2	a	F/EUR	✓	59	17/63	+	59	BL-2	162	328
	b				17/	+	59	CEC-2	2752	251
3	a	F/EUR	✓	75	31/75	+	76	BL-3	200	372
	b				31/	+	77	CEC-3	2946	245
4	a	M/EUR	✓	64	11/77	+	64	BL-4	216	365
	b				11/	+	64	CEC-4	2192	358
5	a	M/EUR	✓	83	27/81	+	83	BL-5	371	472
	b				27/	+	83	CEC-5	2827	246
6	a	F/EUR	✓	69	26/85	+	69	BL-6	282	396
	b				26/	+	69	CEC-6	3625	266
7	a	F/EUR	✓	59	11/90	+	53	BL-7	303	400
	b				11/	+	59	CEC-7	1977	279
8	a	F/EUR	✓	65	11/90	+	65	BL-8	481	373
	b				11/	+	66	CEC-8	2101	247
9	a	M/EUR	✓	58	25/96	+	58	BL-9	408	308
	b				25	+	58	CEC-9	2698	286
CTG18.1 expansion-positive peripheral blood leukocytes (BL)-derived gDNA samples										
10		M/EUR	✓	70	64/82	+	72	BL-10	383	432
11		M/EUR	✓	77	11/85	+	69	BL-11	291	358
12		M/EUR	✓	54	11/94	+	54	BL-12	334	376
13		F/EUR	✓	65	68/95	+	73	BL-13	590	360
14		M/EUR	✓	74	11/101	+	74	BL-14	1039	297
15		M/EUR	✓	67	17/107	+	67	BL-15	996	336
CTG18.1 expansion-negative peripheral blood leukocyte (BL-Ctr)-derived gDNA samples										
16		F/EUR	✗	N/A	11/11	-	57	BL-Ctr1	172	509
17		F/EUR	✗	N/A	11/16	-	29	BL-Ctr2	117	396
18		F/EUR	✗	N/A	11/16	-	50	BL-Ctr3	110	391
19		M/AFR	✓	92	14/21	-	92	BL-Ctr4	109	415
20		M/SAS	✗	N/A	11/22	-	22	BL-Ctr5	134	309
21		M/SAS	✗	N/A	17/22	-	40	BL-Ctr6	135	363
22		M/EUR	✗	N/A	17/23	-	67	BL-Ctr7	167	396
23		M/EAS	✗	N/A	23/24	-	48	BL-Ctr8	184	414
24		F/EUR	✗	N/A	17/32	-	69	BL-Ctr9	142	407
CTG18.1 expansion-negative corneal endothelial cell (CEC)-derived gDNA samples										
25		F/EUR	✓	60	11/14	-	60	CEC-Ctr1	111	470
26		F/EUR	✗	N/A	11/17	-	54	CEC-Ctr2	51	369
27		M/EUR	✗	N/A	15/17	-	70	CEC-Ctr3	164	399
28		F/EUR	✗	N/A	25/27	-	50	CEC-Ctr4	110	198
29		M/EUR	✗	N/A	11/32	-	74	CEC-Ctr5	97	351
30		M/EUR	✗	N/A	11/39	-	58	CEC-Ctr6	182	415
CTG18.1 bi-allelic corneal endothelial cell (CEC)-derived gDNA sample										
31		F/EUR	✓	77	57/93 ^a	+	77	CEC-10	16,803	87

Karyotypic sex, diagnosis, age of first corneal transplant surgery (as an indicator for disease severity), self-reported ancestry (#) and sampling age summarised for all subjects and samples of the study. STR: short-tandem repeat genotyping, M: male, F: female, EUR: European, AFR: African, SAS: South Asian, EAS: East Asian, BL: peripheral blood leukocyte sample, CEC: corneal endothelial cell sample, N/A: not applicable information. TP-PCR + and - indicates the presence or absence of an expanded allele. Subject sampling age shown for both blood collection and excised corneal endothelial tissue. ^aGenotyping data derived from subject's leukocyte-derived gDNA sample.

Table 1: Summary of subjects and gDNA samples analysed by optical genome mapping.

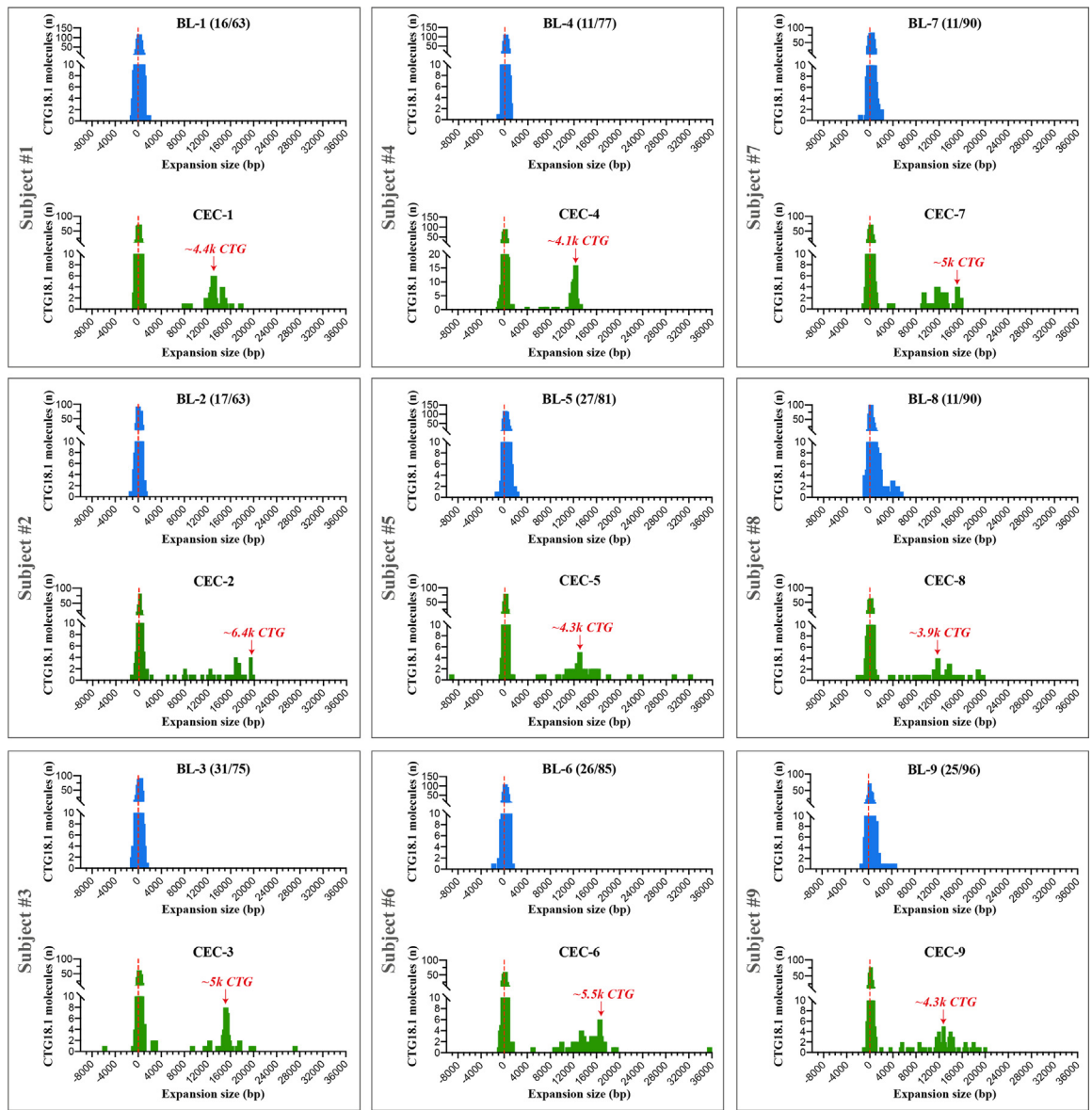


Fig. 2: Diseased corneal endothelial cells (CECs) display increased levels of CTG18.1 somatic instability compared to unaffected leukocytes. A series of peripheral blood leukocyte-derived (BL1-9 in blue) and corneal endothelial-derived (CEC1-9 in green) gDNA samples from nine unrelated FECD patients were analysed by OGM. Each grey box denotes samples from the same individual (subjects 1-9). In each instance, higher levels of somatic instability were detected in affected CECs compared to unaffected blood leukocyte-derived gDNA samples. The size (bp) of the CTG18.1 repeat-containing molecules is plotted (x-axis) against the total number of CTG18.1 molecules detected (y-axis). Red arrows depict the bin with most molecules detected above the 5439 bp threshold observed exclusively within the CEC-derived gDNA samples. Baseline CTG18.1 genotypes determined by STR-PCR analysis of leukocyte gDNA are shown in brackets, for each respective allele.

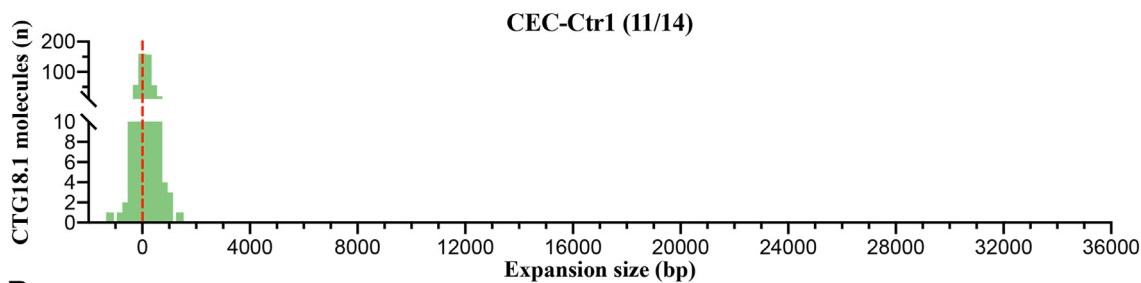
population of molecules ≤ 2000 bp are predominantly attributed to molecules derived from non-expanded CTG18.1 alleles (Fig. 3; Figure S3).

Progenitor repeat size and cellular context both influence CTG18.1 repeat instability

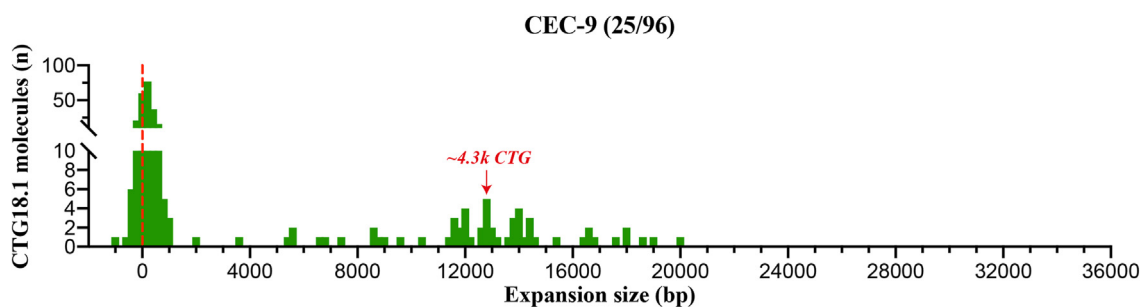
We have previously demonstrated that expanded CTG18.1 alleles are unstable in leukocytes, an effect that

positively correlates with the progenitor CTG18.1 allele length.¹¹ Interestingly, clusters of longer molecules were detected by OGM in samples BL-8 and BL-9, which also had some of the longest progenitor repeat sizes detected by STR-PCR within this series (Fig. 2). Hence, this led us to expand the series of leukocyte gDNA samples to give a more diverse range of progenitor expanded repeat sizes (63–107) to further test the capabilities of OGM to

A CTG18.1 expansion-negative



B Mono-allelic CTG18.1 expansion



C Bi-allelic CTG18.1 expansions

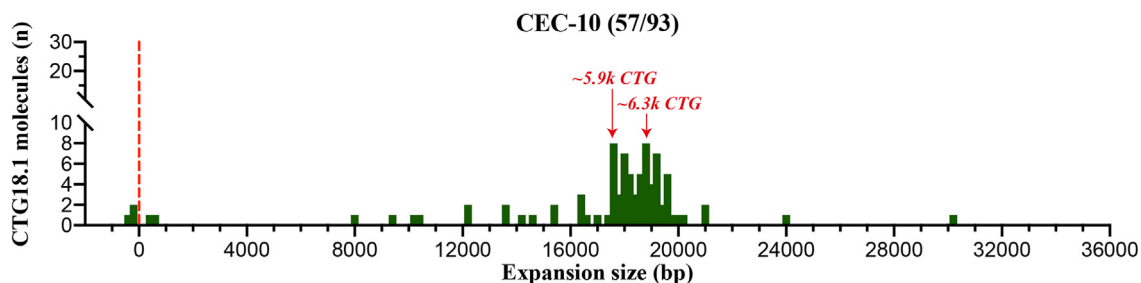


Fig. 3: Analysis of bi-allelic CTG18.1 corneal endothelial cells (CECs) indicates that molecules near the detection threshold of optical genome mapping (OGM) represent non-expanded alleles. CEC-derived gDNA samples from three unrelated FECD cases were analysed by OGM. Based on STR-analysis of leukocyte-derived DNA, we classified cases as CTG18.1 (A) expansion-negative (CEC-Ctr1), (B) mono-allelic expanded (CEC-9) or (C) bi-allelic expanded. Baseline STR genotyping results are shown in brackets for each respective allele. Red arrows depict the bin with most molecules detected >2000 bp threshold, exclusively observed in mono- (CEC-9; B) and bi-allelic (CEC-10; C) expanded CTG18.1 samples.

detect CTG18.1 somatic instability (Figure S4). Molecule counts were plotted by increasing size of CTG18.1 alleles based on the STR genotyping results (Fig. 4A). Spearman correlation analysis revealed a positive correlation between the largest progenitor CTG18.1 allele size (determined by STR-PCR of leukocyte gDNA) and the molecule size measured by OGM (correlation coefficient 0.255, 95% CI: 0.236–0.274) (Table S2). In agreement, a simple linear regression on the expansion-positive leukocyte OGM data suggests that the

progenitor allele size has a strong effect on the detected CTG18.1 molecule size. Given the log transformation of the outcome, the effect of the progenitor allele size variable on the OGM measured CTG18.1 molecule size increases exponentially, with increasing progenitor allele size (adjusted R^2 76.15%, $p < 0.0001$) (Fig. 4B; Table S3). Adding subject sampling age as an independent variable in the model improved the fit, suggesting that age contributes to the increased levels of somatic instability driven by the progenitor allele size

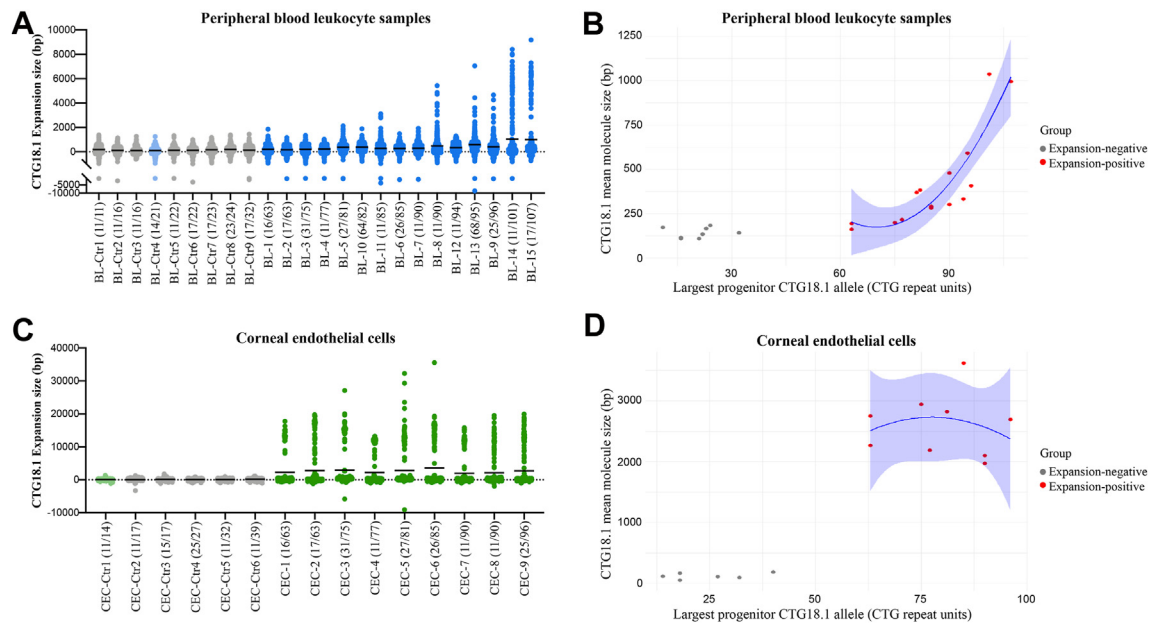


Fig. 4: Expanded CTG18.1 alleles behave dynamically in both peripheral blood leukocytes and corneal endothelial cells. Dot plots of molecules detected per individual capturing the distribution of all CTG18.1 alleles analysed in (A) peripheral blood leukocytes (BL; $n = 24$) and (C) corneal endothelial cell (CEC; $n = 15$) samples are presented. (A,C) Individuals with non-expanded CTG18.1 alleles (<50) are colour-coded in grey (Ctrl). Light blue and light green shades indicate FECD individuals with non-expanded CTG18.1 alleles in their peripheral blood leukocytes or CEC-derived gDNA, respectively. Lines represent the mean CTG18.1 expansion size in base pairs (bp) per sample. (B,D) Scatter plots of CTG18.1 mean molecule size (bp) against the largest progenitor allele is shown with polynomial regression for the expansion-positive dataset (red) and 95% confidence interval (CI) shown in blue for (B) peripheral blood leukocyte and (D) CEC sample series. All data series are arranged in order of the largest allele detected per individual according to baseline CTG18.1 genotype determined by the STR-PCR assay and indicated in brackets for each sample.

(adjusted R^2 91.02%, $p < 0.0001$, Table S3). Adding sex as an independent variable was found to have no effect on the OGM measured CTG18.1 molecule size.

Next, we wanted to explore if the dynamic and extreme range of CTG18.1 length instability observed in the corneal endothelium is driven by the presence of expanded progenitor alleles (≥ 50 CTG repeats, as detected in peripheral blood leukocyte samples) or if it is a phenomenon observed in all endothelial cells irrespective of their baseline CTG18.1 expansion status (Fig. 4C; Figure S3). The corneal endothelium from control tissues was isolated and cultured similarly to the FECD tissues. The control samples included individuals harbouring ≤ 39 CTG expansions (Fig. 4C). In addition, we also analysed CECs from an individual with FECD but without CTG18.1 expanded alleles (blood STR genotype 11/14). Within these control samples, we observed an accumulation of molecules around 0 bp, near the detection threshold of the assay, and a complete absence of longer expanded molecules for both control and FECD non-expanded CEC samples (Fig. 4C; Figure S3). In contrast, remarkable levels of repeat instability were observed for all FECD CEC samples with baseline leukocyte repeat sizes ≥ 63 determined by STR genotyping. Spearman correlation analysis across

the CEC dataset suggested a correlation between the largest progenitor CTG18.1 allele size (determined by STR-PCR of leukocyte gDNA) and the molecule size measured by OGM (0.229, 95% CI: 0.201–0.256) (Table S2). We found no effect of the progenitor allele size, sampling age or sex when we compared CTG18.1 molecule sizes within the expansion-negative and the expansion-positive groups in isolation (Fig. 4D; Table S2). However, when the measured molecule size was compared between the expansion-negative and expansion-positive sample groups, there was a significant difference (Mann Whitney test, $p = 0.001$), suggesting that this is driven by the presence of a CTG18.1 repeat expansion, irrespective of progenitor allele size (Fig. 4D). These data indicate that CTG18.1 repeat instability can be influenced by both the progenitor repeat size and cellular context.

Discussion

FECD is an age-related degenerative eye disease hypothesised to share multiple pathogenic mechanistic parallels with many rare and devastating neurological and neuromuscular diseases also attributed to STR expansions.^{6,33} Here, through the application of OGM,

we have shown that CTG18.1 expansions have an exceptionally high level of somatic instability within affected CECs. This finding provides mechanistic insights into the pathophysiology of FECD. It also highlights the potential utility of this approach to investigate STR instability more broadly across other repeat loci susceptible to high levels of somatic instability.

A strong correlation between CTG18.1 expansion length in peripheral blood leukocytes and molecular hallmarks of pathology in the corneal endothelium, including the presence of nuclear CUG-specific RNA foci and isoform-specific patterns of *TCF4* down-regulation, is well documented.^{7,32,34} Our study has shown that the mean repeat length in mono-allelic expanded CEC lines is 4.4–17.0-fold longer than in leukocyte gDNA samples. Specifically, CECs derived from individuals ($n = 9$) with ≥ 63 baseline CTG18.1 repeats in leukocytes had molecules comprising 1800–11,900 repeats. Thus, probing CTG18.1 repeat length in these CECs has provided new insights during end-stage corneal endothelial disease and highlights the biological relevance of studying repeat instability directly within primarily affected cells. These data, alongside the low levels repeat of instability and lack of molecular hallmarks in expansion-positive fibroblasts, support the hypothesis that high levels of repeat instability could drive downstream molecular events and further reinforces the utility of primary CECs as an appropriate model to investigate FECD pathophysiology *ex vivo*. However, it should be noted that primary CECs are a finite resource, given the scarcity of healthy and diseased corneal endothelial tissues available for research. Regardless of this limitation, significant biological insights have been gained from the restricted sample sizes utilised by this study.

It is now of therapeutic relevance to determine the temporal dynamics of CTG18.1 instability within the corneal endothelium before end-stage disease (i.e., <58 years of age at sampling, reported in this study), given that downstream pathogenic features of CTG18.1 expansion-mediated FECD are hypothesised to be repeat length dependent. Subject sampling age contributed to the increased levels of somatic instability driven by the allele size in the expansion-positive leukocyte-derived samples, in contrast to the expansion-positive CEC subgroup. The lack of correlation in the CECs between subject sampling age and progenitor repeat size may be explained by the small sample size in the series or by somatic instability levels already being saturated in the late-stage disease cells. Future studies are required to determine how instability levels vary within the post-mitotic CECs at earlier pre-symptomatic disease stages. However, this is not a trivial goal given corneal tissues from CTG18.1 expansion-positive individuals without clinical need for corneal transplantation are not easily accessible.

CTG18.1 molecules with approximately 4500 repeats were also detected in the immortalised F35T CEC line, previously reported to have approximately 4500 CTG repeats by Southern blot, suggesting that successive passaging does not grossly alter repeat size. More generally, *in vitro* studies on CAG repeat instability have reported either a minor (1 CAG repeat increase every 12.4 days) or no correlation between cell proliferation and repeat instability.^{35,36} Taken together, these findings suggest that establishing and culturing primary CECs for a single passage, required due to the low cell numbers available from each corneal endothelial explant, is likely to only minimally affect repeat instability and does not explain the magnitude of the repeat instability detected in these cells. Nonetheless, cell culture can introduce artifacts, and in future, as input requirements are reduced for long-read technologies, the need for cell culture may be circumvented.

Interestingly, a link between exposure to ultraviolet light and FECD has been suggested in a CTG18.1 expansion-agnostic setting.^{37,38} We hypothesise that cumulative UV radiation-inducing DNA damage could contribute to the extreme CTG18.1 instability via the DNA repair-dependent mechanisms in the terminally differentiated CECs. Further studies to explore potential associations between CTG18.1 instability and DNA repair pathways are warranted. Importantly, the extreme levels of instability observed in CTG18.1 expansion-positive CECs were absent from both control and expansion-negative FECD CECs, suggesting that dynamic expansion of the repeat element within the corneal endothelium only occurs when the progenitor CTG18.1 repeat size is above a critical threshold (>39 and ≤ 63 CTG repeats). The presence of RNA foci within CECs has been previously reported in patients with ≥ 31 CTG18.1 repeats.⁷ However, control CECs from a 58-year-old individual harbouring 39 CTG18.1 repeats did not display the high levels of somatic instability observed in the expansion-positive group. This could be explained by differences in the sensitivities of the assays, the contribution of genetic modifiers that may alter the levels of somatic instability, and differences in subject sampling ages. Future efforts will likely further refine the threshold for disease and identify individual modifiers of instability that have been effectively delineated for other STR-associated diseases.³⁹

Analysis of a CEC sample with bi-allelic CTG18.1 expansions suggests that CTG18.1 molecules expand dynamically in the vast majority of CECs with progenitor allele size ≥ 50 CTG repeats at end-stage disease. We anticipate advances in long-read sequencing technologies to enable phasing and sequence-level resolution, currently lacking in OGM, and yield further biological insight into the pathophysiology of CTG18.1-mediated FECD. Notably, total molecule counts from the bi-allelic expanded sample were significantly lower than

other samples in the study, suggesting that despite analysing native molecules, OGM is biased against capturing longer molecules.

OGM has an average genome-wide lower detection limit of 500 bp.⁴⁰ On this basis, we queried its utility for the detection of CTG18.1 expansions in the majority of FECD blood leukocyte-derived gDNA samples, given that 500 bp corresponds to approximately 166 CTG repeats, and in 96.7% of our patient cohort, the expanded repeat sizes ranged from 50 to 125 CTG repeats.⁷ However, analysis of leukocyte samples from individuals with a diverse range of expanded CTG18.1 allele sizes (63–107 progenitor repeats) illustrates that OGM can detect an increase in the average CTG18.1 molecule size for progenitor alleles ≥ 63 repeats but cannot discriminate between samples with alleles ≤ 32 repeats. We hypothesise this signal is driven by the dynamic instability of CTG18.1, occurring within a fraction of total peripheral blood leukocytes, which increases with an increase in the progenitor repeat size. We have previously shown that this dynamic shift in instability occurs at >80 and ≤ 91 CTG repeats using an amplification-free long-read sequencing approach.¹¹ The peripheral blood leukocyte gDNA OGM data series presented here is in keeping with this observation, given that the measured molecule size increases exponentially and samples with progenitor model repeats ≥ 81 displayed the most pronounced signature of instability. More generally, these data serve as a benchmark for the wider utility of OGM to explore STR sizing detection and instability thresholds at other loci. To the best of our knowledge, OGM has, to date, only been used to detect and size STR expansions across a limited number of loci and without interrogating the primarily affected tissue.^{22,23,25,41} We believe OGM will offer the greatest utility when modal repeat sizes for any given locus exceed the lower detection limit of 500 bp and large somatically unstable repeats are present, as the method has no upper size detection threshold. Nonetheless, our data illustrate that it is also useful when only a small fraction of total DNA molecules exceed the lower detection limit. This study also illustrates the complementary application of combining OGM with conventional PCR-based fragment size analysis to enable detection of all potential ranges of repeat sizes across somatically unstable STR loci.

In summary, this study extends on the application of OGM for the detection and sizing of large repeat expansions from native DNA molecules. Taking advantage of the single-molecule resolution of the assay, we have demonstrated the importance of investigating STR instability in disease-relevant cell types to probe dynamic and tissue-specific mechanisms of somatic instability. Given the exceptionally high levels of CTG18.1 repeat length instability consistently and exclusively observed in affected CECs, these data also lead us to hypothesise that CTG18.1 instability is a key

driver of FECD pathophysiology. Furthermore, analysis of leukocyte gDNA demonstrates that both progenitor allele size and sampling age can influence CTG18.1 somatic instability levels. These new insights are anticipated to inform and guide the development of translational interventions for this common age-related triplet repeat mediated disease, as well as others.

Contributors

CZ and AED conceptualised and designed the study. CZ and AED generated and analysed the data and interpreted the results; SE and ND shared their OGM expertise; SF performed the bioinformatic analysis of the CTG18.1 molecules; CZ, NB, SL, MAC, AS and ANS performed bio-sample processing, PCR-based genotyping and cell culturing; ASJ's laboratory generated the F35T cell line; EB, HH and AC obtained funding for and granted access to the Saphyr system; PS, LD, KM and SJT contributed to patient recruitment and sample collection; MEC, AJH, PL and SJT edited the manuscript; AED and CZ verified the underlying data and wrote the manuscript. All authors read and approved the final manuscript.

Data sharing statement

The data sets supporting the conclusions of this article are included within the article and its Supplemental Files. All raw CTG18.1 locus-specific molecule OGM data generated in this study are provided in Supplemental File 1. F35T cell line can be obtained by contacting Johns Hopkins Technology Ventures office (ventures.jhu.edu).

Declaration of interests

The authors declare no competing interests. AED has previously acted as a paid consultant for Triplet Therapeutics Ltd, LoQus23 Therapeutics Ltd, Design Therapeutics Ltd and had a research collaboration with ProQR Therapeutics. AED has an ongoing research collaboration with Prime Medicine.

Acknowledgements

We thank all affected individuals for participating in this research. We also acknowledge Beverly Scott for their technical expertise in DNA extractions and bio-sample processing. We also thank Andrew Powers (Design Therapeutics) for his intellectual input and advice on the application of OGM and Polychronis Kemos for his input on statistical analysis. We would also like to acknowledge Miracles for Sight Eye Bank for providing control corneal tissue for our research.

This work was funded by a UKRI Future Leader Fellowship MR/S031820/1 (AED), Moorfields Eye Charity, GR001395, GR001337 (AED), Medical Research Council, R/X006271/1 (SL, AED), Fight for Sight UK, 5171/5172 (AED), Sight Research UK (AED) and The National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology. SE, ND, HH and EB are part of the ICGNMD Consortium, funded by an MRC strategic award to establish an International Centre for Genomic Medicine in Neuromuscular Diseases (ICGNMD) MR/S005021/1. The Saphyr platform was funded by the National Brain Appeal's Innovation Fund and Rosetrees Trust. LD, PS, and PL were supported by GACR 20-19278S, UNCE/24/MED/022 and SVV 260631.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105328>.

References

- 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;5:151–175. <https://doi.org/10.1146/annurev-vision-091718-014852>.

- 2 Mathews P, Benbow A, Corcoran K, DeMatteo J, Philippy B, Van Meter W. 2022 Eye banking statistical report—executive summary. *Eye Bank Corneal Transplant*. 2023;2:e0008. <https://doi.org/10.1097/ebct.0000000000000008>.
- 3 NHS organ and tissue donation and transplantation activity report 2023/2024. Section 10 - Cornea Activity; 2023. <https://www.odt.nhs.uk/statistics-and-reports/annual-activity-report/>.
- 4 Joyce NC. Proliferative capacity of corneal endothelial cells. *Exp Eye Res*. 2012;95:16–23. <https://doi.org/10.1016/j.exer.2011.08.014>.
- 5 Price MO, Mehta JS, Jurkunas UV, Price FW. Corneal endothelial dysfunction: evolving understanding and treatment options. *Prog Retin Eye Res*. 2021;82:100904. <https://doi.org/10.1016/j.preteyeres.2020.100904>.
- 6 Fautsch MP, Wieben ED, Baratz KH, et al. TCF4-mediated Fuchs endothelial corneal dystrophy: insights into a common trinucleotide repeat-associated disease. *Prog Retin Eye Res*. 2021;81:100883. <https://doi.org/10.1016/j.preteyeres.2020.100883>.
- 7 Zarouchlioti C, Sanchez-Pintado B, Tear NJH, et al. Antisense therapy for a common corneal dystrophy ameliorates TCF4 repeat expansion-mediated toxicity. *Am J Hum Genet*. 2018;102:528. <https://doi.org/10.1016/j.ajhg.2018.02.010>.
- 8 Mouro Pinto R, Arning L, Giordano JV, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum Mol Genet*. 2020;29:2551–2567. <https://doi.org/10.1093/hmg/ddaa139>.
- 9 Wheeler VC, Dion V. Modifiers of CAG/CTG repeat instability: insights from mammalian models. *J Huntingt Dis*. 2021;10:123–148. <https://doi.org/10.3233/JHD-200426>.
- 10 Monckton DG. The contribution of somatic expansion of the CAG repeat to symptomatic development in huntington's disease: a historical perspective. *J Huntingt Dis*. 2021;10:7–33. <https://doi.org/10.3233/JHD-200429>.
- 11 Hafford-Tear NJ, Tsai Y-C, Sadan AN, et al. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy-associated TCF4 triplet repeat. *Genet Med*. 2019;21:2092–2102. <https://doi.org/10.1038/s41436-019-0453-x>.
- 12 Wieben ED, Aleff RA, Rinkoski TA, et al. Comparison of TCF4 repeat expansion length in corneal endothelium and leukocytes of patients with Fuchs endothelial corneal dystrophy. *PLoS One*. 2021;16:e0260837. <https://doi.org/10.1371/journal.pone.0260837>.
- 13 Depienne C, Mandel J-L. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am J Hum Genet*. 2021;108:764–785. <https://doi.org/10.1016/j.ajhg.2021.03.011>.
- 14 Nakamori M, Sobczak K, Puwanant A, et al. Splicing biomarkers of disease severity in myotonic dystrophy. *Ann Neurol*. 2013;74:862–872. <https://doi.org/10.1002/ana.23992>.
- 15 Handsaker RE, Kashin S, Reed NM, et al. Long somatic DNA-repeat expansion drives neurodegeneration in Huntington disease 2. *bioRxiv*. 2024. <https://doi.org/10.1101/2024.05.17.592722>.
- 16 Benn CL, Gibson KR, Reynolds DS. Drugging DNA damage repair pathways for trinucleotide repeat expansion diseases. *J Huntingt Dis*. 2021;10:203–220. <https://doi.org/10.3233/JHD-200421>.
- 17 Ciosi M, Cumming SA, Chatzi A, et al. Approaches to sequence the HTT CAG repeat expansion and quantify repeat length variation. *J Huntingt Dis*. 2021;10:53–74. <https://doi.org/10.3233/JHD-200433>.
- 18 Ciosi M, Maxwell A, Cumming SA, et al. A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine*. 2019;48:568–580. <https://doi.org/10.1016/j.ebiom.2019.09.020>.
- 19 Erdmann H, Schöberl F, Giurgiu M, et al. Parallel in-depth analysis of repeat expansions in ataxia patients by long-read sequencing. *Brain J Neurol*. 2023;146:1831–1843. <https://doi.org/10.1093/brain/awac377>.
- 20 Höjjer I, Tsai Y-C, Clark TA, et al. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum Mutat*. 2018;39:1262–1272. <https://doi.org/10.1002/humu.23580>.
- 21 Yuan Y, Chung CY-L, Chan T-F. Advances in optical mapping for genomic research. *Comput Struct Biotechnol J*. 2020;18:2051–2062. <https://doi.org/10.1016/j.csbj.2020.07.018>.
- 22 Facchini S, Dominik N, Manini A, et al. Optical genome mapping enables detection and accurate sizing of RFC1 repeat expansions. *Biomolecules*. 2023;13:1546. <https://doi.org/10.3390/biom13101546>.
- 23 Morato Torres CA, Zafar F, Tsai Y-C, et al. ATTCT and ATTCC repeat expansions in the ATXN10 gene affect disease penetrance of spinocerebellar ataxia type 10. *HGG Adv*. 2022;3:100137. <https://doi.org/10.1016/j.xhgg.2022.100137>.
- 24 Maroille T, Tsai M-H, Mascarenhas R, et al. A novel FAME1 repeat configuration in a European family identified using a combined genomics approach. *Epilepsia Open*. 2023;8:659–665. <https://doi.org/10.1002/epi4.12702>.
- 25 Efthymiou S, Lemmers RJLF, Vishnu VY, et al. Optical genome mapping for the molecular diagnosis of facioscapulohumeral muscular dystrophy: advancement and challenges. *Biomolecules*. 2023;13:1567. <https://doi.org/10.3390/biom13111567>.
- 26 Peh GSL, Chng Z, Ang H-P, et al. Propagation of human corneal endothelial cells: a novel dual media approach. *Cell Transplant*. 2015;24:287–304. <https://doi.org/10.3727/096368913X675719>.
- 27 Saha S, Skeie JM, Schmidt GA, et al. TCF4 trinucleotide repeat expansions and UV irradiation increase susceptibility to ferroptosis in Fuchs endothelial corneal dystrophy 2022. 2022:497862. <https://doi.org/10.1101/2022.06.27.497862>.
- 28 Angelbello AJ, Benhamou RI, Rzuczek SG, et al. A small molecule that binds an RNA repeat expansion stimulates its decay via the exosome complex. *Cell Chem Biol*. 2021;28:34–45.e6. <https://doi.org/10.1016/j.chembiol.2020.10.007>.
- 29 Wieben ED, Aleff RA, Tosakulwong N, et al. A common trinucleotide repeat expansion within the transcription factor 4 (TCF4, E2-2) gene predicts Fuchs corneal dystrophy. *PLoS One*. 2012;7:e49083. <https://doi.org/10.1371/journal.pone.0049083>.
- 30 Mootha VV, Gong X, Ku H-C, Xing C. Association and familial segregation of CTG18.1 trinucleotide repeat expansion of TCF4 gene in Fuchs' endothelial corneal dystrophy. *Invest Ophthalmol Vis Sci*. 2014;55:33–42. <https://doi.org/10.1167/iovs.13-12611>.
- 31 Mantere T, Neveling K, Pebrel-Richard C, et al. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet*. 2021;108:1409–1422. <https://doi.org/10.1016/j.ajhg.2021.05.012>.
- 32 Bhattacharyya N, Chai N, Hafford-Tear NJ, et al. Deciphering novel TCF4-driven mechanisms underlying a common triplet repeat expansion-mediated disease. *PLoS Genet*. 2024;20:e1011230. <https://doi.org/10.1371/journal.pgen.1011230>.
- 33 Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol*. 2021;22:589–607. <https://doi.org/10.1038/s41580-021-00382-6>.
- 34 Hu J, Rong Z, Gong X, et al. Oligonucleotides targeting TCF4 triplet repeat expansion inhibit RNA foci and mis-splicing in Fuchs' dystrophy. *Hum Mol Genet*. 2018;27:1015–1026. <https://doi.org/10.1093/hmg/ddy018>.
- 35 Gomes-Pereira M, Fortune MT, Monckton DG. Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. *Hum Mol Genet*. 2001;10:845–854. <https://doi.org/10.1093/hmg/10.8.845>.
- 36 Goold R, Flower M, Moss DH, et al. FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Hum Mol Genet*. 2019;28:650–661. <https://doi.org/10.1093/hmg/ddy375>.
- 37 Liu H, Cheng J, Zhuang X, Qi B, Li F, Zhang B. Genomic instability and eye diseases. *Adv Ophthalmol Pract Res*. 2023;3:103–111. <https://doi.org/10.1016/j.aopr.2023.03.002>.
- 38 Liu C, Miyajima T, Melangath G, et al. Ultraviolet A light induces DNA damage and estrogen-DNA adducts in Fuchs endothelial corneal dystrophy causing females to be more affected. *Proc Natl Acad Sci U S A*. 2020;117:573–583. <https://doi.org/10.1073/pnas.1912546116>.
- 39 Rajagopal S, Donaldson J, Flower M, Hensman Moss DJ, Tabrizi SJ. Genetic modifiers of repeat expansion disorders. *Emerg Top Life Sci*. 2023;7:325–337. <https://doi.org/10.1042/ETLS20230015>.
- 40 Zhang Q, Wang Y, Xu Y, et al. Optical genome mapping for detection of chromosomal aberrations in prenatal diagnosis. *Acta Obstet Gynecol Scand*. 2023;102:1053–1062. <https://doi.org/10.1111/aogs.14613>.
- 41 Ghorbani F, de Boer-Bergsma J, Verschuuren-Bemelmans CC, et al. Prevalence of intronic repeat expansions in RFC1 in Dutch patients with CANVAS and adult-onset ataxia. *J Neurol*. 2022;269:6086–6093. <https://doi.org/10.1007/s00415-022-11275-9>.