nature genetics

Technical Report

https://doi.org/10.1038/s41588-024-02067-0

Synchronized long-read genome, methylome, epigenome and transcriptome profiling resolve a Mendelian condition

Received: 26 September 2023

Accepted: 19 December 2024

Published online: 29 January 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Resolving the molecular basis of a Mendelian condition remains challenging owing to the diverse mechanisms by which genetic variants cause disease. To address this, we developed a synchronized long-read genome, methylome, epigenome and transcriptome sequencing approach, which enables accurate single-nucleotide, insertion–deletion and structural variant calling and diploid de novo genome assembly. This permits the simultaneous elucidation of haplotype-resolved CpG methylation, chromatin accessibility and full-length transcript information in a single long-read sequencing run. Application of this approach to an Undiagnosed Diseases Network participant with a chromosome X;13-balanced translocation of uncertain significance revealed that this translocation disrupted the functioning of four separate genes (*NBEA*, *PDK3*, *MAB21L1* and *RB1*) previously associated with single-gene Mendelian conditions. Notably, the function of each gene was disrupted via a distinct mechanism that required integration of the four 'omes' to resolve. These included fusion transcript formation, enhancer adoption, transcriptional readthrough silencing and inappropriate X-chromosome inactivation of autosomal genes. Overall, this highlights the utility of synchronized long-read multi-omic profiling for mechanistically resolving complex phenotypes.

The diagnosis of Mendelian conditions is challenged by the ability to detect pathogenic genetic variation accurately and to determine whether an identified genetic variant has a functional consequence. Recent advances in accurate long-read sequencing have markedly improved our ability to detect genetic variants[1–4], yet our ability to determine whether an identified genetic variant has a functional consequence remains quite limited. This challenge is particularly acute when evaluating noncoding genetic variants owing to the >4 million noncoding variants harbored in each genome[5] and the diverse mechanisms by which noncoding variants can cause disease. Multi-omic approaches have shown promise in resolving this challenge by integrating genomic information with functional information from the same sample, such as RNA sequencing (RNA-seq) and CpG methylation information[6–10]. However, the broader adoption of these multi-omic approaches for resolving Mendelian conditions has remained limited, in part due to experimental and analytical shortcomings with current multi-omic approaches. Specifically, current multi-omic approaches are grounded in short-read sequencing, which limits their ability to comprehensively characterize the functional output of individual haplotypes within a sample. In addition, current approaches combine data from independently prepared samples, which necessitates batch effect corrections between each modality assessed.

Accurate long-read sequencing has the potential to resolve these limitations through haplotype-phased epigenetic and transcriptomic information. Specifically, accurate long-read sequencing permits the simultaneous detection of CpG methylation information on each sequenced read[11,12], which provides a direct evaluation of the impact of genetic variants on CpG methylation patterns[13–15]. In addition,

e-mail: absterga@uw.edu

long-read sequencing can be leveraged to enable the simultaneous identification of chromatin architectures[16–20], such as the occupancy of nucleosomes, transcription factors and accessible chromatin patches along each sequenced read. For example, with single-molecule chromatin fiber sequencing (Fiber-seq) before DNA extraction, permeabilized cells are treated with a nonspecific N^6-methyladenine methyltransferase (m^6A-MTase), which stencils the architecture of each chromatin fiber onto its underlying DNA molecule in the form of m^6A-modified bases that can be directly read out during standard single-molecule, real-time (SMRT) sequencing. Furthermore, recent advances in Fiber-seq enable the co-identification of both accurate chromatin features[20,21] and CpG methylation along each sequenced read[22], providing genomic sequence, CpG methylation and chromatin epigenetic information in the same sequencing run. Finally, long-read, full-length transcript sequencing is emerging as a powerful tool for resolving the transcript impact of genetic variants[23,24], and recent cDNA (cDNA) concatenation workflows for processing full-length transcript data (multiplexed arrays isoform sequencing (MAS-seq))[25] provide an order-of-magnitude higher throughput identification of full-length transcript data. Overall, these improvements in the throughput, accuracy and utility of long-read sequencing over the past few years have made long-read sequencing a tractable tool for genomic studies.

We sought to leverage recent advances in both Fiber-seq and full-length long-read transcript sequencing to create a synchronized long-read multi-ome, which provides an accurate, haplotype-phased, long-read genome, CpG methylome, chromatin epigenome and transcriptome without the need for tiered sampling or redundant sequencing (Fig. 1a). We benchmarked the performance of this synchronized multi-ome assay against Genome-in-a-Bottle (GIAB) and Human Pangenome Reference Consortium (HPRC) samples with known ground truths. We then applied it to solve the genetic and molecular basis of a participant within the Undiagnosed Diseases Network (UDN), demonstrating how each modality contributes to revealing the molecular basis of this individual's disease in a single sequencing run.

Results

Synchronized long-read multi-ome profiling

We first tested the accuracy of this approach for identifying genetic variants by applying long-read multi-ome profiling to two GIAB and two HPRC cell lines that have been extensively characterized (GM12878 (HG001), GM24385 (HG002), HG02630 and GM20129). Fiber-seq-modified genomic DNA (gDNA) was generated by treating permeabilized cells with a nonspecific m^6A-MTase for 10 min, followed by gDNA extraction, shearing to ~18 kb and barcoding using standard SMRTbell library prep. In parallel, RNA was isolated from the same cells, and polyA-primed cDNA was generated and concatenated into ~18 kb molecules using a modified version of the MAS-seq approach[25] that we optimized for bulk RNA. Concatenating multiple cDNA molecules together into a ~18 kb barcoded DNA molecule enables the gDNA and concatenated cDNA to have similar lengths, overcoming common issues that arise from pooling libraries of significantly different lengths into the same sequencing run (Fig. 1a). Pooling Fiber-seq gDNA and MAS-seq concatenated cDNA at a molar ratio of 15:1 (GIAB samples) or 9:1 (HPRC samples) resulted in a total of over 5 million sequencing reads for each sample, with 5–15% of these reads deriving from MAS-seq concatenated cDNA while the Fiber-seq gDNA reads permitted 26–31× genome-wide coverage (Table 1).

Validation of synchronized long-read multi-ome profiling

Overall, this sequencing strategy did not detract from the ability to accurately identify genetic variants. Specifically, Fiber-seq-modified gDNA from GM24385 cells retained highly accurate variant calling with a single nucleotide variant (SNV) F1 of 99.94%, an insertion–deletion (indel) F1 of 98.81% and a structural variant (SV) F1 of 92.5% against established gold-standard GIAB variant annotations from HG002 (Table 1). In addition, as this approach is grounded in accurate long-read sequencing, 87–98% of Fiber-seq sequencing reads could be accurately phased to their respective haplotype (higher values from the samples that are of African genetic ancestry)[26], and these reads were compatible with generating high-quality long-read de novo diploid genome assemblies (median haplotype contig N50s 29 Mb and consensus accuracy quality value (QV) 50).

This approach also enabled the identification of epigenetic information. Specifically, CpG methylation information directly obtained using long-read multi-ome profiling correlated with existing whole-genome bisulfite sequencing (WGBS) and Oxford Nanopore (ONT) sequencing data (mean Pearson correlation of 0.87 with existing ONT CpG methylation datasets; Table 1). In addition, m^6A modifications were obtained using long-read multi-ome profiling demarcated on average 377 million nucleosome footprints (average size 146 bp). Using these single-molecule m^6A-marked chromatin features, we identified on average 113,846 accessible elements, with 93.7% of these peaks mapping to known regulatory elements as defined by ENCODE[27]. Finally, this approach enabled the identification of 2.3–6.5 million full-length nonconcatenated transcripts (higher values obtained from samples pooled at a ratio of 9:1), which arise from, on average 63,329 unique splicing isoforms derived from on average 13,088 genes. Overall, these data demonstrate strong agreement between synchronized long-read multi-ome profiling and that obtained using traditional short-read-based transcript and epigenetic studies.

Long-read multi-ome profiling of GIAB cells identified haplotype-phased genetic variation, CpG methylation, chromatin accessibility and transcript data between both haplotype-phased genomes contained within these cells (Fig. 1b and Extended Data Fig. 1). On average, using our ~30-fold genomic coverage datasets, we identified 1,366 accessible chromatin elements within each sample with nominally significant haplotype-specific chromatin accessibility (21 per sample with genome-wide significance), many of which were contained within imprinted loci (Extended Data Fig. 2). Pairing the haplotype-resolved chromatin and transcriptomic data together at these loci enabled the identification of specific regulatory elements that were subjected to imprinting, as well as the transcriptional output of these regulatory elements (Fig. 1b). Furthermore, this approach identified noncoding genetic variants within challenging-to-map regions of the genome, such as the *HLA* locus, that appear to disrupt gene regulatory patterns (Extended Data Fig. 3).

Identifying the molecular basis of a Mendelian condition

We next determined whether this approach could identify the molecular basis for a previously unsolved Mendelian condition. Specifically, we applied synchronized long-read multi-ome profiling to skin fibroblasts from an undiagnosed 9-month-old female with bilateral retinoblastomas, developmental delay, polymicrogyria, sensorineural hearing loss (SNHL), lactic acidosis, hypotonia and dysmorphic facial features who was enrolled within the UDN (UDN318336; Fig. 2a and Supplementary Table 1). Prior trio genome sequencing did not reveal any genetic variants thought to be causing the phenotype, and an SNP array was normal (Supplementary Table 2). Clinical RNA-seq of *RB1* revealed that bulk *RB1* expression was at the low end of normal, but *RB1* Sanger sequencing was normal. Karyotype revealed an apparently balanced de novo X;13 translocation (46,XX,t(X;13)(p22.1;q14.1)) of uncertain significance (Fig. 2b), and fluorescence in situ hybridization studies showed that *RB1* was unlikely to be disrupted at the breakpoint. Balanced translocations are common structural chromosomal rearrangements that occur in 0.52% of the population[28] and result in copy-neutral changes in the genome. While the majority of balanced translocations do not result in disease[29], certain balanced translocations are known to cause Mendelian conditions when translocations disrupt genes, create fusion

https://doi.org/10.1038/s41588-024-02067-0

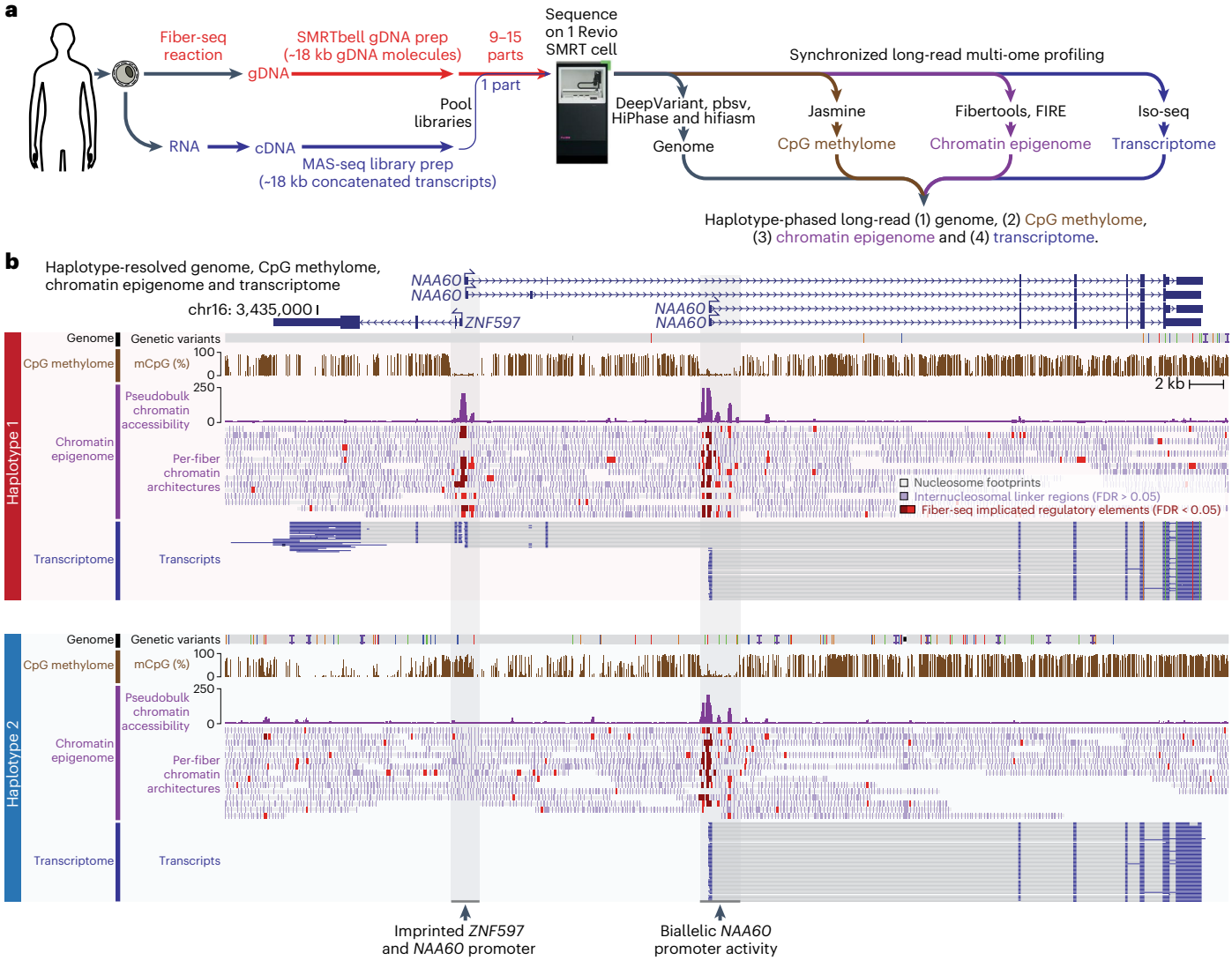


Fig. 1 | Synchronized long-read genome, methylome, epigenome and transcriptome sequencing. a, Schematic representation describing the experimental and computational workflow for synchronized multi-ome profiling. Specifically, cells are subjected to a Fiber-seq reaction followed by gDNA extraction and SMRTbell library preparation, and, in parallel, cells are subjected to an RNA extraction followed by cDNA synthesis and MAS-seq library preparation. The two libraries are then mixed together and sequenced simultaneously using a single sequencing run, enabling the simultaneous detection of the genome, CpG methylome, chromatin epigenome and transcriptome from the sample. **b**, Example genomic region showing the haplotype-resolved genome, CpG methylome, chromatin epigenome and transcriptome from GM12878 cells at a known imprinted locus.

transcripts, alter gene regulatory programs or involve chromosome X, which can also result in the extension of X-chromosome inactivation (XCI) into autosomal DNA[30]. These mechanisms often require an array of diverse functional studies to independently evaluate.

Long-read genome exposes *NBEA* disruption

Long-read multi-ome profiling of fibroblast cells from this individual resulted in 28× genome-wide haplotype-resolved coverage of the genome, CpG methylome and chromatin epigenome, as well as 2,196,793 full-length transcripts (Table 1). Genome assembly of these sequencing reads delineated the precise translocation breakpoints (Fig. 2c), which resulted in the creation of a ~50 Mb derivative chromosome 13, der(13), that contains all of 13p and part of 13q and Xp, as well as a ~210 Mb derivative chromosome X, der(X), that contains all of Xq and part of Xp and 13q (Fig. 2d). The translocation breakpoint on 13q is located in intron 41 of the 58-exon gene *NBEA*. Consequently, the first 41 exons of *NBEA* are located on der(13), and the terminal 17 exons are located on der(X) (Fig. 2e). The haplotype-phased multi-ome functional data demonstrated that although the *NBEA* promoters on the intact chromosome 13 ($chr13_{int}$) and der(13) showed similar chromatin accessibility (Extended Data Fig. 4a), full-length *NBEA* transcripts would only be able to arise from $chr13_{int}$, consistent with this balanced translocation resulting in disrupted *NBEA* transcripts arising from the der(X). To further explore whether these disrupted transcripts result in a potential truncated protein or are readily removed via nonsense-mediated decay (NMD), we performed long-read transcript sequencing on UDN318336 fibroblasts with and without treatment with the NMD inhibitor cycloheximide (Extended Data Fig. 4b). This revealed that this translocation results in the formation of numerous distinct noncanonical *NBEA* transcripts that are not subjected to NMD and are predicted to result in the formation of a truncated *NBEA* protein product (Fig. 2f). This truncated *NBEA* protein retains all N-terminal domains, including the protein kinase A-binding region, but lacks the PH-BEACH and N-terminal WD40 domains (Fig. 2g), and prior experimental studies suggest that such a truncation could have unique impacts on dendrite formation[31].

https://doi.org/10.1038/s41588-024-02067-0

Table 1 | Overview of genetic, epigenetic and transcript features of long-read multi-ome samples

	Analysis metric	Sample				
		HG001	HG002	HG02630	GM20129	UDN318336
Sequencing statistics	Pooling ratio (gDNA:cDNA)	15:1	15:1	9:1	9:1	15:1
	HiFi read yield	98.8 Gb	90.1 Gb	85.2 Gb	92.1 Gb	90.3 Gb
	HiFi reads	6,268,744	6,000,613	5,015,661	5,074,241	5,902,290
	Insert length N50	18,932 bp	17,443 bp	16,953 bp	18,081 bp	17,221 bp
	Median HiFi read QV	Q31	Q32	Q33	Q33	Q31
Genome	gDNA reads	5,716,157	5,359,077	4,167,840	4,362,733	5,563,966
	gDNA read length N50	18,785 bp	17,664 bp	17,531 bp	18,379 bp	16,793 bp
	Total gDNA yield	89.92 Gb	80.34 Gb	73.23 Gb	80.71 Gb	83.34 Gb
	Median coverage of hg38	30×	26×	27×	31×	28×
	SNVs (F1)	4,416,098 (99.90%)	4,413,305 (99.94%)	5,342,920 (na)	5,255,964 (na)	4,396,931 (na)
	Indels (F1)	974,414 (98.75%)	973,416 (98.81%)	1,120,565 (na)	1,108,174 (na)	971,312 (na)
	SVs ≥50 bp (F1)	23,458 (na)	23,836 (92.5%)	23,834 (na)	24,015 (na)	23,201 (na)
	Percentage of reads haplotype phased	87.0%	90.0%	97.6%	98.1%	87.7%
	Assembly size (hap1/hap2)	3,016 Mb/3,018 Mb	3,010 Mb/3,010 Mb	3,048 Mb/3,031 Mb	3,033 Mb/3,042 Mb	3,036 Mb/3,038 Mb
	Number of contigs (hap1/hap2)	1,129/1,131	776/919	1,244/1,121	748/669	705/762
	Contig N50 (hap1/hap2)	30.6 Mb/30.7 Mb	36.3 Mb/28.8 Mb	15.4 Mb/14.6 Mb	26.0 Mb/29.3 Mb	40.5 Mb/27.3 Mb
	QV (hap1/hap2)	49.9/49.7	50.3/50.0	49.6/49.8	51.3/51.6	na
mCpG epigenome	Methylated CpG sites (%)	12,818,712 (45%)	20,252,720 (70%)	19,253,116 (66%)	20,426,659 (70%)	20,301,742 (71%)
	WGBS correlation (Pearson)	0.79	0.92	na	na	na
	ONT correlation (Pearson)	0.82	0.93	0.86	0.87	na
Chromatin epigenome	Nucleosome size	146 bp	147 bp	145 bp	144 bp	138 bp
	Number of nucleosomes	441,504,701	332,061,724	347,741,836	385,658,725	360,468,819
	FIREs in clusters[a]	40.0%	41.0%	38.6%	42.5%	46.0%
	FIREs in ENCODE peaks	66.0%	67.0%	67.7%	69.3%	75.0%
	Number of FIRE peaks[b]	122,016	84,121	112,555	136,690	167,752
	Percentage of FIRE peaks in ENCODE peaks	91.6%	97.3%	93.7%	92.1%	96.3%
Transcriptome	MAS-seq reads	518,921	604,272	825,212	695,914	297,248
	Transcript reads	2,347,741	3,716,982	6,494,682	5,395,840	2,196,793
	Mean transcript length	2,579 bp	2,041 bp	1,747 bp	2,020 bp	2,798 bp
	Mean array size (concatenation factor)	4.52	6.15	7.87	7.75	7.39
	Total unique genes	10,536	11,512	14,822	15,483	10,383
	Total unique isoforms	40,618	48,967	81,621	82,110	32,430

[a]Percentage of FIREs preferentially clustered along the genome. [b]Number of significant Fiber-seq peaks in the genome. na, not available.

https://doi.org/10.1038/s41588-024-02067-0

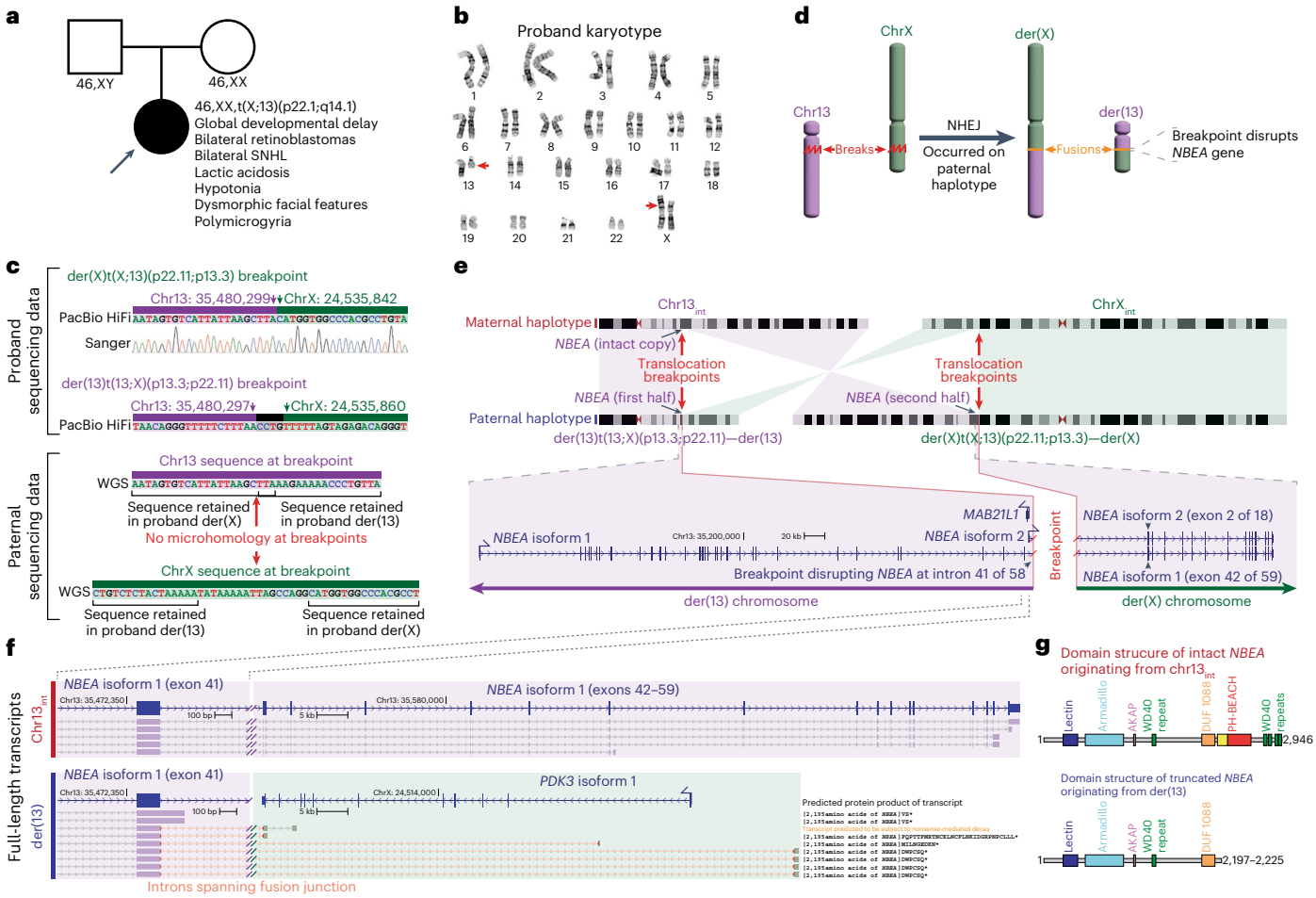


Fig. 2 | Long-read multi-ome for resolving the genetic basis of an unsolved Mendelian condition. a, Pedigree for the proband, as well as the clinical features of the proband and the results of her karyotype and that of her parents. **b**, Image of the proband's karyotype with the der(13) and der(X) chromosomes marked by red arrows. **c**, Sequence of the breakpoints on der(X) and der(13), as well as the sequence of this same region in chr13 and chrX in her father. Sanger trace showing validation of the der(X) breakpoint junction. **d**, Schematic representation showing the breakpoint and fusion event that occurred selectively on the paternal haplotype. **e**, Top, idiogram showing the $chr13_{int}$ and $chrX_{int}$, as well as the derivative chr13 and chrX in the proband. Translocation breakpoints and the location of the gene *NBEA* are highlighted. Bottom, gene model for both *NBEA* isoforms that differ in their transcriptional start site, showing the portion of *NBEA* that is located on der(13) versus der(X). **f**, Individual *NBEA* transcript reads from the intact chr13 and der(13) showing fusion transcripts along der(13), as well as the predicted protein product of each transcript and whether it is predicted to result in NMD. **g**, NBEA protein domains from NBEA transcripts arising from the intact chr13 or der(13).

Long-read transcriptome exposes *PDK3–MAB21L1* fusion transcript

PDK3 on Xp encodes two isoforms that differ by the inclusion of a small terminal exon in isoform 2 that adds seven amino acids to the protein sequence (Fig. 3a). The translocation breakpoint on Xp is located in this terminal intron for *PDK3* isoform 2, placing *PDK3* upstream of the gene *MAB21L1* with transcription for both of these genes in the same direction. The full-length transcript data revealed that this resulted in the formation of fusion transcripts between the splice donor site on exon 11 of *PDK3* isoform 2 and the splice acceptor site on exon 2 of *MAB21L1* isoform 2 that is not subjected to NMD (Extended Data Fig. 4c) and adds 66 amino acids at the C-terminal end of PDK3. PDK3 is one of four kinases essential for phosphorylating and subsequently inhibiting pyruvate dehydrogenase complex (PDC) activity[32]. This fusion transcript results in a protein with intact kinase activity that undergoes differential protein modification relative to the intact PKD3 protein (Extended Data Fig. 5a,b). Furthermore, this translocation abrogated the ability of der(13) to produce transcripts corresponding to *PDK3* isoform 2 (Fig. 3b), which contains a substantially longer 3′ UTR. However, despite these alterations, endogenous PDK3 protein levels remained low in these patient-derived fibroblast cells, consistent with low-level expression of *PDK3* in fibroblasts (Extended Data Fig. 5c).

Long-read CpG methylome exposes *MAB21L1* readthrough silencing

Production of this fusion kinase transcript on der(13) results in transcriptional readthrough of the *MAB21L1* promoter. Transcriptional readthrough can result in complex epigenetic changes[33,34], including epigenetic silencing, as is seen at the *MSH2* promoter in individuals with *EPCAM* deletions[34]. Notably, we observed that the *MAB21L1* gene was epigenetically and transcriptionally silenced selectively along der(13), as demonstrated by an allelic imbalance in both *MAB21L1* promoter CpG hypermethylation and *MAB21L1* full-length transcripts (Fig. 3c–e). To validate this finding, we generated retinal organoids from these patient-derived fibroblasts, as *MAB21L1* is highly expressed in the retina (Extended Data Fig. 6a). Fiber-seq on these patient-derived retinal organoids similarly demonstrated that the *MAB21L1* promoter is hypermethylated and shows reduced chromatin accessibility selectively on der(13) (Fig. 4a and Extended Data Fig. 6b), consistent with transcriptional readthrough silencing of *MAB21L1* on der(13).

https://doi.org/10.1038/s41588-024-02067-0

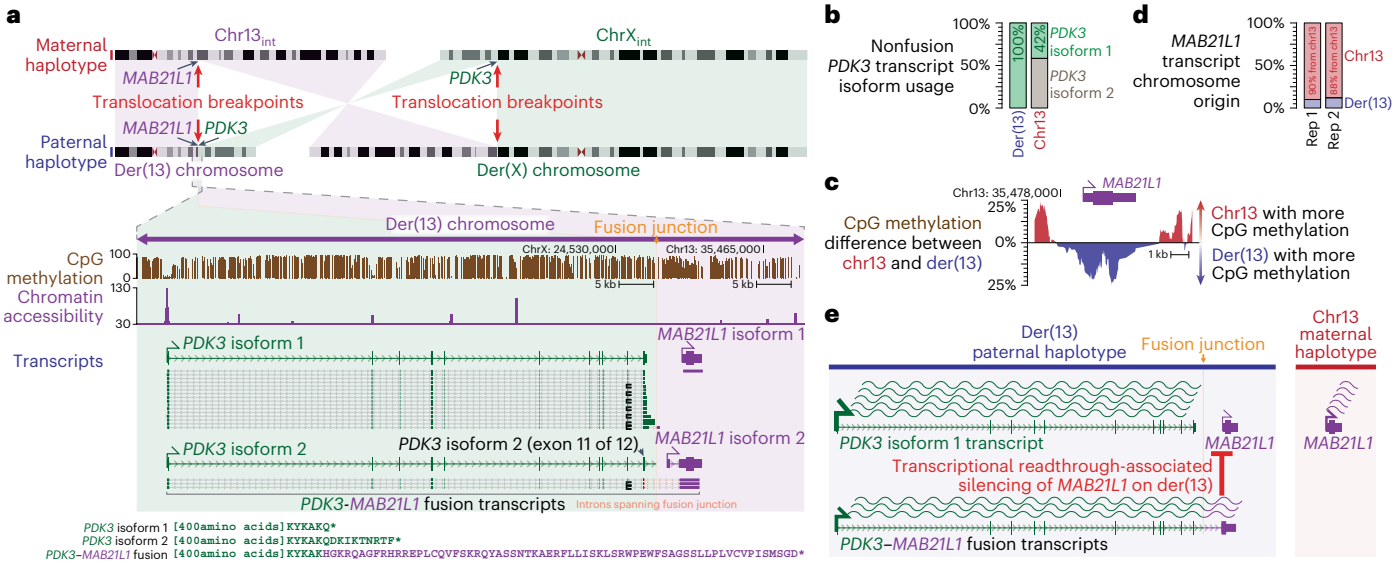


Fig. 3 | The results of der(13) in a *PDK3–MAB21L1* fusion transcript and *MAB21L1* silencing. a, Top, ideogram showing the intact chr13$_{int}$ and chrX$_{int}$, as well as the derivative chr13 and chrX in the proband. Translocation breakpoints and the location of the genes *PDK3* and *MAB21L1* are highlighted. Bottom, CpG methylation, chromatin accessibility and full-length transcript data selectively on the der(13) haplotype are displayed, highlighting the formation of a fusion transcript between *PDK3* and *MAB21L1*. **b**, Relative abundance of nonfusion *PDK3* transcripts arising from chr13$_{int}$ and der(13). **c**, CpG methylation differences at the *MAB21L1* promoter between chr13$_{int}$ and der(13) demonstrating selective hyper-CpG methylation of the *MAB21L1* promoter along der(13). **d**, Allelic imbalance of full-length *MAB21L1* transcripts between chr13$_{int}$ and der(13) in two long-read transcript sequencing replicates from patient fibroblasts demonstrating silencing of *MAB21L1* along der(13). **e**, Schematic representation for transcriptional readthrough silencing of the *MAB21L1* gene selectively along der(13).

Long-read chromatin epigenome exposes *PDK3* enhancer adoption

Fiber-seq of the patient-derived retinal organoids also identified nine distinct accessible elements within 10 kb of the *MAB21L1* transcriptional start site, with element 9 demonstrating significant codependent actuation with the *MAB21L1* promoter, as well as bulk H3K27ac signal across multiple tissues (Fig. 4a), consistent with it being a putative *MAB21L1* enhancer. Notably, the accessibility of element 9 is selectively increased within the retinal organoids, and element 9 remains accessible on 76% of the chromatin fibers deriving from der(13) in these cells, suggesting that element 9 is a cell-selective *MAB21L1* enhancer that escapes transcriptional readthrough silencing (Fig. 4b), likely owing to its position after the transcription termination signal for the *PDK3–MAB21L1* fusion transcript. As the translocation positions this putative enhancer adjacent to the *PDK3* promoter, which is located only ~70 kb away on der(13), we sought to see whether the *PDK3* promoter is adopting this cell-selective *MAB21L1* enhancer. Consistent with this model of enhancer adoption, we found that patient-derived induced pluripotent stem (iPS) cells and retinal organoids show long-range chromatin interactions between the *PDK3* promoter and this cell-selective *MAB21L1* enhancer (Extended Data Fig. 7) and that patient-derived retinal organoids demonstrated significantly elevated PDK3 protein levels compared to control retinal organoids (Fig. 4c). However, there was a nonsignificant change in pyruvate dehydrogenase (PDH) activity in these patient-derived retinal organoids (Extended Data Fig. 8). Consequently, the chromosome X;13 translocation appears to simultaneously result in *NBEA* truncation, a *PDK3–MAB21L1* fusion kinase transcript, abrogation of *PDK3* isoform 2 production, transcriptional readthrough silencing of *MAB21L1* and cell-selective enhanced *PDK3* expression via adoption of *MAB21L1* regulatory elements (Fig. 4d)—distinct perturbations to three genes known to be involved in Mendelian conditions.

Long-read multi-ome exposes inappropriate XCI of autosomal genes

RB1 is the only gene associated with hereditary bilateral retinoblastomas, yet *RB1* is located 13.5 Mb away from the breakpoint along der(X). A chromosome X;13 translocation involving the opposite arm of chromosome X has previously been identified in an individual with retinoblastomas and was found to result in the inappropriate inactivation of the *RB1* locus in ~10% of cells[30] (Extended Data Fig. 9). As such, we sought to evaluate the haplotype-phased epigenetic data from this patient to determine whether der(X) is being subjected to XCI and whether this extended into autosomal DNA along der(X). Overall, we observed that patient-derived fibroblasts demonstrated markedly skewed XCI, with the intact chromosome X (chrX$_{int}$) being preferentially subjected to XCI (Fig. 5a), consistent with prior observations in autosome-X translocations[35]. However, the autosomal DNA portion of der(X) did exhibit features of being epigenetically silenced in a subset of cells (Fig. 5b). Specifically, the 13p13.3 to 13qter region of chromosome 13, which is located on both der(X) and chr13$_{int}$, selectively and significantly exhibited an imbalance in allelic chromatin accessibility and transcript production consistent with this region on der(X) being silenced in 3–10% of cells (Fig. 5c). Notably, this imbalance in allelic chromatin accessibility and transcript production was not seen in the two GIAB samples (Extended Data Fig. 10a,b), consistent with this being a feature driven by the balanced X;13 translocation. Consequently, we estimate that in 3–10% of fibroblast cells from this patient, the *RB1* locus would be epigenetically silenced (Fig. 5d).

Discussion

We present a synchronized long-read genomic, epigenomic and transcriptomic sequencing approach and demonstrate the accuracy and utility of this approach for characterizing the mechanistic basis by which genetic variants cause disease. In addition to eliminating redundant sequencing and batch effects that are inherent to current multi-ome methods, this synchronized long-read multi-ome approach enables the haplotype phasing of 90–98% of all genetic and epigenetic features, as well as full-length transcripts that contain heterozygous variants. As such, this approach enables the comparison of epigenetic and transcript signals between the two haplotypes within diploid individuals, providing an internally controlled comparison to identify heterozygous or biallelic variants that alter the epigenome or

https://doi.org/10.1038/s41588-024-02067-0

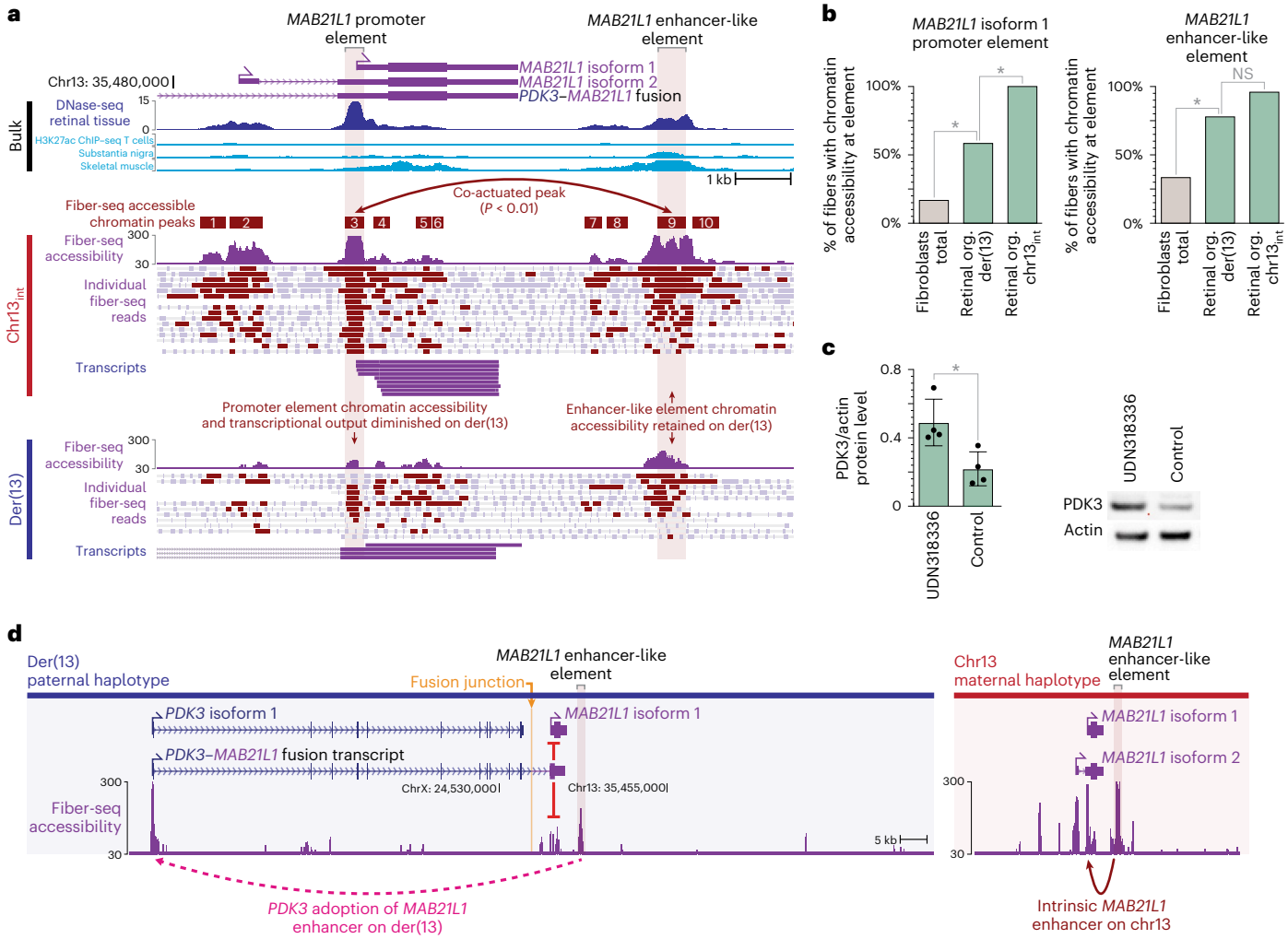


Fig. 4 | Placement of a *MAB21L1* enhancer-like element adjacent to *PDK3* along der(13). a, *MAB21L1* locus showing bulk DNase-seq and H3K27ac ChIP–seq (top), as well as haplotype-resolved Fiber-seq chromatin accessibility and full-length cDNA transcripts from patient-derived retinal organoids (middle and bottom). Chromatin accessibility peak pair with significant single-molecule co-actuated chromatin accessibility is shown, exposing a downstream enhancer-like element for the *MAB21L1* isoform 1 promoter. ChIP–seq track ranges are all from zerofold to tenfold change over control. **b**, Chromatin accessibility of the *MAB21L1* isoform 1 promoter, as well as the downstream enhancer-like element in patient-derived fibroblasts and retinal organoids (left), as well as their haplotype-specific accessibility in patient-derived retinal organoids (right). **P* = 0.0447, 0.0067 and 0.0088 (values in order; one-sided Fisher's exact test without adjustment for multiple comparisons). **c**, Bar plot and western blot showing PDK3 and β-actin protein levels within patient-derived retinal organoids, as well as age-matched control retinal organoids. Data are presented as mean ± s.e.m. **P* = 0.0183 (unpaired two-sided *t* test; *n* = 4) **d**, *MAB21L1* locus along the der(13) and chr13$_{int}$ chromosomes showing the placement of a strong *MAB21L1* enhancer-like element in proximity to the *PDK3* promoter selectively along the der(13) haplotype. ChIP–seq, chromatin immunoprecipitation followed by sequencing. NS, not significant.

transcriptome[20]. As most Mendelian conditions are caused by heterozygous or biallelic variants, this internally controlled approach is well suited for resolving the functional impact of disease-associated noncoding variants, in addition to phasing them relative to surrounding coding variants.

Application of this approach to an individual with a presumed Mendelian condition and a chromosome X;13-balanced translocation revealed that this translocation disrupted the functioning of four separate genes associated with the condition, with each gene being disrupted via a distinct mechanism. While the genomic data were able to identify one of these genes as being disrupted, the paired epigenetic and transcript data were essential for identifying the disruptions to the other three genes. This highlights the pivotal role that paired functional data can have when evaluating an individual with a presumed Mendelian condition. Furthermore, by pairing CpG methylation, chromatin and transcript information together, this approach enabled the identification of quite distinct disease mechanisms, including the formation of fusion transcripts, enhancer adoption, transcriptional readthrough silencing and inappropriate XCI of autosomal DNA.

This study defines a new Mendelian condition mediated by a single genetic variant, a 46,XX,t(X;13)(p22.11;q13.3)-balanced translocation. We demonstrate that this balanced translocation disrupts the functioning of four separate genes that are independently associated with different monogenic conditions (Online Mendelian Inheritance in Man (OMIM) numbers 614041, 619157, 618479 and 300905). The phenotype observed in the patient appears to arise from complex functional alterations to these genes in a manner that is specific to this balanced translocation (Table 2). Specifically, it is possible that the disrupted *NBEA* transcripts are resulting in a unique molecular mechanism that extends beyond that observed in individuals with *NBEA* haploinsufficiency, which is known to cause an autosomal dominant neurodevelopmental disorder[36] (OMIM 619157). Specifically, these disrupted NBEA transcripts are predicted to form a truncated NBEA protein that retains binding domains that could compete with that of the intact NBEA protein[31],

https://doi.org/10.1038/s41588-024-02067-0

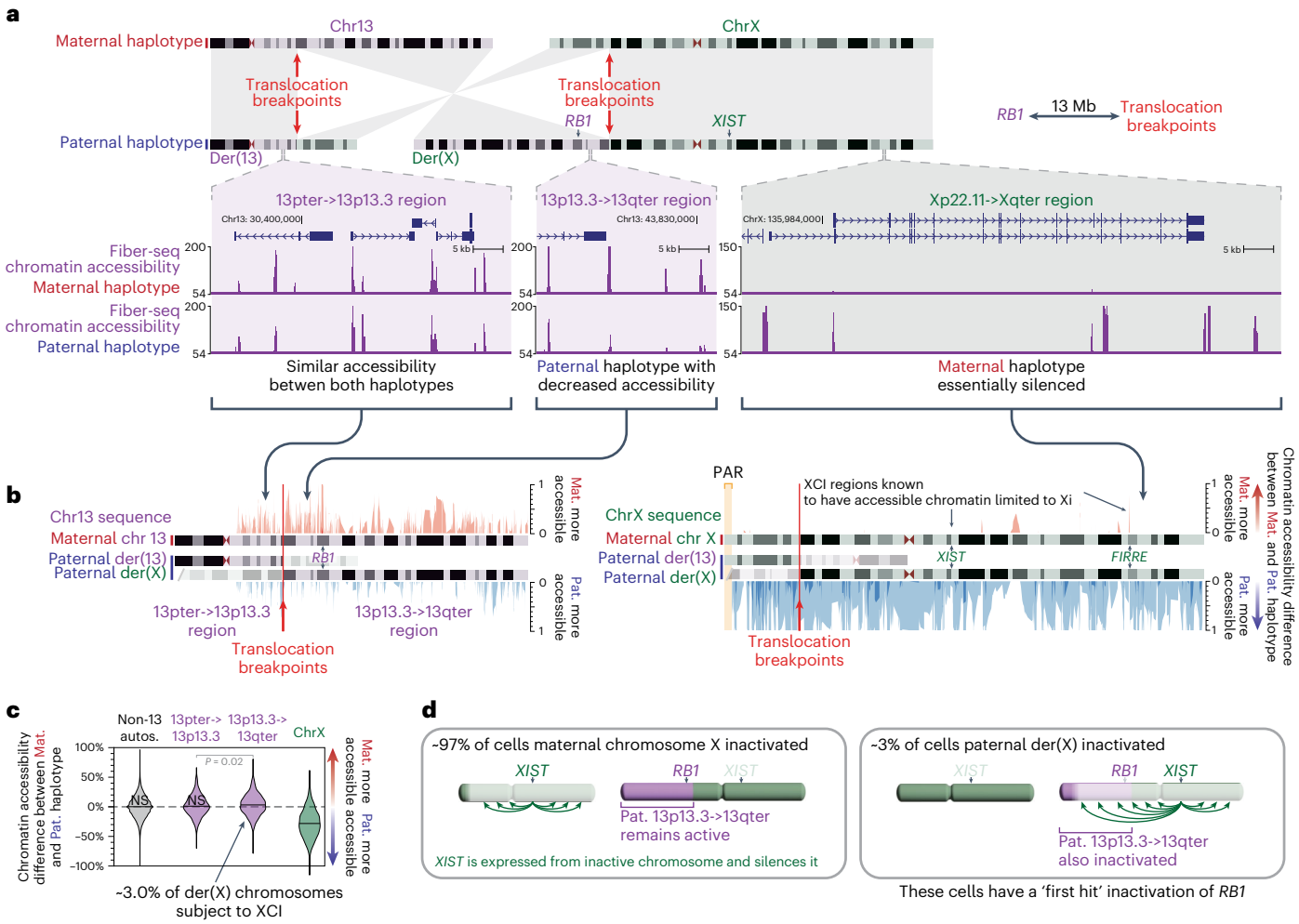


Fig. 5 | Inappropriate XCI of the *RB1* locus along der(X). a, Top, idiogram showing the chr13$_{int}$ and chrX$_{int}$, as well as the derivative chr13 and chrX in the proband. Translocation breakpoints and the location of the genes *RB1* and *XIST* are highlighted. Bottom, haplotype-resolved chromatin accessibility is displayed for loci across the der(13) and der(X) chromosomes. **b**, Difference in chromatin accessibility between the maternal and paternal haplotype across loci along chr13 (left) and chrX (right). Regions with more red than blue signals have more chromatin accessibility along chr13$_{int}$ versus der(13) or chrX$_{int}$ versus der(X). **c**, Swarm plot showing the overall haplotype imbalance in chromatin accessibility along autosomes (except for chr13), chrX and two portions of chr13. Specifically, the 13pter->13p13.3 region is present along chr13$_{int}$ and der(13), whereas the 13p13.3->13qter region is present along chr13$_{int}$ and der(X). *P* value ($P = 0.02$) was calculated using a one-sided Mann–Whitney *U* test without adjustments for multiple comparisons. **d**, Model showing inappropriate XCI of the autosomal region along der(X) that contains the *RB1* locus as the first hit for the development of bilateral retinoblastomas in this patient. Mat., maternal; pat., paternal.

thereby having either a gain-of-function or dominant-negative impact. Of note, the *NBEA*–chrX fusion transcripts were not originally detected with the lower transcript coverage in the 15:1 pooled multi-ome, prompting us to test a 9:1 pooling ratio, which resulted in a ~110% increase in full-length transcript reads with equivalent genomic coverage. Although the translocation results in the readthrough transcriptional silencing of *MAB21L1* on der(X), heterozygous *MAB21L1* loss of function (LOF) variants are not known to cause disease[37,38], consistent with the phenotype having little resemblance to that seen in individuals with biallelic *MAB21L1* LOF variants, which causes cerebellar, ocular, craniofacial and genital syndrome (OMIM 601280).

This patient's bilateral retinoblastomas appear to result from the inactivation of the *RB1* locus on der(X), which can be seen in 3–10% of her fibroblast cells, and may be higher or lower in different primary tissues. This is an extension of local XCI onto the autosomal portion of der(X) and results in the heterozygous LOF of *RB1* (OMIM 614041). This would serve as the 'first hit' in the progression to biallelic *RB1* LOF. This finding suggests that the clinical course would mirror that of individuals with mosaic 'first hit' *RB1* LOF variants, which is consistent with the clinical history at 3 years of age. In combination with the prior case report of an individual with bilateral retinoblastomas and a balanced 46,X,t(X;13)(q28;q14.1) translocation[30], these findings demonstrate that chromosome 13;X translocations involving either the p or the q arm of chromosome X can result in a predisposition to bilateral retinoblastomas. Notably, although the inactivation of the autosomal region of der(X) extended into the *RB1* locus, which is located 13.5 Mb away from the translocation breakpoint, not all of the der(X) appeared to be equally impacted by the spreading of XCI along der(X) (Extended Data Fig. 10c). This indicates that the location of the breakpoints on both X and 13 may have a significant role In the extent of autosomal DNA that is subjected to XCI and, as such, the phenotypic outcomes of the translocation.

Finally, our proband's translocation results in a unique combination of alterations to both the transcript and expression patterns of *PDK3* that is distinct from that of previously described *PDK3* gain-of-function and PDC loss-of-function variants. A gain-of-function variant in *PDK3* has previously been shown to cause X-linked dominant Charcot–Marie–Tooth disease (CMTX6)[39] (OMIM 300905), which is characterized by childhood-onset muscle weakness, muscle atrophy, sensory

Table 2 | Overview of molecular variants identified in UDN318336 via multi-ome long-read sequencing

Molecular event	Ome(s) required for identification	Associated clinical phenotypes in UDN318336	Proposed mechanism	Overlapping Mendelian condition
NBEA–chrX fusion transcripts	Genome; transcriptome	Polymicrogyria, SNHL, developmental delay, lactic acidosis and hypotonia.	*NBEA* haploinsufficiency and/or production of truncated protein	OMIM 619157
PDK3–MAB21L1 fusion kinase transcript	Genome; transcriptome		Overexpression of *PDK3* in tissue that endogenously expresses *MAB21L1*. Potentially altered regulation of PDK3–MAB21L1 fusion	OMIM 300905; OMIM 312170
PDK3 adoption of *MAB21L1* enhancer and subsequent *PDK3* ectopic GOE	Chromatin epigenome			
XCI of *RB1* locus	Chromatin epigenome	Bilateral retinoblastomas	'First hit' in the development of biallelic *RB1* LOF	OMIM 180200
Transcriptional readthrough silencing of *MAB21L1*	CpG methylome; chromatin epigenome; transcriptome	No effect on patient phenotype as only one *MAB21L1* haplotype impacted, with other haplotypes demonstrating intact gene regulation	N/A	OMIM 618479

abnormalities and SNHL. This gain-of-function variant increases PDK3 affinity for the PDC, locking PDC in an inactive state. In contrast, complete PDC deficiency (OMIM 312170) is associated with lactic acidosis, developmental delay, hypotonia, seizures, polymicrogyria and dysmorphic features[40], a phenotype that is more severe than that of CMTX6 likely owing to the tissue-specific expression of PDK3 protein, and the presence in many tissues of other kinases that regulate PDC. Our proband's der(X) results in an active fusion PDK3 kinase protein that has ectopic gain of expression (GOE) via enhancer adoption of an *MAB21L1* enhancer. As the intrinsic expression of *MAB21L1* is largely limited to neural tissues, it is likely that this enhancer adoption is selectively upregulating *PDK3* in the brain, which typically expresses *PDK3* at low levels.

Notably, certain genetic, epigenetic and transcriptomic features may be cell-selective. For example, multi-ome sequencing on the proband-derived fibroblasts uncovered the genomic translocation, the *MAB21l1* transcriptional readthrough silencing, the *NBEA*–chrX fusion transcript, the *PDK3–MAB21L1* fusion transcript and the inappropriate inactivation of autosomal genes along der(X). However, the *PDK3* enhancer adoption was only observable in the proband-derived retinal organoids owing to the tissue selectivity of *MAB21L1* and *PDK3* expression. Consequently, when applying functional studies, such as long-read multi-ome sequencing, it is important to consider what cell type is best suited for such a study.

Overall, we demonstrate that synchronized long-read multi-ome sequencing enables the highly accurate discovery of haplotype-phased genetic variants, in addition to the simultaneous identification of the functional consequence of these genetic variants. It is anticipated that this approach will enable clinicians and researchers to better understand how diverse classes of genetic variation mechanistically drive human diseases, as well as potential molecular targets for modulating these diseases.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-02067-0.

References

1. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
2. Merker, J. D. et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet. Med.* **20**, 159–163 (2018).
3. Hiatt, S. M. et al. Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *HGG Adv.* **2**, 100023 (2021).
4. Cohen, A. S. A. et al. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes. *Genet. Med.* **24**, 1336–1348 (2022).
5. Aganezov, S. et al. A complete reference genome improves analysis of human genetic variation. *Science* **376**, eabl3533 (2022).
6. Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
7. Aref-Eshghi, E. et al. Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *Am. J. Hum. Genet.* **106**, 356–370 (2020).
8. Kreitmaier, P., Katsoula, G. & Zeggini, E. Insights from multi-omics integration in complex disease primary tissues. *Trends Genet.* **39**, 46–58 (2023).
9. Kremer, L. S. et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8**, 15824 (2017).
10. Cummings, B. B. et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017).
11. Flusberg, B. A. et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
12. Sigurpalsdottir, B. D. et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol.* **25**, 69 (2024).
13. Sanford Kobayashi, E. et al. Approaches to long-read sequencing in a clinical setting to improve diagnostic rate. *Sci. Rep.* **12**, 16945 (2022).
14. Fukuda, H. et al. Father-to-offspring transmission of extremely long NOTCH2NLC repeat expansions with contractions: genetic and epigenetic profiling with long-read sequencing. *Clin. Epigenetics* **13**, 204 (2021).
15. Mahmoud, M., Doddapaneni, H., Timp, W. & Sedlazeck, F. J. PRINCESS: comprehensive detection of haplotype resolved SNVs, SVs, and methylation. *Genome Biol.* **22**, 268 (2021).
16. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
17. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).

18. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
19. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
20. Vollger, M. R. et al. A haplotype-resolved view of human gene regulation. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.14.599122 (2024).
21. Jha, A. et al. DNA-m6A calling and integrated long-read epigenetic and genetic analysis with *fibertools*. *Genome Res.* **34**, 1976–1986 (2024).
22. Dubocanin, D. et al. Conservation of chromatin organization within human and primate centromeres. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.20.537689 (2023).
23. Aneichyk, T. et al. Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* **172**, 897–909 (2018).
24. Stergachis, A. B. et al. Full-length isoform sequencing for resolving the molecular basis of Charcot–Marie–Tooth 2A. *Neurol. Genet.* **9**, e200090 (2023).
25. Al'Khafaji, A. M. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* **42**, 582–586 (2024).
26. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
27. Meuleman, W. et al. Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251 (2020).
28. Bache, I., Brondum-Nielsen, K. & Tommerup, N. Genetic counseling in adult carriers of a balanced chromosomal rearrangement ascertained in childhood: experiences from a nationwide reexamination of translocation carriers. *Genet. Med.* **9**, 185–187 (2007).
29. Giardino, D. et al. De novo balanced chromosome rearrangements in prenatal diagnosis. *Prenat. Diagn.* **29**, 257–265 (2009).
30. Tsutsumi, M. et al. A female patient with retinoblastoma and severe intellectual disability carrying an X;13 balanced translocation without rearrangement in the RB1 gene: a case report. *BMC Med. Genomics* **12**, 182 (2019).
31. Repetto, D. et al. Molecular dissection of neurobeachin function at excitatory synapses. *Front. Synaptic Neurosci.* **10**, 28 (2018).
32. Holness, M. J. & Sugden, M. C. Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem. Soc. Trans.* **31**, 1143–1151 (2003).
33. Heinz, S. et al. Transcription elongation can affect genome 3D structure. *Cell* **174**, 1522–1536 (2018).
34. Ligtenberg, M. J. L. et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. *Nat. Genet.* **41**, 112–117 (2009).
35. Lyon, M. F. X-chromosome inactivation and human genetic disease. *Acta Paediatr. Suppl.* **91**, 107–112 (2002).
36. Mulhern, M. S. et al. NBEA: developmental disease gene with early generalized epilepsy phenotypes. *Ann. Neurol.* **84**, 788–795 (2018).
37. Bruel, A. L. et al. Autosomal recessive truncating MAB21L1 mutation associated with a syndromic scrotal agenesis. *Clin. Genet.* **91**, 333–338 (2017).
38. Rad, A. et al. MAB21L1 loss of function causes a syndromic neurodevelopmental disorder with distinctive cerebellar, ocular, cranio facial and genital features (COFG syndrome). *J. Med. Genet.* **56**, 332–339 (2019).
39. Kennerson, M. L. et al. A new locus for X-linked dominant Charcot–Marie–Tooth disease (CMTX6) is caused by mutations in the pyruvate dehydrogenase kinase isoenzyme 3 (PDK3) gene. *Hum. Mol. Genet.* **22**, 1404–1416 (2013).
40. Patel, K. P., O'Brien, T. W., Subramony, S. H., Shuster, J. & Stacpoole, P. W. The spectrum of pyruvate dehydrogenase complex deficiency: clinical, biochemical and genetic features in 371 patients. *Mol. Genet. Metab.* **105**, 34–43 (2012).

Mitchell R. Vollger[1,2], Jonas Korlach[3], Kiara C. Eldred[4], Elliott Swanson[1], Jason G. Underwood[3], Stephanie C. Bohaczuk[1], Yizi Mao[2], Yong-Han H. Cheng[1], Jane Ranchalis[2], Elizabeth E. Blue[2,5,6], Ulrike Schwarze[7], Katherine M. Munson[1], Christopher T. Saunders[3], Aaron M. Wenger[3], Aimee Allworth[2], Sirisak Chanprasert[2], Brittney L. Duerden[8], Ian Glass[6,9], Martha Horike-Pyne[2], Michelle Kim[3], Kathleen A. Leppig[10], Ian J. McLaughlin[3], Jessica Ogawa[11], Elisabeth A. Rosenthal[2], Sam Sheppeard[2], Stephanie M. Sherman[2], Samuel Strohbehn[2], Amy L. Yuen[10], Andrew W. Stacey[12], University of Washington Center for Rare Disease Research*, Undiagnosed Diseases Network*, Thomas A. Reh[4], Peter H. Byers[2,7], Michael J. Bamshad[6,9], Fuki M. Hisama[2,6], Gail P. Jarvik[1,2,6], Yasemin Sancak[13], Katrina M. Dipple[6,9] & Andrew B. Stergachis[1,2,6] ✉

[1]University of Washington School of Medicine Department of Genome Sciences, Seattle, WA, USA. [2]University of Washington School of Medicine Department of Medicine, Seattle, WA, USA. [3]PacBio, Menlo Park, CA, USA. [4]University of Washington School of Medicine Department of Biological Structure, Seattle, WA, USA. [5]Institute for Public Health Genetics, University of Washington, Seattle, WA, USA. [6]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [7]University of Washington School of Medicine Department of Laboratory Medicine and Pathology, Seattle, WA, USA. [8]Mary Bridge/MultiCare, Tacoma, WA, USA. [9]University of Washington Department of Pediatrics, Seattle, WA, USA. [10]Genetic Services, Kaiser Permanente Washington, Seattle, WA, USA. [11]Case Western Reserve University, Cleveland, OH, USA. [12]University of Washington Department of Ophthalmology, Seattle, WA, USA. [13]University of Washington School of Medicine Department of Pharmacology, Seattle, WA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: absterga@uw.edu

University of Washington Center for Rare Disease Research

Elizabeth E. Blue[2,5,6], Martha Horike-Pyne[2], Michael J. Bamshad[6,9] & Gail P. Jarvik[1,2,6]

A full list of members and their affiliations appears in the Supplementary Information.

Undiagnosed Diseases Network

Jane Ranchalis[2], Elizabeth E. Blue[2,5,6], Ulrike Schwarze[7], Aimee Allworth[2], Sirisak Chanprasert[2], Ian Glass[6,9], Martha Horike-Pyne[2], Kathleen A. Leppig[10], Elisabeth A. Rosenthal[2], Sam Sheppeard[2], Stephanie M. Sherman[2], Samuel Strohbehn[2], Peter H. Byers[2,7], Michael J. Bamshad[6,9], Fuki M. Hisama[2,6], Gail P. Jarvik[1,2,6], Katrina M. Dipple[6,9] & Andrew B. Stergachis[1,2,6]

A full list of members and their affiliations appears in the Supplementary Information.

Methods

Ethics statement and patient consent

The patient was referred to the University of Washington and participated in an Institutional Review Board (IRB)-approved protocol. All participants or their legal guardians provided informed written consent and did not receive compensation. This study was approved by the National Institutes of Health (NIH) IRB (15HG0130).

gDNA preparation

GM12878 (HG001), GM24385 (HG002), HG02630 and GM20129 cells were obtained as cells from Coriell. UDN318336 cultured fibroblast cells were obtained via a skin punch biopsy from participant UDN318336. UDN318336 retinal organoids were generated by first converting UDN318336 cultured fibroblast cells into iPS cells using the CytoTune-iPS Sendai reprogramming method (Invitrogen) and then differentiating these iPS cells into retinal organoids as previously described[41]. Retinal organoids were collected at day 145, then made into a single-cell suspension using papain and DNAse for 20 min and then stopped with ovomucoid (Worthington Biochemical, LK003150) before permeabilization. Cells were permeabilized and treated with Hia5 enzyme as previously described[42]. Specifically, 1 million cells were washed with PBS and then resuspended in 60 µl buffer A (15 mM Tris (pH 8.0), 15 mM NaCl, 60 mM KCl, 1 mM EDTA (pH 8.0), 0.5 mM EGTA (pH 8.0) and 0.5 mM spermidine) and 60 µl of cold 2× lysis buffer (0.1% IGEPAL CA-630 in buffer A for GM12878 and GM24385; 0.2% IGEPAL CA-630 in buffer A for UDN318336 fibroblasts and retinal organoid cells) was added and mixed by gentle flicking and then kept on ice for 10 min. Samples were then pelleted, supernatant removed and then resuspended in 57.5 µl buffer A and moved to a 25 °C thermocycler. In total, 0.5 µl of Hia5 MTase (100 U) and 1.5 µl of 32 mM *S*-adenosylmethionine (NEB B9003S; 0.8 mM final concentration) were added, then carefully mixed by pipetting the volume up and down ten times with wide-bore tips. The reactions were incubated for 10 min at 25 °C, then stopped with 3 µl of 20% SDS (1% final concentration) and transferred to new 1.5 ml microfuge tubes. High-molecular-weight DNA was then extracted using the Promega Wizard HMW DNA Extraction Kit A2920. PacBio SMRTbell libraries were then constructed using these Fiber-seq-treated gDNAs following the manufacturer's SMRTbell prep kit 3.0 procedure.

RNA preparation

Bulk mRNA was isolated from GM24385 (HG002) cell cultures using the TRIzol Reagent and Phasemaker Tubes Complete System (Thermo Fisher Scientific). Bulk mRNA from GM12878 (HG001), HG02630, GM20129 and UDN318336 fibroblast RNA was isolated using the Qiagen RNeasy Plus Micro Kit (74034). RNA was subjected to cDNA synthesis as described in the standard PacBio Iso-Seq protocol, with the modification of using new cDNA amplification primers that combine the protocol's sequence with the MAS-seq capture primer sequence[25] to enable subsequent MAS-seq concatenation. Following cDNA amplification, MAS-seq libraries were generated by adapting the PacBio MAS-seq protocol to an 8-mer array.

Pooling Fiber-seq and MAS-seq

Fiber-seq libraries and bulk MAS-seq libraries from the same sample (that is, GM12878, GM24385 or UDN318336 fibroblasts) were pooled in a 15:1 molar ratio, respectively, and sequenced using a single SMRT Cell on the PacBio Revio system. Fiber-seq libraries and bulk MAS-seq libraries from the same sample (that is, HG02630 and GM20129) were pooled in a 9:1 molar ratio, respectively, and sequenced using a single SMRT Cell on the PacBio Revio system. UDN318336 retinal organoid Fiber-seq was sequenced on a Sequel II system.

Separating MAS-seq and Fiber-seq reads

HiFi read BAM files were computed and segregated into gDNA and MAS-seq reads on the Revio system, taking advantage of the different adapters used in the two library protocols. MAS-seq reads were deconcatenated into Iso-seq reads using SMRT Link (v12.0).

Identification of genomic variants

Variant calling was performed with DeepVariant for SNVs and indels[43] (1.5.0) and pbsv for SVs (https://github.com/PacificBiosciences/pbsv). Reads were haplotype-phased using a custom pipeline (https://github.com/mrvollger/k-mer-variant-phasing (v0.0.1)) that uses SNVs identified with DeepVariant and then runs a variant-based phaser (HiPhase v 1.2.1) to bin reads into phase blocks, which are assigned to either the maternal or paternal haplotype using parental short-read genome sequencing combined with the trio *k*-mer-based phaser meryl. F1 statistics were calculated using GIAB variant calls from HG001 and HG002 as the gold standard. Specifically, for SNV and indel benchmarking, we used the v4.2.1 GRCh38 benchmark for both NA12878_HG001 and HG002_NA24385_son. For SV benchmarking in HG002, we used the NIST_SV_v0.6 benchmark lifted to GRCh38 with NCBI Remap. De novo assembly was performed using hifiasm[44]; the concordance QVs of the assemblies against HG001 and HG002 GIAB references were computed with yak v0.1.

Identification of CpG methylation

Base-level CpG methylation was called using jasmine (PacBio). The percentage of CpG methylation at each genomic position was identified from a pileup of reads using pb-CpG-tools (https://github.com/PacificBiosciences/pb-CpG-tools). Pearson correlations for HG001 and HG002 were calculated using methylKit[45], compared to WGBS data from EpiQC[46], and ONT data from Epi2me (https://labs.epi2me.io/giab-2023.05/, using the 'super accuracy' analysis level 60× BAM files).

Identification of chromatin architectures

m^{6}A methylation and methyltransferase-sensitive patches (MSPs) were identified using fibertools-rs[21]. MSPs correspond to regions between nucleosome footprints and represent a combination of internucleosomal linker regions, as well as accessible regulatory elements. MSPs were then processed using the Fiber-seq inferred regulatory element (FIRE) pipeline[20] (https://github.com/fiberseq/fiberseq-fire, multi-ome-v0.1). Briefly, the FIRE pipeline using a semi-supervised machine learning approach[47] trained to identify the difference in signal between a regulatory element and an internucleosomal linker region. This method assigns an estimated precision based on validation data to every MSP, indicating the precision of the model for assigning a given MSP as a regulatory element versus an internucleosomal linker region. We then define significant MSPs (precision over 90%) to be FIREs and aggregate the FIRE signal over multiple Fiber-seq molecules to create a track representing genome-wide accessibility using the following formula:

$$S_g = -\frac{10}{C_g}\sum_{i=1}^{C_g}\log_{10}(1-p_i),$$

where S is the aggregate FIRE signal at a genomic position (g), C is the genomic coverage of Fiber-seq data at that position and p is the precision assigned to an individual FIRE. We then apply a genome-wide Bonferroni correction (at $\alpha = 0.01$) to call peaks within this signal track. Known imprinted loci were identified from ref. 48.

Identification of full-length transcripts

Transcriptome statistics were generated using the Iso-seq analysis workflow in SMRT Link (v12.0). Fusion transcripts were identified using pbfusion (v0.1). Each full-length transcript read was considered a 'transcript' in this manuscript, with the understanding that some transcripts may represent PCR duplicates.

Sanger validation of der(X) fusion junction

The der(X) fusion junction was amplified by PCR with primers annealing to sequences on chromosomes 13 and X that yielded a 403 bp

amplicon (Supplementary Table 3). The product was checked by PAGE and sequenced in both directions with BigDye Terminator (v3.1) on the ABI 3500 Genetic Analyzer. The data were analyzed with Chromas software.

MAB21L1 expression

MAB21L1 GTEx expression data were accessed from https://www.gtexportal.org/home/gene/MAB21L1. *NBEA*, *MAB21L1* and *PDK3* expression during retinal organoid differentiation was quantified using publicly available data[49]. *MAB21L1* full-length transcripts from UDN318336 were haplotype-resolved using heterozygous variants contained within the transcript.

Identification of *MAB21L1* putative enhancer peaks

Codependency scores for every neighboring FIRE peak with the *MAB21L1* promoter using UDN318336 retinal organoid Fiber-seq data were calculated. Specifically, we defined codependency scores as the observed rate of co-accessibility (that is, accessible at both peaks along an individual chromatin fiber) minus the expected rate of co-accessibility given independence between the two peaks. Specifically, we identified overlapping chromatin fibers and accessible patches (FIRE precision ≤0.05) using bedtools intersect (v2.30.0) and calculated the proportion of fibers that are accessible at each peak. We performed all subsequent codependency analyses in Python (v3.9.12). For each pair of FIRE peaks, we calculated the expected co-accessibility as the product of their accessible proportions, while the observed co-accessibility was calculated as the proportion of fibers spanning both peaks that are accessible at each. We quantified the significance of this codependency using Fisher's exact test. H3K27ac chromatin immunoprecipitation followed by sequencing (ENCFF365WHC, ENCFF522ANT and ENCFF962HWM) and retinal DNase-seq (ENCFF064DIM) data were downloaded from ENCODE.

Quantification of chromosomes X and 13 allelic imbalance

To quantify chromatin allelic imbalance on each chromosome, we measured the number of accessible molecules (FIRE precision ≤0.10) for each haplotype at every FIRE peak on chromosomes 13 and X compared to the whole genome. We then calculated the difference in percent accessibility between the maternal and paternal haplotypes at every FIRE peak and built a distribution of these values to identify regions of imbalance. We tested for statistical significance by comparing the distributions of chromosomes X and 13 against the rest of the genome using a Mann–Whitney *U* test. FIRE peaks with unusually low coverage (coverage less than the median minus three s.d. or coverage less than ten reads) were excluded from the calculation to remove mapping, assembly and phasing artifacts. Code to repeat this analysis is made available on GitHub as part of the Fiber-seq FIRE pipeline.

Similarly, to quantify the transcriptomic allelic imbalance on each chromosome, we measured the maternal–paternal ratio of phased Iso-seq reads for genes with at least 30% phasing and at least ten Iso-seq reads. We then calculated the percentage difference in the number of transcripts between the maternal and paternal haplotypes for every gene and built a distribution of these values to identify regions of imbalance. We tested for statistical significance by comparing the distributions of chromosomes X and 13 against the rest of the genome using a two-sided Mann–Whitney *U* test. Code to repeat this analysis is made available on GitHub as part of the Fiber-seq FIRE pipeline. Individual Iso-seq reads were phased using whatshap haplotag (v2.3) with default settings.

iPS cell culture

iPS cells were maintained by dissociation and propagation in StemFlex media (Thermo Fisher Scientific, A3349401) on Matrigel (Corning, CLS354277)-coated plates at 37 °C in 5% CO_2. iPS cell colonies were passaged by dissociation with ReLeSR (STEMCELL Technologies, 100-0484).

Retinal organoid culture

Retinal organoids were grown as previously described[41] with minor alterations mentioned herein. The cell culture media used were neural induction media (NIM)—484.5 ml DMEM/F12 (Life Technologies, 11330-057), 5 ml N2 supplement (Life Technologies, 17502048), 5 ml MEM Non-Essential Amino Acids Solution (NEAA) (Life Technologies, 11140050) and 5 ml penicillin–streptomycin (Life Technologies, 15240062). Retinal differentiation media (RDM)—240 ml DMEM/F12 (Life Technologies, 11330-057), 240 ml DMEM (Life Technologies, 12430062), 10 ml B27 supplement (Life Technologies, 17504001), 5 ml MEM NEAA (Life Technologies, 11140050), 5 ml penicillin–streptomycin (Life Technologies, 15240062) + 10% FBS (Corning, 35-011 CV) based on protocol needs. In brief, confluent stem cell colonies were lifted from the plate using dispase (2 mg ml^{-1}; Life Technologies, 17105041) for 5–10 min and then forcefully removed by pipetting 2 ml per well of DMEM (Life Technologies, 12430062) directly onto cells using a 1 ml pipette. Colonies were removed from the six-well plates and allowed to settle to the bottom of a 15 ml collection tube by gravity. The supernatant was removed, and colonies were transferred to a T25 tissue culture flask in a 3:1 StemFlex:NIM mix for a total of 10 ml per flask. Full media change was performed the following day (day 1) with 10 ml of 1:2 StemFlex:NIM. On days 2–5, a full media change was performed and replaced with 10 ml of NIM. On day 7, a full media change was performed and replaced with 10 ml of NIM, and BMP4 (1.5 nM; R&D Systems, 314-BP-050) was added to the media to a final concentration of 1.5 nM. On day 8, EBs were then evenly distributed between the six wells of a six-well plate. The media was not changed, and EBs were moved and cultured in the BMP4-containing NIM media they were in. In total, 200 ml of FBS was added to allow for EBs to stick to the bottom of the plate. On days 10–20, every other day, a full media change was performed with fresh NIM. On days 14 and 16, a full media change was performed with NIM, and the signaling factor modulator CHIR99021 (3 mM; Cedarlane Laboratories, 04-0004-02) was added. EBs were then manually lifted from the plates on day 20 or 21 by forcefully pipetting 2 ml per well of DMEM directly onto EBs using a 1 ml pipette. EBs were removed from the six-well plates and allowed to settle to the bottom of a 15 ml collection tube by gravity. From this point on, we consider the cell's retinal organoids. The supernatant was removed, and then organoids were resuspended in RDM with 10% FBS. From this point on, cells were maintained with a full media change of RDM + 10% FBS every 2–3 days.

Evaluation of transcripts subjected to NMD

RNA from patient-derived fibroblast cells was separately isolated from untreated cells, as well as cells treated with 0.1 mg ml^{-1} cycloheximide for 6 h. These RNA samples were subjected to cDNA synthesis, as detailed above, and then pooled and concatenated as above and sequenced on a single SMRT cell.

Western blotting

Retinal organoid tissue was derived as described above from UDN318336 iPS cells (day 72) and from AICS-0088 cl.83 iPS cells (control sample, day 66). Fibroblasts cell lines used were UDN318336, UND374570 (control 1) and UND359892 (control 2). Cells were lysed with radioimmunoprecipitation assay (RIPA) lysis buffer supplemented with protease inhibitor tablets and processed as previously described[50]. Samples were normalized to 0.4–1 μg μl^{-1}, and 10 μl of the sample was loaded on an SDS–PAGE gel. Antibodies (PDHA1—ABclonal, A13687; p-PDHA1 S293—ABclonal, AP1022; PDK3—ABclonal, A8028; PDK3—ABclonal, A12480 and β-actin—Cell Signaling Technology, 3700S) were used at 1:1,000 dilution for western blotting in 1% bovine serum albumin prepared in Tris-buffered saline with 0.1% Tween. Horseradish

peroxidase-conjugated secondary antibodies were used at 1:10,000 dilution, and blots were imaged using iBrightCL1000. Quantification of western blotting was done using FIJI. The smallest rectangle possible was drawn around a band for a protein and then was 'measured' using FIJI's measurement tool. This same-sized rectangle was used to measure all other bands.

Calcium uptake assay

The cells used were UDN318336 iPS cells and AICS-0088 cl.83 iPS cells (control). Calcium uptake assays were done as previously described[50]. Cell number was determined using a Coulter counter. One million cells were used for each calcium uptake assay. The rate of calcium uptake was determined using data from the linear range of the uptake curves (40–60 s). Data from three replicates were normalized to the average of three control samples. A two-tailed *t* test was used for statistical analysis.

Pyruvate dehydrogenase activity assay

Stem cells AICS-0088 cl.83 iPS cells (control) and UDN318336 iPS cells were grown to 90% confluency in StemFlex media and then dissociated using Accutase for 7 min at 37 °C. Cells were collected with equal volume StemFlex media to Accutase and then pelleted at 700*g* for 5 min. Cells were then resuspended in 5 ml of StemFlex media for counting. After the total cell number was determined, cells were spun down again at 1,000*g* for 3 min and resuspended in PDH assay buffer at 1 million cells per 20 µl. The pyruvate dehydrogenase activity assay kit (Sigma-Aldrich, MAK183) was used, and the procedure was followed as the documentation details. Cells were assayed at a concentration of 0.75 million cells per well. Plates were assayed in a Synergy H1 plate reader at 37 °C. Each sample was run in triplicate. Mean with s.d. is displayed.

Chromatin conformation capture (3C) assay

3C was performed according to previous protocols[51,52]. In brief, patient iPS cells or 60-day-old retinal organoids were washed in Dulbecco's Phosphate-Buffered Saline (DPBS) (no calcium, no magnesium; Thermo Fisher Scientific, 14190144). In total, 1 ml of accutase (Sigma-Aldrich, A6964-100ML) was added to the cultures. The cultures were incubated at 37 °C on a nutator for 7 min. After 7 min, organoids were pipetted up and down as needed (every 5–10 min). The extent of dissociation was determined by bright-field microscopy. The dissociation was complete by 30–45 min after Accutase addition when clusters of cells could no longer be seen. The cells were then spun down at 700*g* for 5 min and then resuspended in a single-cell suspension in RDM media for processing.

The 10 million iPS cells or 3 million cells from retinal organoids were fixed in 10 ml of 1% formaldehyde diluted in PBS for 10 min. The reaction was quenched with 540 µl of 2.5 M glycine, incubated at room temperature for 5 min, on ice for 15 min, centrifuged at 800*g* for 10 min and lysed in 10 mM Tris (pH 8), 10 mM NaCl and 0.2% IGEPAL with protease inhibitors (Promega, G6521). Cells were incubated on ice for 20 min and homogenized in a Dounce homogenizer (pestle A). Pellets were collected by centrifugation at 2,200*g* in a microfuge for 5 min, washed twice with 1.1× DpnII buffer (NEB, B0543SVIAL) and resuspended in 1.1× DpnII buffer. In total, 20% SDS was added to a final concentration of 0.3%, and nuclei were incubated at 65 °C for 10 min. Triton X-100 was added for a final concentration of 2%, and then the sample was incubated at 37 °C for 1 h with shaking (900 rpm). In total, 10 µl of predigested sample was collected for digest efficiency. In total, 400 U (iPS cells) or 200 U (retinoid cells) of DpnII (NEB R0543T) was added, and the sample was incubated overnight at 37 °C with shaking (900 rpm). In total, 10 µl of postdigested sample was collected for digest efficiency. SDS was added to a final concentration of 2%, and the sample was incubated at 65 °C for 30 min to inactivate the restriction enzyme. The sample was added to a solution containing 1.15× ligation buffer (Promega, C1263), 1% Triton X-100 and 100 U (iPSCs) or 50 U (retinoid cells) T4 ligase (Promega, M1794) for a final volume of approximately 1 ml per 1M cells and incubated for 2–3 h at 16 °C. In total, 300 µg (iPS cells) or 150 µg (retinoid cells) of proteinase K was added to the sample, incubated at 65 °C for 4 h and then treated with an additional 300 or 150 µg of proteinase K followed by overnight incubation at 65 °C. In total, 300 µg (iPS cell) or 150 µg (retinoid) of RNase A was added to each sample, followed by incubation at 37 °C for 45 min. One sample volume of phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol; Thermo Fisher Scientific, 15593031) was added, mixed vigorously and centrifuged at 4,300*g* for 15 min. The aqueous phase was transferred to a new tube and combined with 400 or 200 µg of glycogen (Thermo Fisher Scientific, R0561), 1/10 volume sodium acetate (pH 5.2) and 2.6 volumes of ice-cold ethanol, incubated overnight for 4 h at −80 °C, centrifuged for 30 min at 4 °C, rinsed with 70% ethanol, air dried and resuspended in 300 µl of elution buffer (PacBio 101-633-500). The sample was added to a Nanosep centrifugal filter (MWCO 3 kDa; Pall, OD003C33) and centrifuged at 16,000*g* for 10 min or until ~20 µl remained above the filter. The sample was diluted with 400 µl and centrifuged as described an additional three times. The predigest and postdigest samples were treated with 20 µg of proteinase K and 1 µg of RNase A and purified with phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol) and ethanol precipitation, as described.

qPCR primers were designed with Primer3 to anneal to the 3′ ends of the predicted DpnII restriction digest products (Supplementary Table 3). The MABenh primer and a custom MGB Taqman probe (Supplementary Table 3; Applied Biosystems, 4316034) designed to anneal to the *MAB21L1* enhancer peak were common to each reaction. qPCR was performed with TaqMan Fast Advanced Master Mix for qPCR, no uracil-*N*-glycoslyase (Thermo Fisher Scientific, A44360), according to the manufacturer's instructions. A standard curve consisting of BAC DNA from CH17-6D16 (chrX) and CH17-139G1 (chr13) clones (https://bacpacresources.org/library.php?id=231) digested with Sau3AI (NEB, R0169S) and ligated according to previously published methods[52] was included for each primer set. Sau3AI is an isoschizomer of DpnII but is not sensitive to bacterial adenine methylation. Enrichment was calculated from the standard curve for each primer pair and scaled by the enrichment of the S0/MABenh product for relative enrichment.

Viral expression of PDK3 fusion in HEK293T cells

pLYS1 lentiviral vectors were generated that contained either the wild-type PDK3 isoform 1 protein sequence or the PDK3–MAB21L1 protein sequence. HEK293T cells were infected with virus for stable expression using an empty control vector, the PDK3 vector or the PDK3–MAB21L1 vector. Lysates from these stable cell lines were generated and used for western blotting to detect PDK3, PDH and phospho-PDH levels.

Sample sizes, statistical assumptions and randomization

No statistical methods were used to predetermine sample sizes because they are irrelevant to this study, as this is a single case report. Where applied data met the assumptions of the statistical tests used, including whether normality and equal variances were formally tested. We used an *F* test to compare variances to ensure they met the statistical assumptions. Blinding and randomization were not relevant to this study, as this is a single case report.

Data collection

Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw and processed sequencing data generated for GM12878, GM23485, GM20129 and HG02630 in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession PRJNA1124997. Restrictions apply to the availability of some data generated or analyzed during this study to preserve subject confidentiality. Genomic data for our case report are accessible to the scientific community through the UDN (dbGaP). Individuals interested in accessing UDN data through dbGaP should submit a data access request. Detailed instructions for this process can be found on the NIH Scientific Data Sharing website (https://sharing.nih.gov/accessing-data/accessing-genomic-data/how-to-request-and-access-datasets-from-dbgap). Cell lines obtained from the National Institute of General Medical Sciences Human Genetic Cell Repository at the Coriell Institute for Medical Research include GM12878, GM23485, GM20129 and HG02630. The UDN318336 fibroblast line was obtained directly from UDN participant UDN318336. Hia5 enzyme and plasmid are available upon request. ONT data for the CpG comparison are available from Epi2me (https://labs.epi2me.io/giab-2023.05/, super accuracy analysis level 60× BAM files).

Processed FIRE results for GM12878, GM23485, GM20129, HG02630, UDN318336 and retinal organoid derived from UDN318336 are publicly available through https://doi.org/10.5281/zenodo.14511246, as well as through GitHub (https://github.com/StergachisLab/Fiber-seq-publication-data). Source data are provided with this paper.

Code availability

All software used in the study is publicly available. Specifically, the custom FIRE pipeline multi-ome-v0.1 (ref. 53; https://doi.org/10.5281/zenodo.12701477 and https://github.com/fiberseq/FIRE) and the custom *k*-mer-variant-phasing pipeline v0.0.1 (ref. 54; https://doi.org/10.5281/zenodo.10655504 and https://github.com/mrvollger/k-mer-variant-phasing) are publicly available and accessioned on Zenodo.

References

41. Tresenrider, A. et al. Single-cell sequencing of individual retinal organoids reveals determinants of cell-fate heterogeneity. *Cell Rep. Methods* **3**, 100548 (2023).
42. Dubocanin, D. et al. Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin. Preprint at *bioRxiv* https://doi.org/10.1101/2022.05.09.491186 (2022).
43. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983 (2018).
44. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
45. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
46. Foox, J. et al. The SEQC2 epigenomics quality control (EpiQC) study. *Genome Biol.* **22**, 332 (2021).
47. Fondrie, W. E. & Noble, W. S. Machine learning strategy that leverages large data sets to boost statistical power in small-scale experiments. *J. Proteome. Res.* **19**, 1267–1274 (2020).
48. Akbari, V. et al. Genome-wide detection of imprinted differentially methylated regions using nanopore sequencing. *eLife* **11**, e77898 (2022).
49. Eldred, K. C. et al. Thyroid hormone signaling specifies cone subtypes in human retinal organoids. *Science* **362**, eaau6348 (2018).
50. MacEwen, M. J. S. et al. Evolutionary divergence reveals the molecular basis of EMRE dependence of the human MCU. *Life Sci. Alliance* **3**, e202000718 (2020).
51. Hagege, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
52. Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods* **58**, 192–203 (2012).
53. Vollger, M. R. et al. Release version for the multi-ome manuscript (multi-ome-v0.1). *Zenodo* https://doi.org/10.5281/zenodo.12701477 (2024).
54. Vollger, M. R. et al. mrvollger/k-mer-variant-phasing: 0.0.1 (0.0.1). *Zenodo* https://doi.org/10.5281/zenodo.10655527 (2024).

Acknowledgements

We thank C. C. Lambert, A. Souppe, H. Dhillon and S. Zhang (all from PacBio) for assistance with library and sequencing preparations. A.B.S. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew Biomedical Scholar. This study was supported by the NIH (grants 1DP5OD029630 and 1U01HG010233), in addition to funds from the Collagen Diagnostic Laboratory, University of Washington, and the Brotman Baty Institute for Precision Medicine. Sequencing and data analysis were provided by the University of Washington Center for Rare Disease Research, with support from National Human Genome Research Institute (grants U01 HG011744 and U24 HG011746). Research from the UDN reported in this publication was supported by the National Institute of Neurological Disorders and Stroke of the NIH (awards U01HG010233, U01HG007942 and U01HG007530). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. M.R.V. was supported by a training grant (T32) from the NIH (2T32GM007454-46). The Eichler Lab provided the CH17-6D16 (chrX) and CH17-139G1 (chr13) BAC clones.

Author contributions

A.B.S., M.R.V. and J.K. conceived the overall method design, coordinated activities from co-authors and wrote the first draft of the manuscript. M.R.V., E.S. and A.B.S. contributed to the analysis of Fiber-seq and MAS-seq data. K.C.E. and T.A.R. contributed to the generation of patient-derived retinal organoids, as well as the evaluation of RNA-seq data in retinal organoids. Y.S. contributed to the western blots, as well as mitochondrial functional studies. Y.M. created the Hia5 enzyme. E.E.B., S. Sheppeard, S. Strohbehn, S.M.S., E.A.R. and A.M.W. contributed to the analysis of genomic data. U.S. and P.H.B. contributed to the Sanger validation of the translocation breakpoints. K.M.D., G.P.J., F.M.H., M.J.B., A.L.Y., J.O., K.A.L., I.G., B.L.D., S.C. and A.A. contributed to patient phenotyping. M.H.-P. contributed to getting IRB approval for this study. J.R., A.B.S., K.C.E., M.K. and Y.H.C. contributed to the generation of Fiber-seq and mRNA samples. J.G.U. contributed to the mRNA concatenation protocol development and transcriptome data analysis. C.T.S. and A.M.W. contributed to CpG methylation calling and IGV visualization. I.J.M. and K.M.M. coordinated and performed the sequencing runs. All authors revised the manuscript.

Competing interests

J.K., J.G.U., C.T.S., A.M.W., M.K. and I.J.M. are full-time employees at PacBio, a company developing single-molecule sequencing technologies. A.B.S. is a co-inventor on a patent relating to the Fiber-seq method (US17/995,058). The remaining authors declare no competing interests.

Additional information

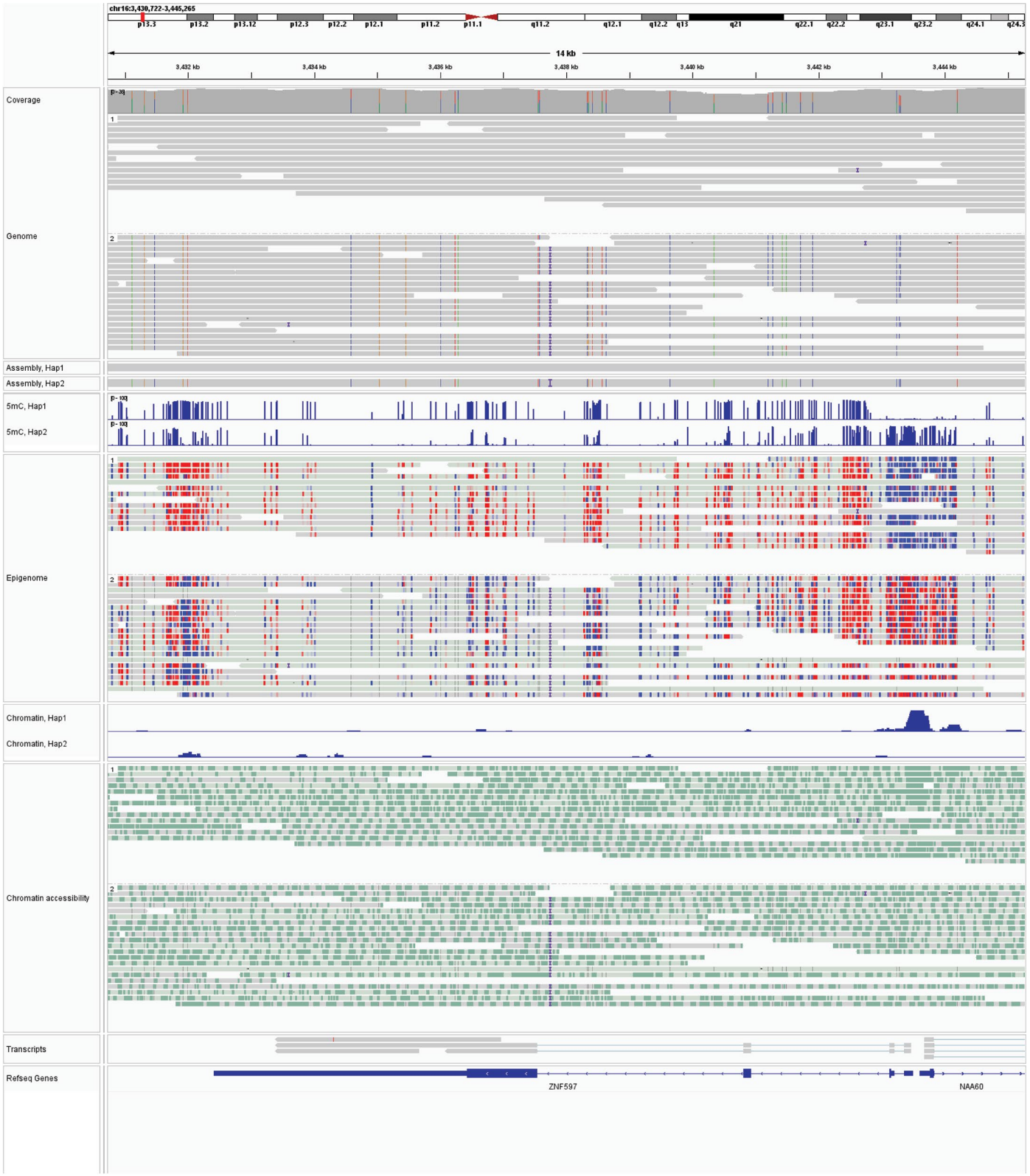
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-02067-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-02067-0.

Correspondence and requests for materials should be addressed to Andrew B. Stergachis.

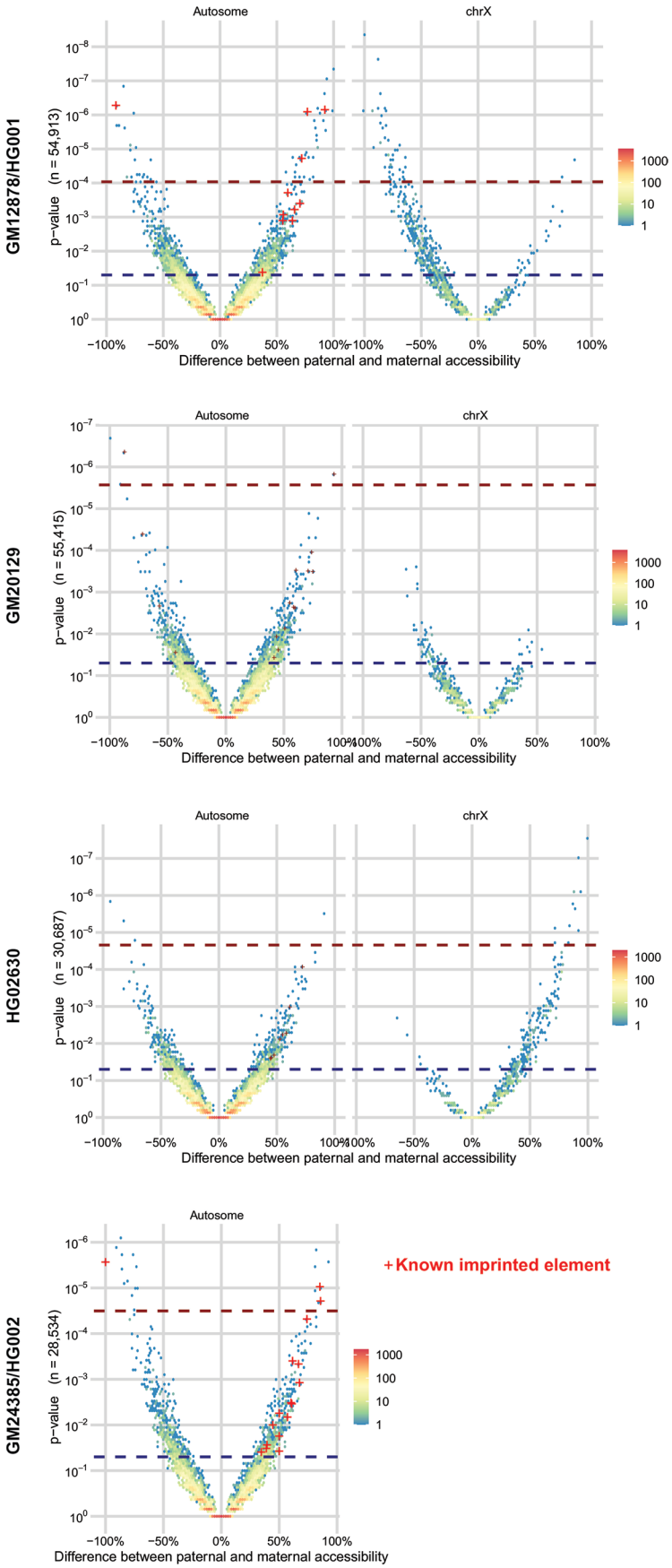
Peer review information *Nature Genetics* thanks Sarah Stenton and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-024-02067-0



Extended Data Fig. 1 | IGV view of integrated long-read multi-ome data. IGV view showing an example genomic region for GM12878 (from top to bottom) the haplotype-resolved genome, CpG methylome, chromatin and full-length transcript annotations.

https://doi.org/10.1038/s41588-024-02067-0

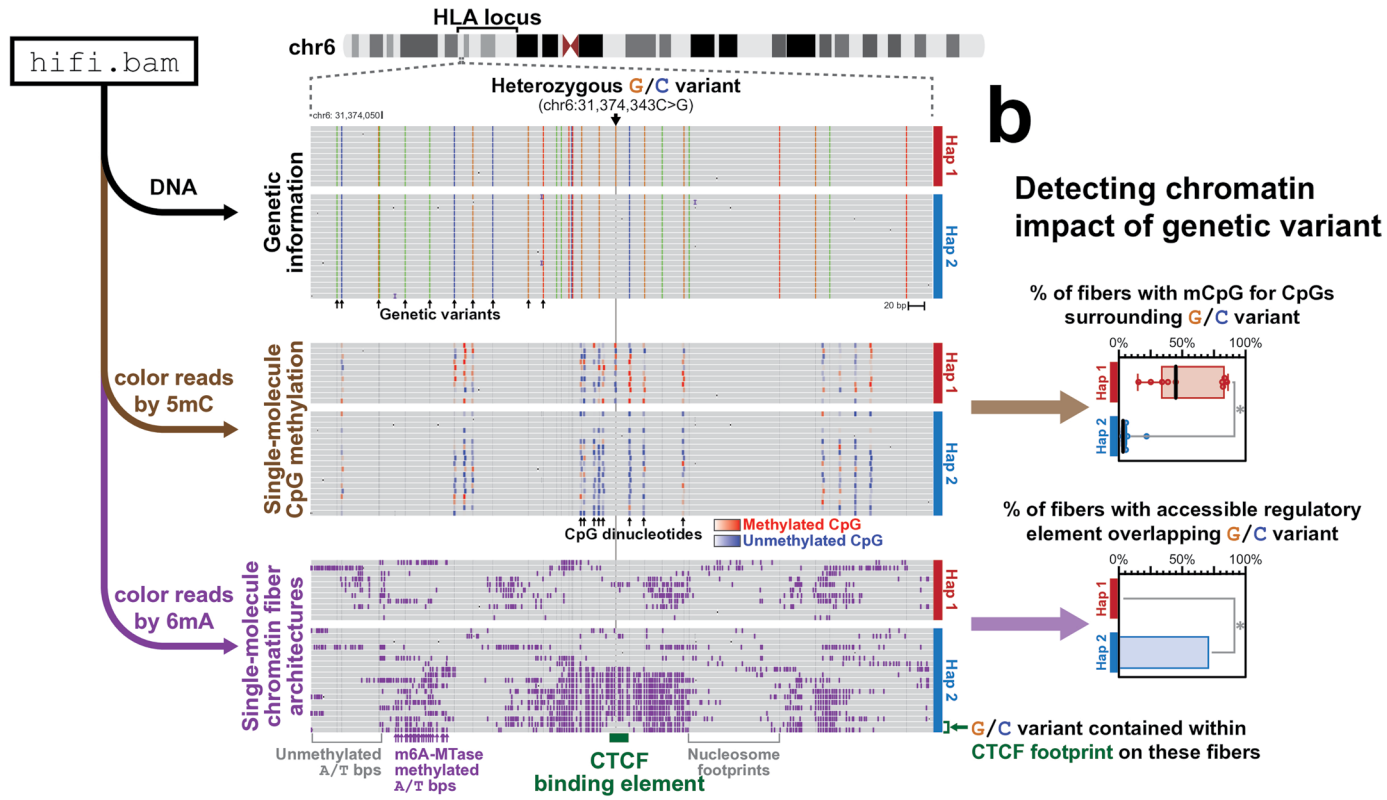


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-024-02067-0

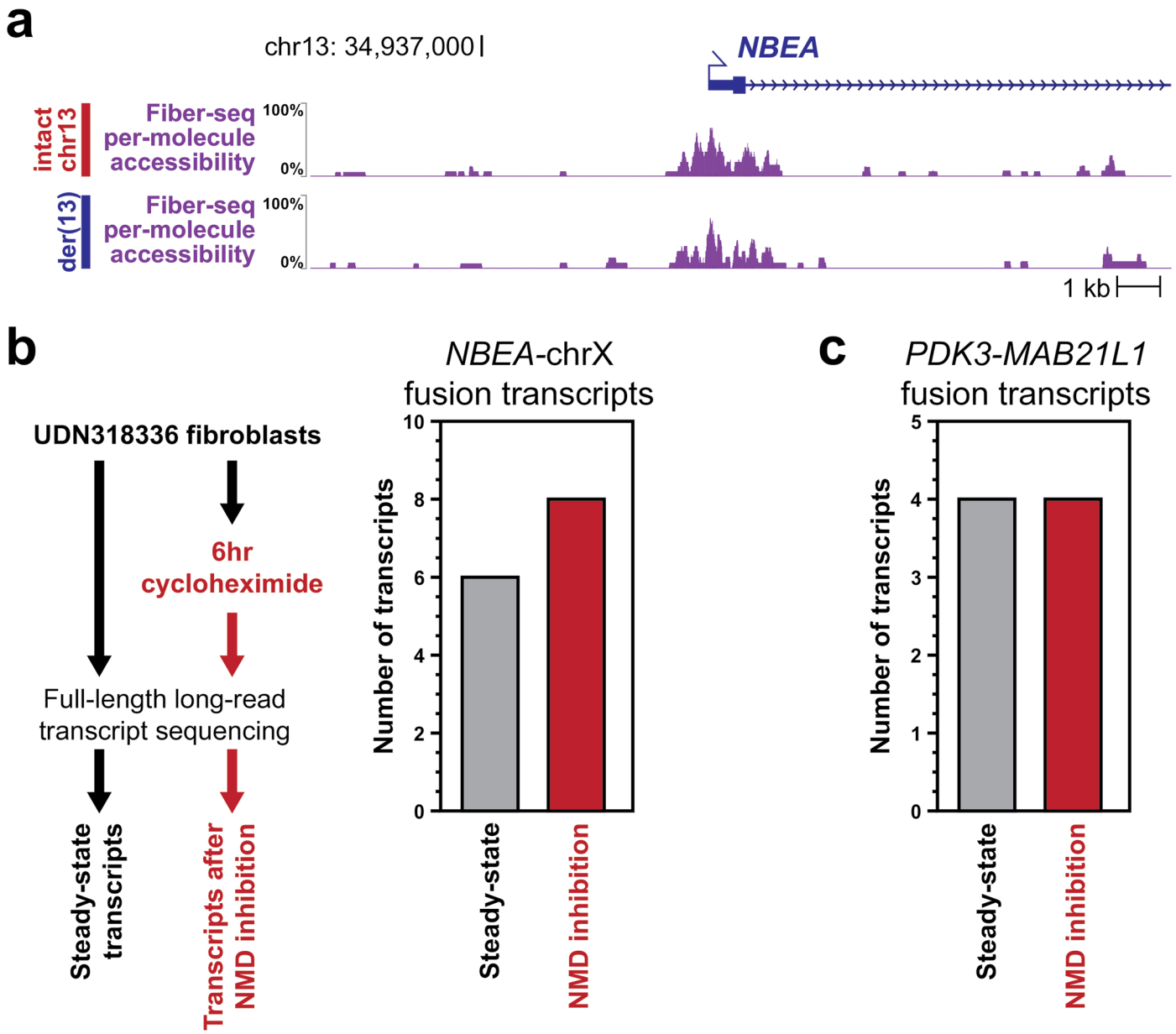
Extended Data Fig. 2 | Haplotype-specific chromatin architectures. Volcano plot showing the absolute difference in the percentage of chromatin fibers with chromatin accessibility for each peak genome-wide between the paternal and maternal haplotypes for GM12878, GM20129, HG02630 and GM24385. Peaks are divided into whether they are present on an autosome or the X chromosome, with the exception of GM24385, which is 46,XY. P value was calculated using a two-sided Fisher's exact test without adjustment for multiple comparisons. Blue dash represents nominal significance line ($p < 0.05$), and red dash represents the Benjamini–Hochberg FDR correction significance line (FDR < 0.05). Peaks corresponding to known imprinted loci with nominally significant scores are denoted by red crosses.

https://doi.org/10.1038/s41588-024-02067-0



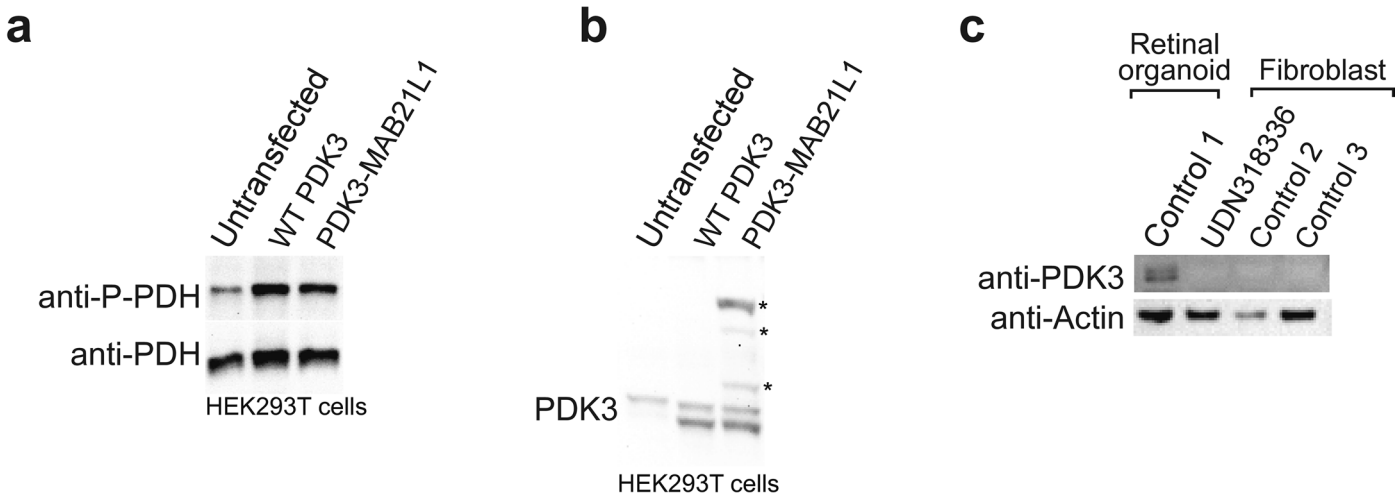
Extended Data Fig. 3 | Identification of haplotype-specific chromatin architectures within HLA locus. a, View of single-molecule haplotype-resolved genetic information, CpG methylation information and chromatin accessibility information for a heterozygous single-nucleotide variant (SNP) identified for GM12878 within the HLA locus. Note that all information is derived from the same sequencing reads. Denoted below is the location of a predicted CTCF binding element, and immediately above it are two fibers that demonstrate single-molecule protein occupancy at this site. **b**, Quantification of CpG methylation surrounding this SNP (top), as well as the percentage of fibers with a FIRE element overlapping that SNP (bottom), by haplotype. The box bounds the interquartile range (IQR) divided by the median, and the whiskers extend to the minima and maxima of the data. *Top, p value = 0.00077 (one-sided paired t test without adjustment for multiple comparisons, n = 8). *Bottom, p value = 0.00061 (one-sided Fisher's exact test without adjustment for multiple comparisons).

https://doi.org/10.1038/s41588-024-02067-0



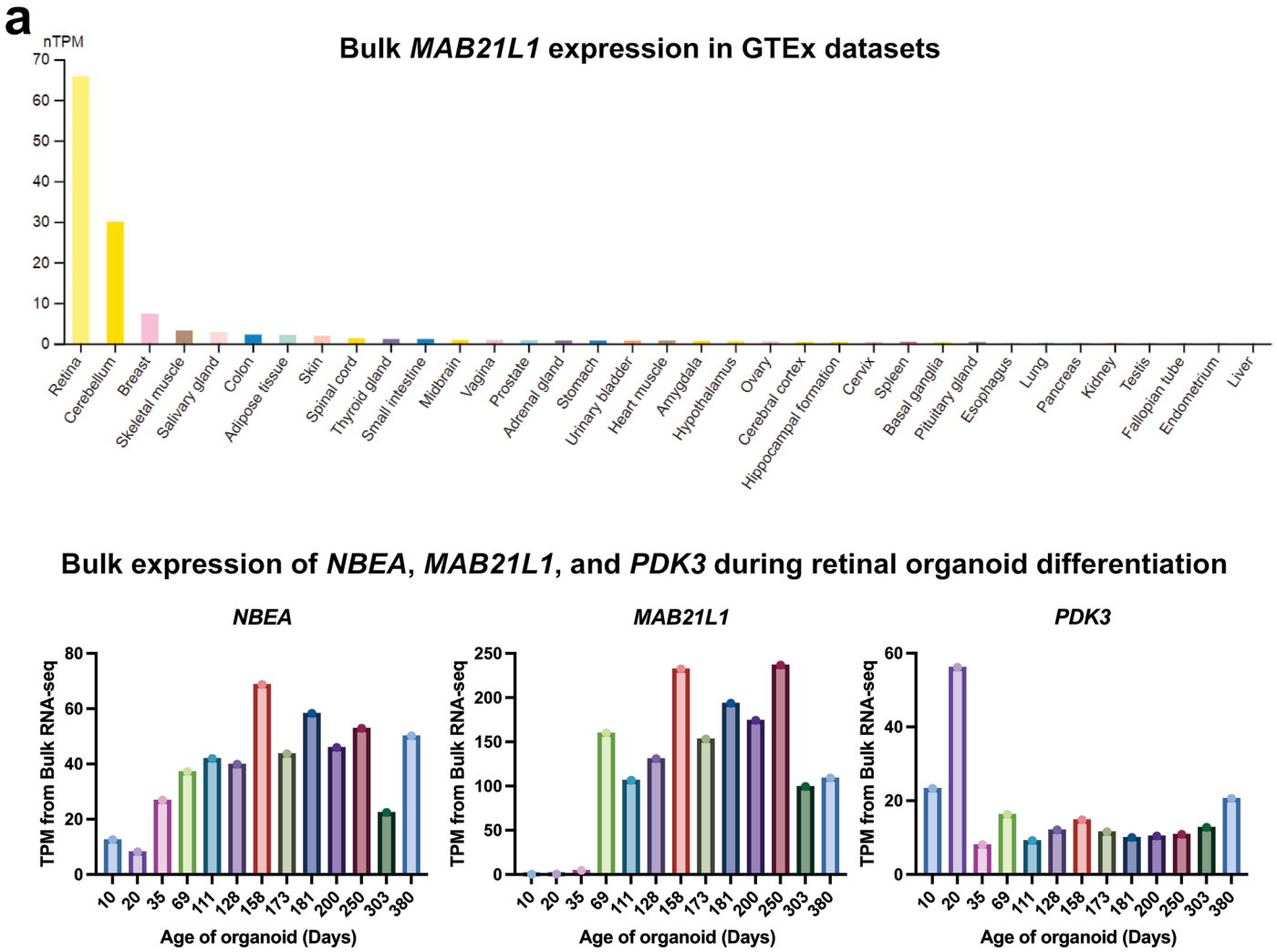
Extended Data Fig. 4 | Haplotype-specific chromatin accessibility of *NBEA* promoter and transcript NMD. a, Per-molecule chromatin accessibility of the *NBEA* promoter along both the der(13) and intact chr13 chromosomes within patient-derived fibroblasts. Chromatin accessibility was measured using Fiber-seq. **b**, Schematic representation for resolving whether the fusion transcripts are being subjected to nonsense-mediated decay (NMD). Specifically, long-read transcript sequencing was performed on samples before and after treatment with the NMD inhibitor cycloheximide. The libraries were sequenced to similar depths, and the total number of transcripts corresponding to the *NBEA*–chrX fusion transcripts was quantified in both samples. **c**, Same as **b**, but for the *PDK3–MAB21L1* fusion transcript.

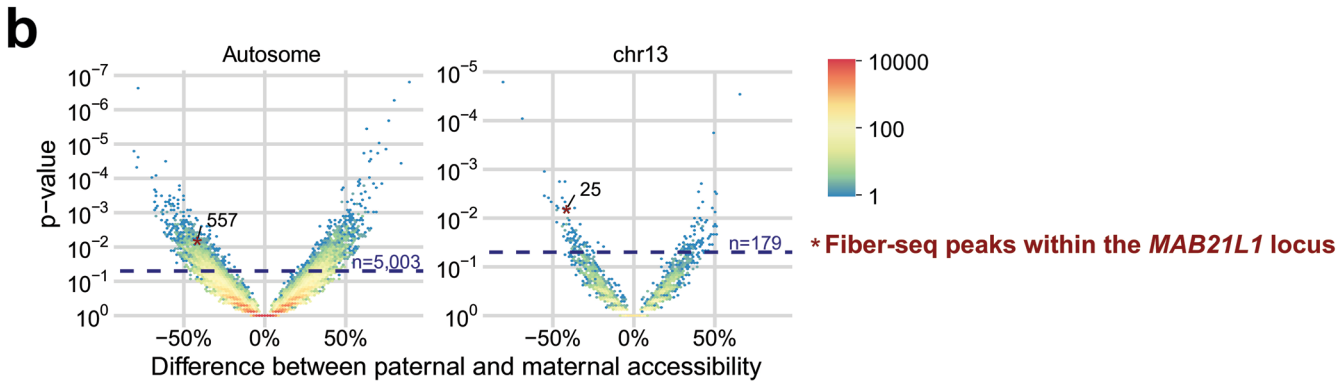
https://doi.org/10.1038/s41588-024-02067-0



Extended Data Fig. 5 | PDK3 protein levels in patient cells. a, Western blot of phospho-PDH and PDH in HEK293 cells, as well as HEK293 cells containing a viral expression vector for the intact PDK3 protein, or the PDK3–MAB21L1 fusion protein. **b**, Western blot of PDK3 in HEK293 cells, as well as HEK293 cells containing a viral expression vector for the intact PDK3 protein, or the PDK3–MAB21L1 fusion protein. * indicates protein bands selectively present in the PDK3-MAB21L1 sample. **c**, Western blot of PDK3 and actin protein levels in patient-derived fibroblasts, as well as control age-matched fibroblast cultures and retinal organoids. All experiments in the figure were performed once.

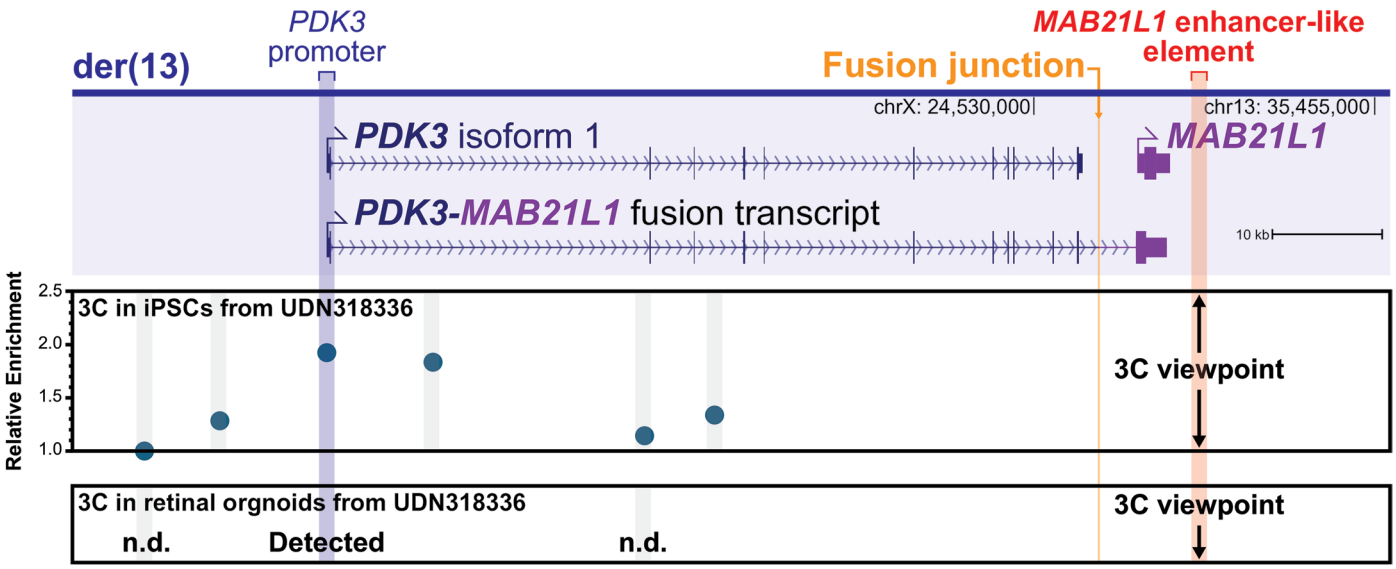
https://doi.org/10.1038/s41588-024-02067-0





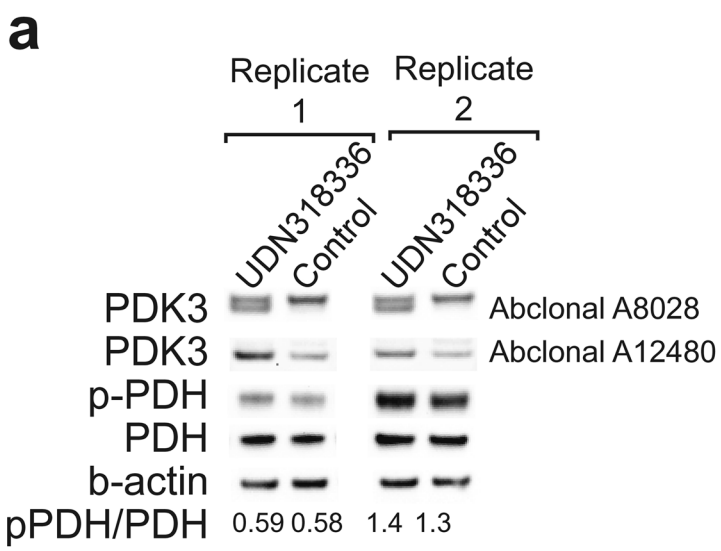
Extended Data Fig. 6 | Cell-selective expression of *MAB21L1*. **a**, (Top) Bar plot showing the bulk tissue RNA expression of *MAB21L1* (ENSG00000180660) from various GTEx samples are displayed by www.proteinatlas.org. (Bottom) Bulk expression of *NBEA*, *MAB21L1* and *PDK3* during retinal organoid differentiation as a function of the age of the organoid. TPM, transcripts per million. **b**, Volcano plot showing the absolute difference in the percentage of chromatin fibers with chromatin accessibility for each peak genome-wide between the paternal and maternal haplotypes for the retinal organoid Fiber-seq data from UDN318336. The MAB21L1 regulatory elements that are disrupted on der(13) are highlighted in red. Blue dash represents the nominal significance line ($p < 0.05$; two-sided Fisher's exact test without adjustment for multiple comparisons).

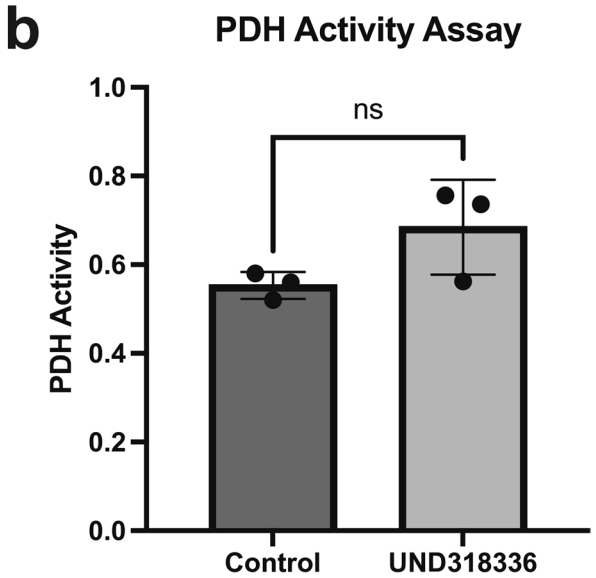
https://doi.org/10.1038/s41588-024-02067-0



Extended Data Fig. 7 | 3C chromatin interactions between PDK3 promoter and MAB21L1 enhancer. Results from chromatin conformation capture (3C) experiment using induced pluripotent stem cells (iPSCs) and retinal organoids from UDN318336. For the iPSC 3C experiment, displayed is the relative enrichment of different regions of the PDK3 gene for interactions with the MAB21L1 enhancer. For the retinal organoid 3C experiment, as a limited amount of 3C library was obtained, displayed are the qualitative results of qPCR for each of the tested interactions.

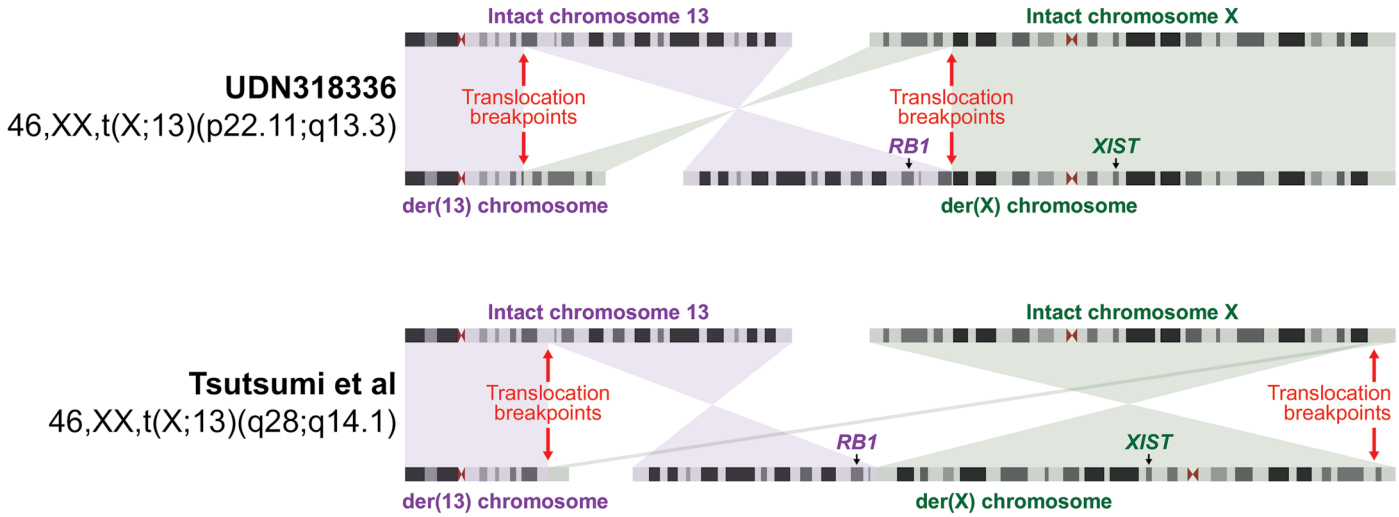
https://doi.org/10.1038/s41588-024-02067-0





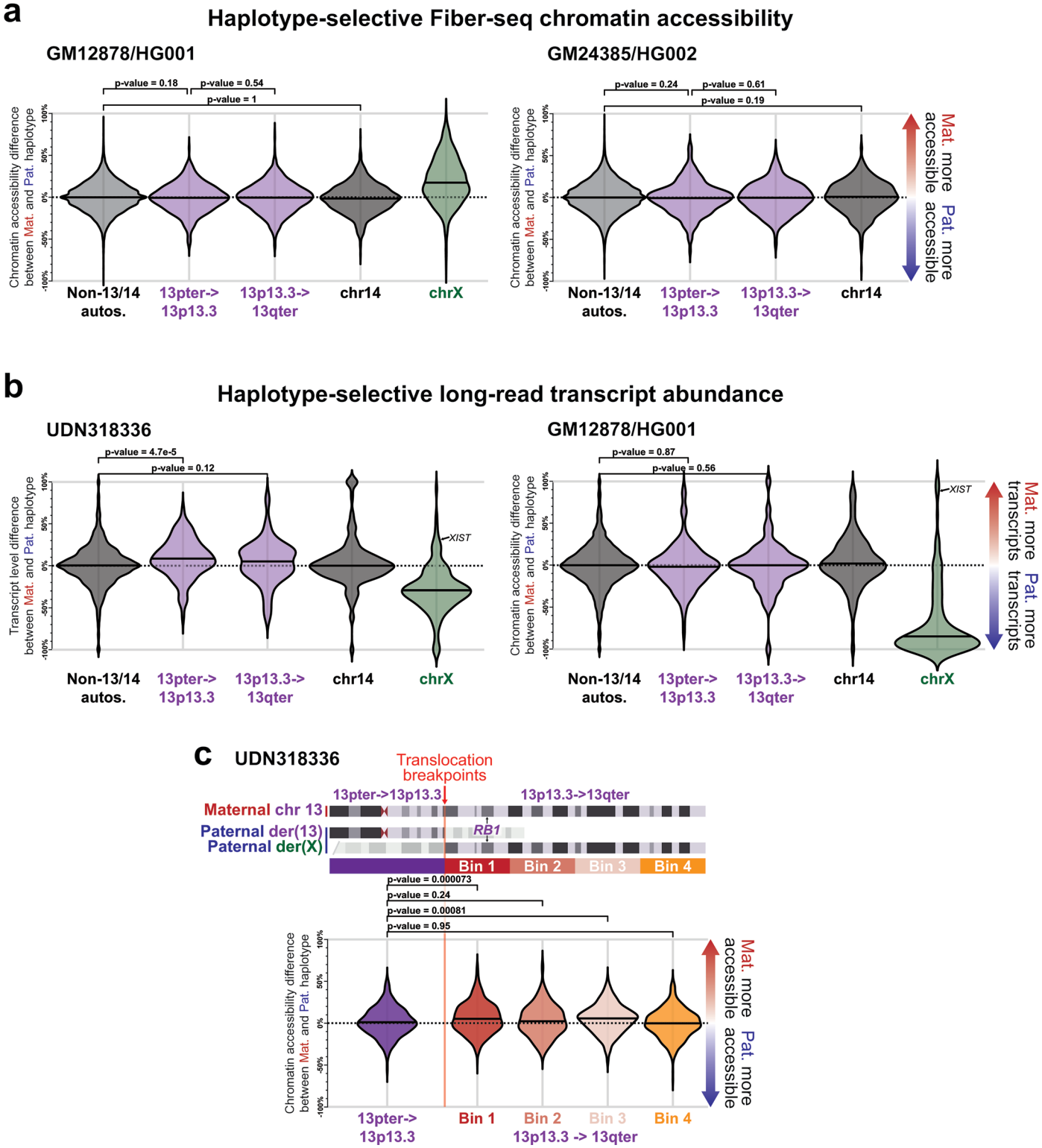
Extended Data Fig. 8 | PDK3 protein levels in patient retinal organoids and PHD activity in iPSCs. a, Western blot images showing quantification of PDK3, phospho-PDH, PDH and β-actin levels in patient-derived retinal organoids, as well as age-matched retinal organoids from a separate unaffected and unrelated individual. Data from two separate protein extractions are displayed. Quantification of phospho-PDH to PDH level is shown at the bottom. **b**, Pyruvate dehydrogenase activity assay on iPSCs from UND318336, as well as control iPSCs. Data are presented as mean values ± SEM, and results are from three replicates (n = 3). NS, not significant (p value = 0.11), unpaired two-sided t test.

https://doi.org/10.1038/s41588-024-02067-0



Extended Data Fig. 9 | Comparison with prior chromosome X;13 translocations. Ideogram showing the translocation breakpoints and derivative chromosomes for this case, as well as a previously published case who similarly had bilateral retinoblastomas. The translocation breakpoints for the previously published case are estimated based on the karyotype.

https://doi.org/10.1038/s41588-024-02067-0



Extended Data Fig. 10 | Allelic imbalance differences along chr13. a, Swarm plot showing the overall haplotype imbalance in chromatin accessibility along autosomes (except for chromosomes 13 and 14), chromosome X, chromosome 14 and two portions of chromosome 13 within GM12878 and GM24385 cells. P value calculated using a one-sided Mann–Whitney U test without adjustments for multiple comparisons. **b**, Swarm plot showing the overall haplotype imbalance in transcript production along autosomes (except for chromosomes 13 and 14), chromosome X, chromosome 14 and two portions of chromosome 13 within UDN318336 and GM12878 cells. P value calculated using a one-sided Mann–Whitney U test without adjustments for multiple comparisons. **c**, Swarm plot showing the haplotype imbalance in chromatin accessibility at different regions of chromosome 13 in fibroblasts from UDN318336. P value calculated using a one-sided Mann–Whitney U test without adjustments for multiple comparisons.

nature portfolio

Corresponding author(s): Andrew B. Stergachis

Last updated by author(s): Oct 22, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: No software was used for data collection

Data analysis	Custom software was created for the analysis described in the methods of the manuscript. Below we list the software and links to their permanent versions on Zenodo. - FIRE multi-ome-v0.1: https://doi.org/10.5281/zenodo.12701477, https://github.com/fiberseq/FIRE - k-mer-variant-phasing v0.0.1: https://doi.org/10.5281/zenodo.10655504, https://github.com/mrvollger/k-mer-variant-phasing Additionally the following tools were used: - DeepVariant v1.5.0 - pbsv v2.9.0 - Hiphase v1.2.1 - yak v0.1 - pb-CpG-tools v2.2 - SMRT Link v12.0 - pbfusion v0.1 - Python v3.9 - whatshap v2.3 - hifiasm v0.19

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All raw and processed sequencing data generated for GM12878, GM23485, GM20129, and HG02630 in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA1124997. Restrictions apply to the availability of some data generated or analyzed during this study to preserve subject confidentiality. Genomic data for our case report are accessible to the scientific community through the Undiagnosed Disease Network (dbGaP). Individuals interested in accessing UDN data through dbGaP should submit a data access request (DAR). Detailed instructions for this process can be found on the NIH Scientific Data Sharing website: https://sharing.nih.gov/accessing-data/accessing-genomic-data/how-to-request-and-access-datasets-from-dbgap. Cell lines obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research include GM12878, GM23485, GM20129, and HG02630. The UDN318336 fibroblast line was obtained directly from UDN participant UDN318336. Hia5 enzyme and plasmid is available upon request. ONT data for the CpG comparison from Epi2me (https://labs.epi2me.io/giab-2023.05/, super accuracy analysis level 60x bam files).

Processed FIRE results for GM12878, GM23485, GM20129, HG02630, UDN318336, and retinal organoid derived from UDN318336 are publicly available as track hubs hosted on AWS and can be directly linked into the UCSC genome browser:
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/GM12878_pacbiome/trackHub/hub.txt
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/HG002_pacbiome/trackHub/hub.txt
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/GM20129_PS00447/trackHub/hub.txt
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/HG02630_PS00445/trackHub/hub.txt
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/UDN318336/trackHub/hub.txt
https://s3-us-west-1.amazonaws.com/stergachis-manuscript-data/2023/Vollger_et_al_long-read_multi-ome/UDN318336_retinal/trackHub/hub.txt

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Chromosome count and sex assigned at birth for the five samples used in this study are provided.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity information are not relevant to this study.
Population characteristics	Additional population covariates are not relevant for this rare disease case report.
Recruitment	Participant was recruited through the Pacific Northwest Undiagnosed Diseases Network site using an IRB approved protocol.
Ethics oversight	This study was approved by the National Institutes of Health Institutional Review Board (IRB) (IRB # 15HG0130), and written informed consent was obtained from all participants in the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size as this study involved genetic testing of a single individual with a unique disease.
Data exclusions	No data were excluded from any of the experiments described.
Replication	To confirm the reproducibility of the multi-ome sequencing and the computational results we repeated the experiments on four well characterized cell lines (GM12878, GM23485, GM20129, and HG02630) and found consistent results as reported in Table 1.
Randomization	Randomization is not relevant to this study as this is a single case report.
Blinding	Blinding is not relevant to this study as this is a single case report.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	PDHA1: Abclonal A13687 p-PDHA1 S293: Abclonal AP1022 PDK3: Abclonal A8028 PDK3: Abclonal A12480 Beta Actin: Cell Signaling Technology 3700S Anti rabbit IgG, HRP linked Antibody, Cell Signaling Technology #7074S Anti-mouse IgG, HRP-linked Antibody. Cell Signaling Technology #7076S
Validation	Antibodies were validated to ensure they were detecting the correct sized protein product, per manufacture website.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cell lines GM12878, GM23485, GM20129, and HG02630 were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research. The UDN318336 fibroblast line was obtained directly from UDN participant UDN318336.
Authentication	GM12878 and GM23485 were authenticated using GIAB variant calls. GM20129 and HG02630 were authenticated using publicly available CpG methylation and parental short-read data.
Mycoplasma contamination	No Mycoplasma was detected in the sequencing data.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA