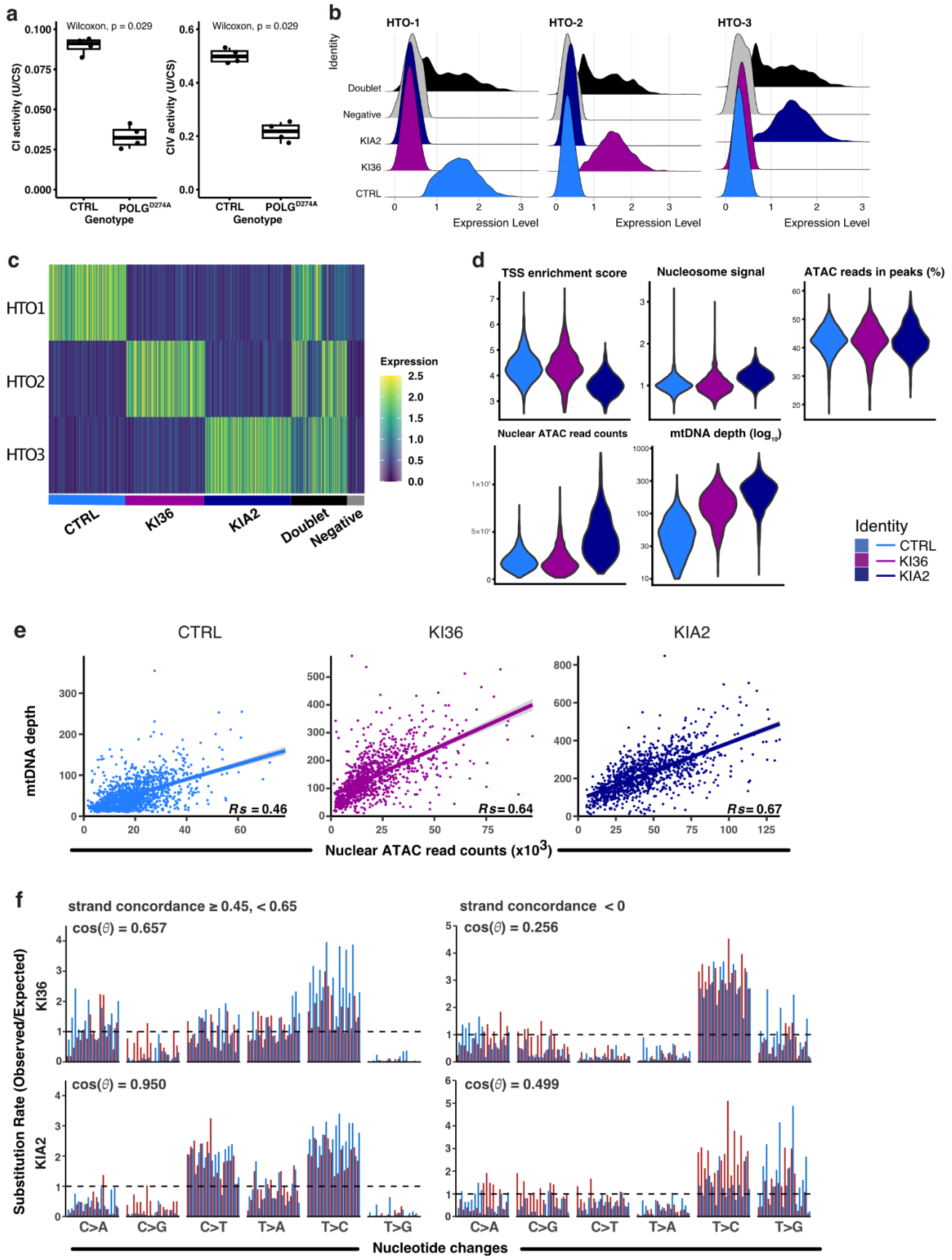
Supplementary Information

Single-cell multi-omic analysis of mitochondrial mutational mosaicism and dynamics

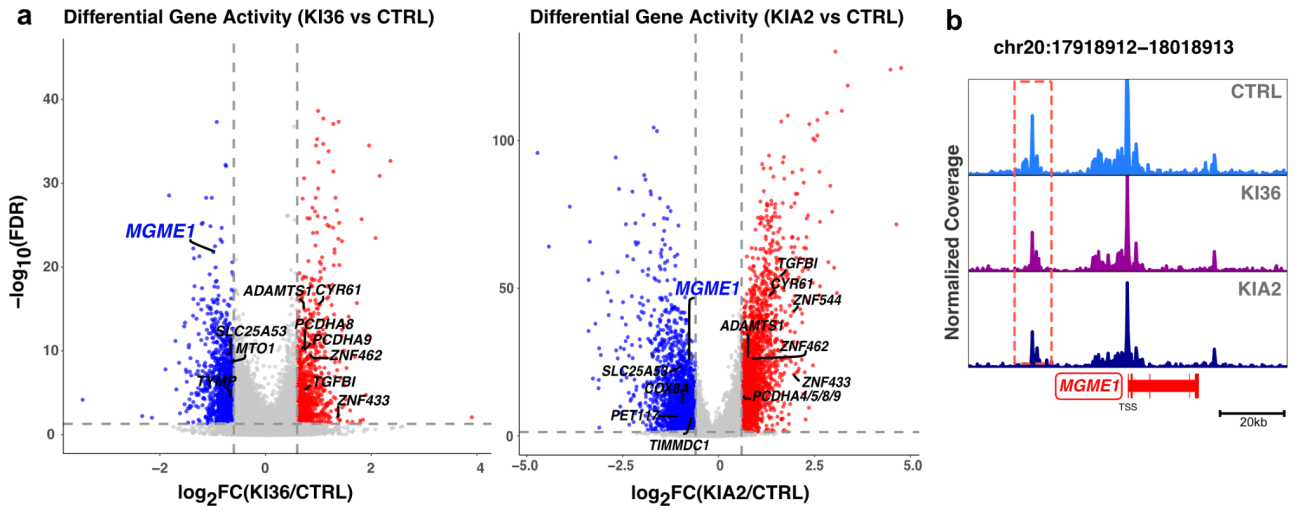
Yu-Hsin Hsieh, Pauline Kautz, Lena Nitsch, Ambre M. Giguelay, Janet Liebold, Veronika Dimitrova, Stephania Contreras Castillo, Freya Jungen, Gabor Zsurka, Genevieve Trombly, Markus Schuelke, Wolfram S. Kunz, Caleb A. Lareau, Leif S. Ludwig*

*To whom correspondence should be addressed: leif.ludwig@bih-charite.de

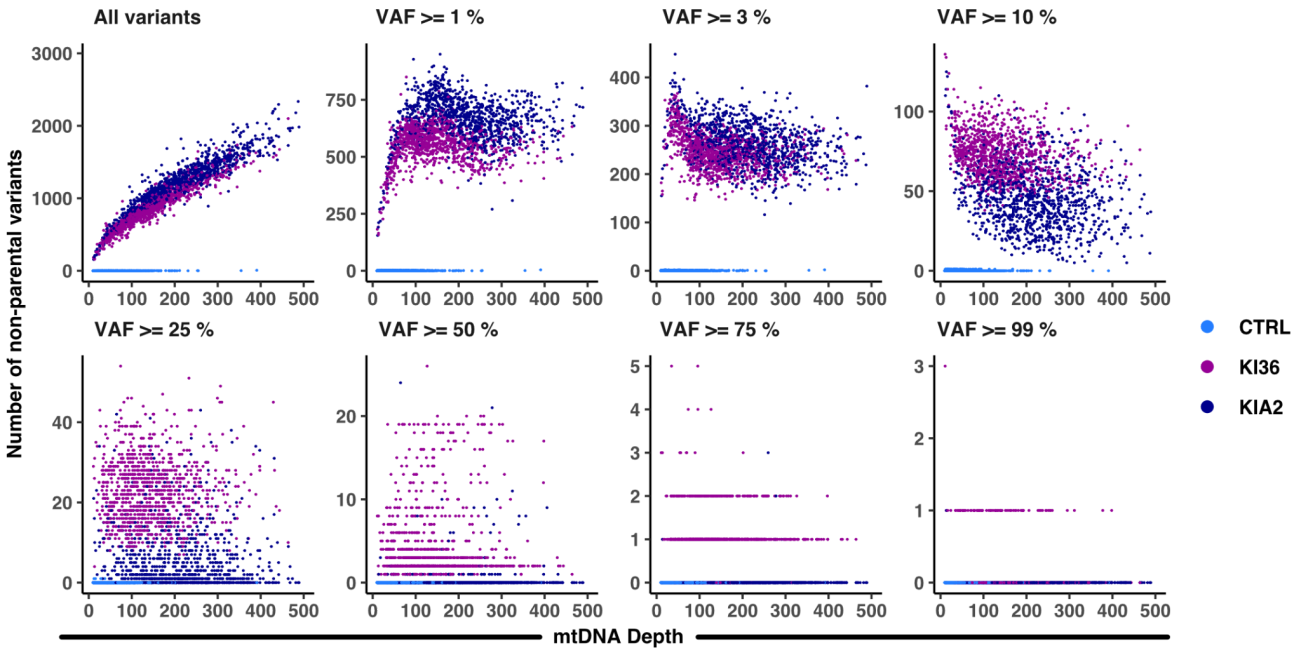


Supplementary Fig. 1 – mtscATAC-seq quality metrics of *POLG*^{D274A} HEK293 cell lines. (a) Enzyme activity of mitochondrial complex I (CI, left) and IV (CIV, left) in CTRL (n=4) and *POLG*^{D274A} (n=4) cells. Data were represented as units normalized to citrate synthase (CS) activity (U/CS).

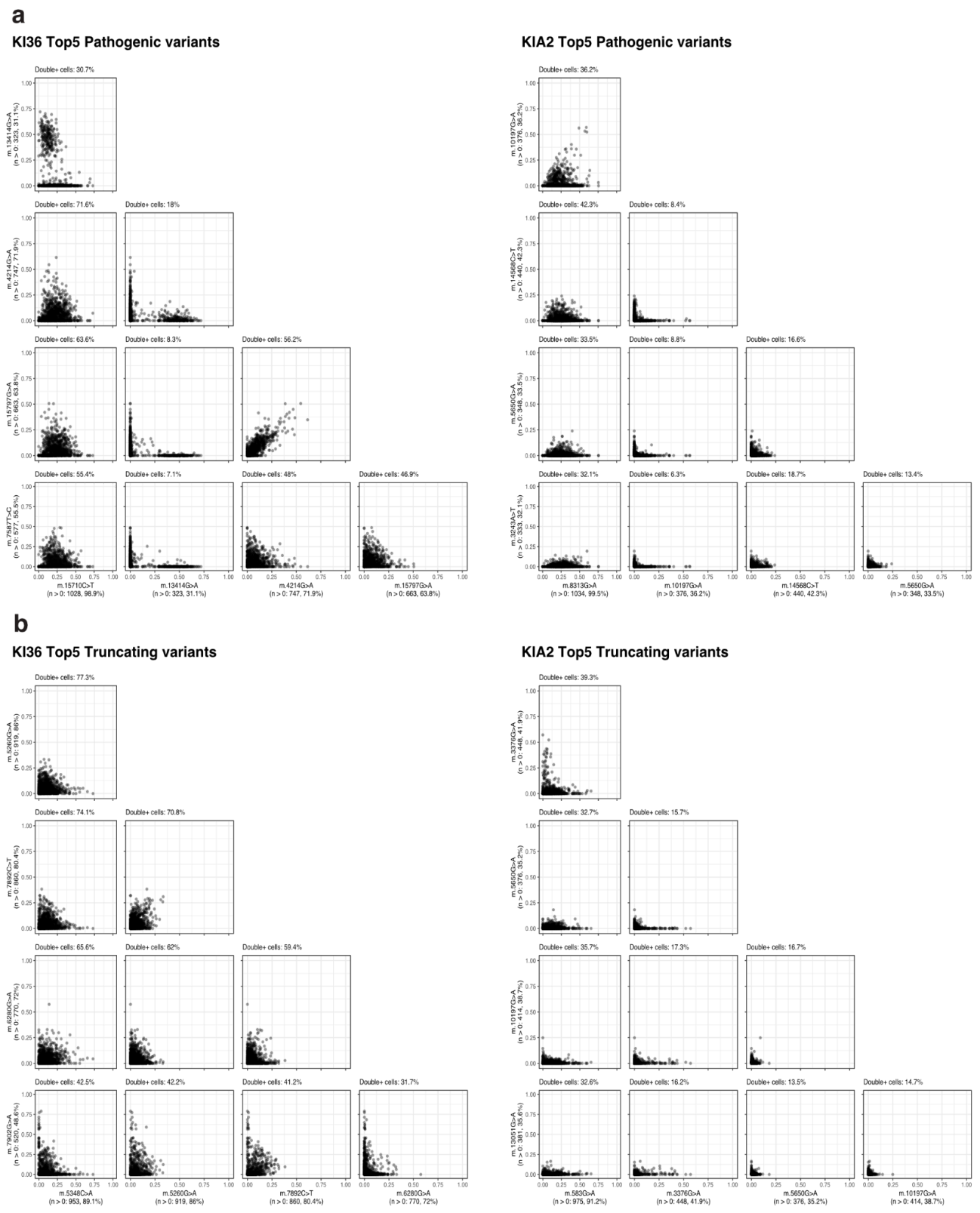
Statistical test: two-sided Wilcoxon rank-sum test. Boxes indicate the median (middle line), 25th to 75th percentiles (box). **(b)** Ridge plots illustrating the hashtag-derived oligo (HTO) expression of indicated cell lines after demultiplexing. **(c)** Heatmap of centered log-ratio (CLR)–normalized HTO expression. Rows represent the three HTOs, and each column corresponds to a single cell. **(d)** Violin plot of TSS enrichment score, nucleosome signal, percent reads in peaks, unique ATAC read counts, and mitochondrial DNA depth in single cells. **(e)** Correlations between ATAC read counts and mtDNA depth. Regression lines with the Spearman coefficient (R_s) are shown in the figure. **(f)** Mutational signature of mtDNA variants with an intermediate (≥0.45, <0.65, left) or negative (<0, right) strand concordance in both *POLG*^{D274A} (upper, KI36; lower, KIA2) cell lines. The cosine similarity (cos(θ)) to the mutational signature of high-confidence mtDNA variants (strand concordance > 0.65) is shown.



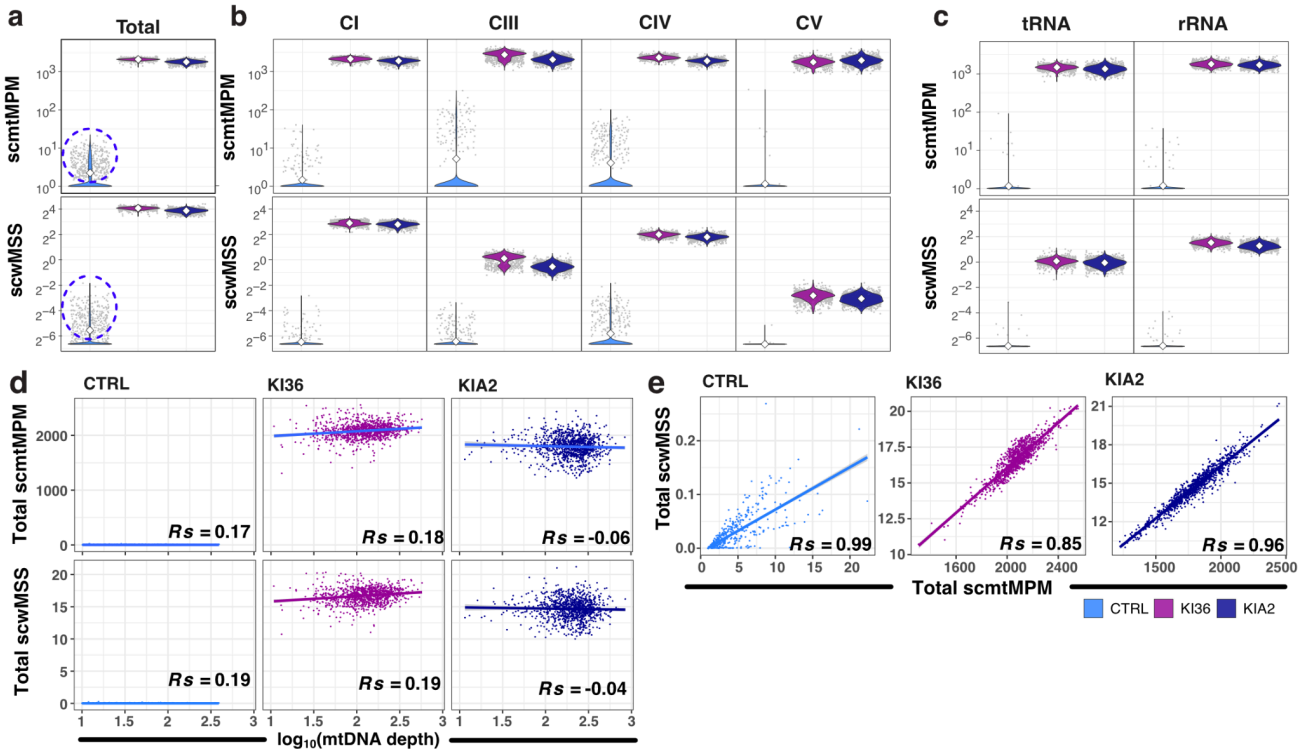
Supplementary Fig. 2 – Differential accessible gene analysis in *POLG*^{D274A} knock-in HEK293 cell lines. (a) Volcano plots of differentially accessible genes (DAGs) in KI36 (left) and KIA2 (right) compared to CTRL. Statistically significant up-regulated (FDR ≤ 0.05 with $\log_2$FC >=0.6) DAGs are highlighted in red, and down-regulated (FDR ≤ 0.05 with $\log_2$FC >=0.6) DAGs are highlighted in blue. **(b)** Pseudobulk chromatin accessibility track plot of *MGME1*. The red dashed box highlights a differentially accessible peak.



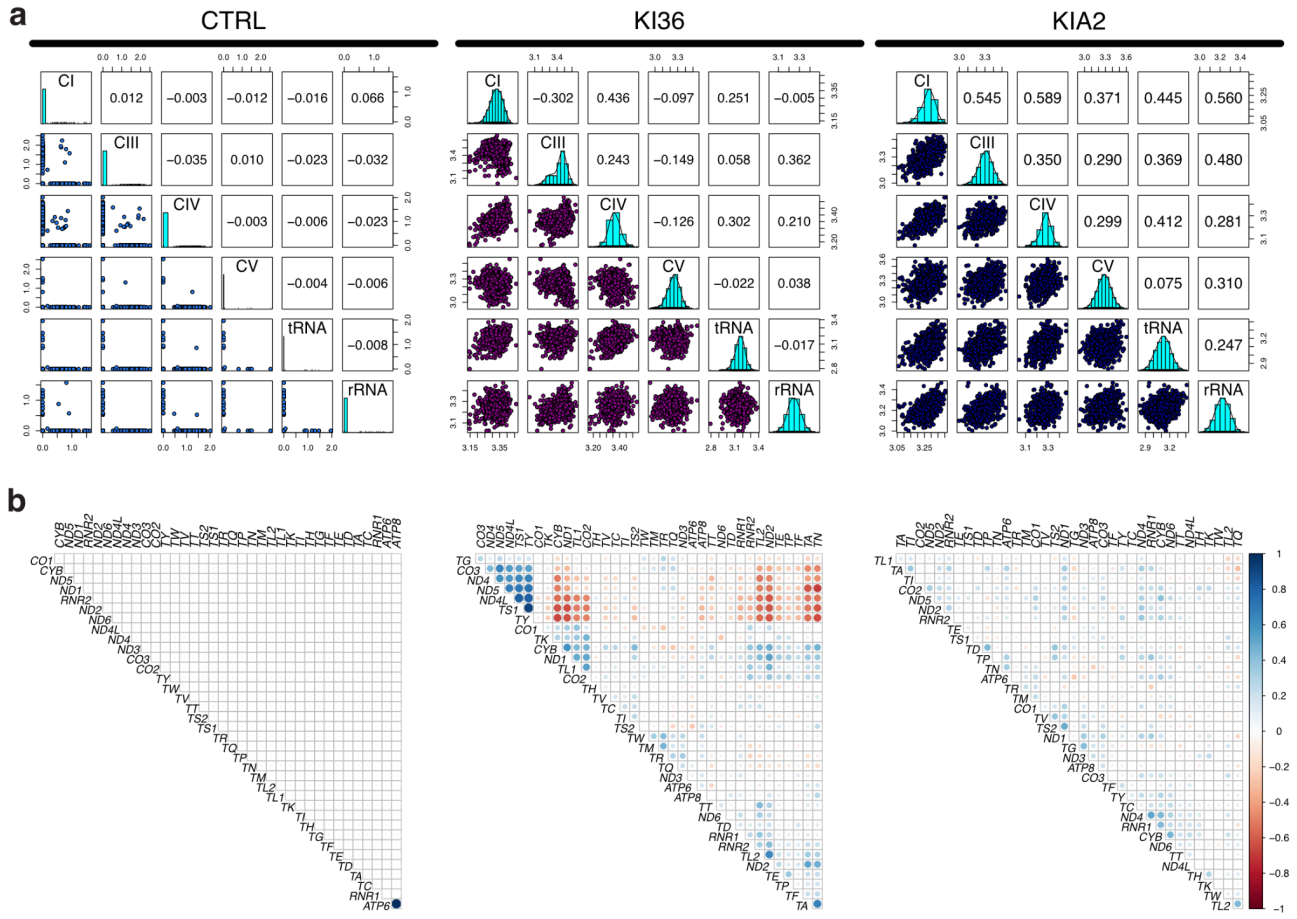
Supplementary Fig. 3 – Correlations between mitochondrial DNA depth and the number of detected high-confidence non-parental somatic mtDNA variants in *POLG*^{D274A} cell lines. The number of high-confidence *de novo* mtDNA variants at different variant allele frequency cut-offs versus mtDNA depth in single cells is plotted and colored by the identity of control or *POLG*^{D274A} cell lines.



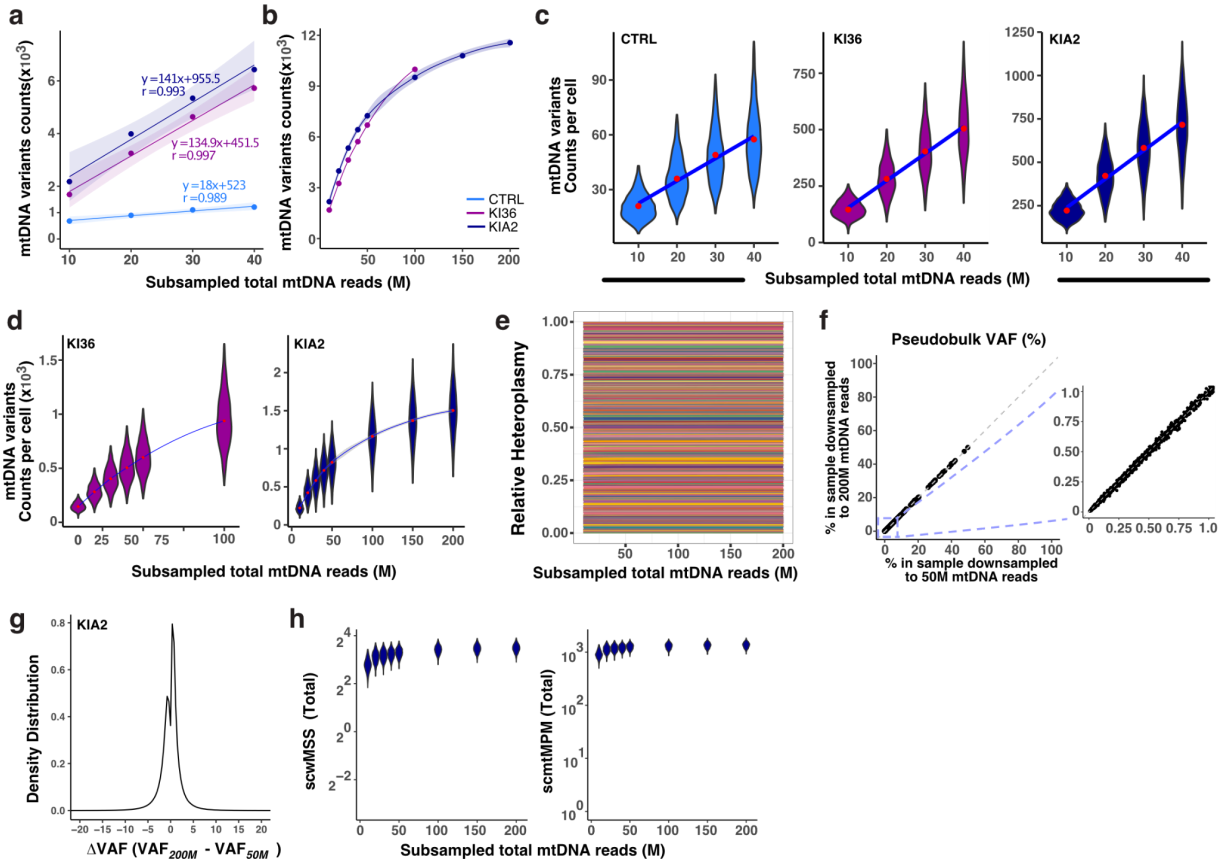
Supplementary Fig. 4 – Pairwise distribution of top-ranked truncating and pathogenic mtDNA variants. (a) Scatter plot matrices show pairwise comparisons of the top five mtDNA pathogenic variants. Points represent individual cells, with x- and y-axes showing variant allele frequencies (VAFs), along with the number and frequency of cells carrying each variant. Each plot is annotated with the percentage of double-positive cells (cells with non-zero VAF for both variants). **(b)** Same as (a), but for the top five mtDNA truncating variants from each dataset.



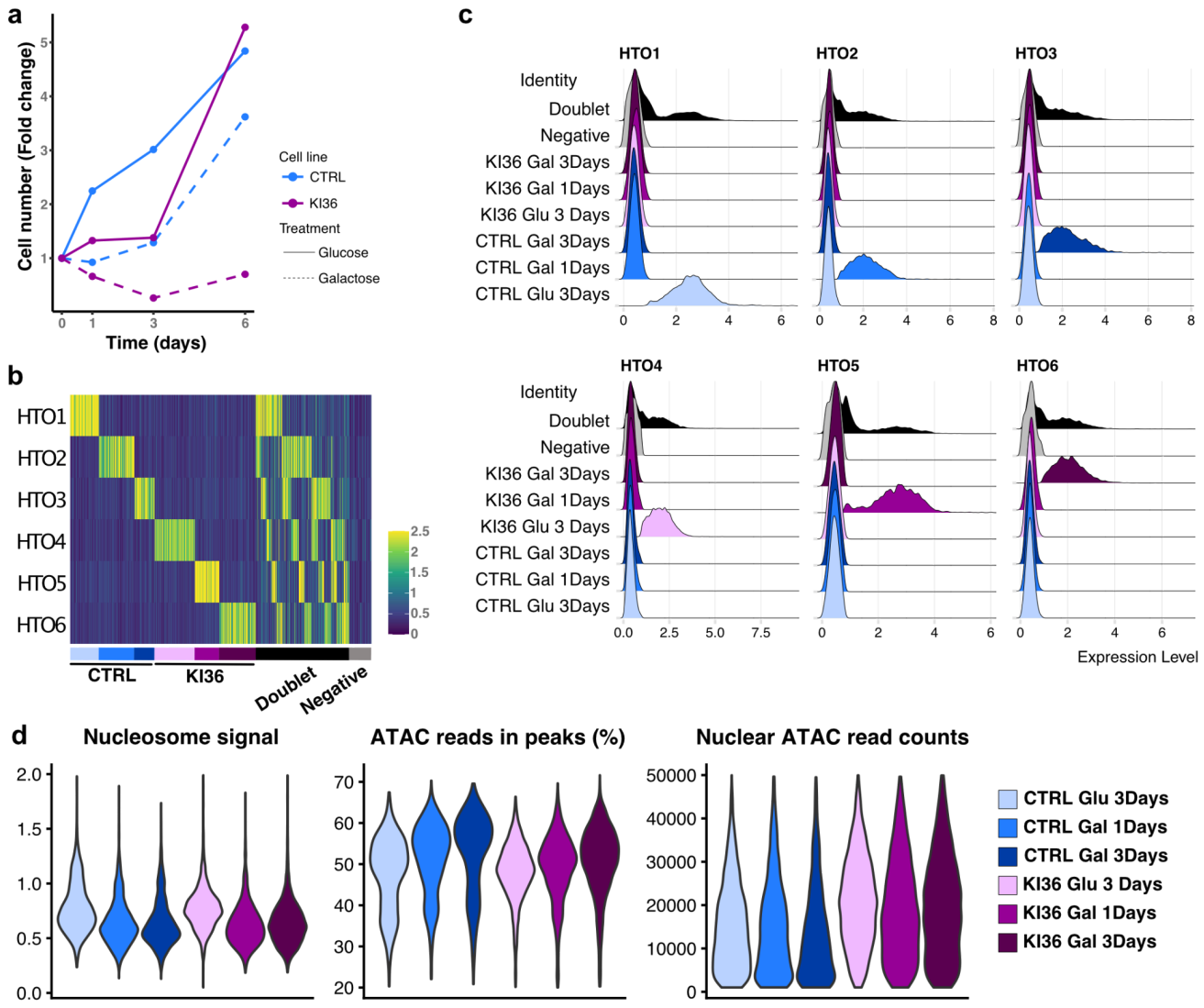
Supplementary Fig. 5 – Supporting information on mtDNA mutational burden quantifications in *POLG*^{D274A} HEK293 cell lines. (a) Violin plots of Mutations per million bases (scmtMPM) and Heteroplasmy-weighted sum of mitochondrial constraint scores (scwMSS) of total mtDNA in *POLG*^{D274A} cell lines. Dotted circles highlight elevated scores for complex I and IV, as well as rRNA in the control cell line. **(b)** Same as (a), but for the indicated respiratory chain complexes. **(c)** Same as (a), but for tRNA and rRNA. Dotted circles highlight elevated scores for complex I and IV, as well as rRNA in the control cell line. **(d)** Correlations between mitochondrial DNA depth and quantifications of mtDNA mutational burden. **(e)** Correlations between scmtMPM and scwMSS. The gray area around the regression lines represents confidence intervals, with the Spearman coefficient (R_s) being indicated in the figure.



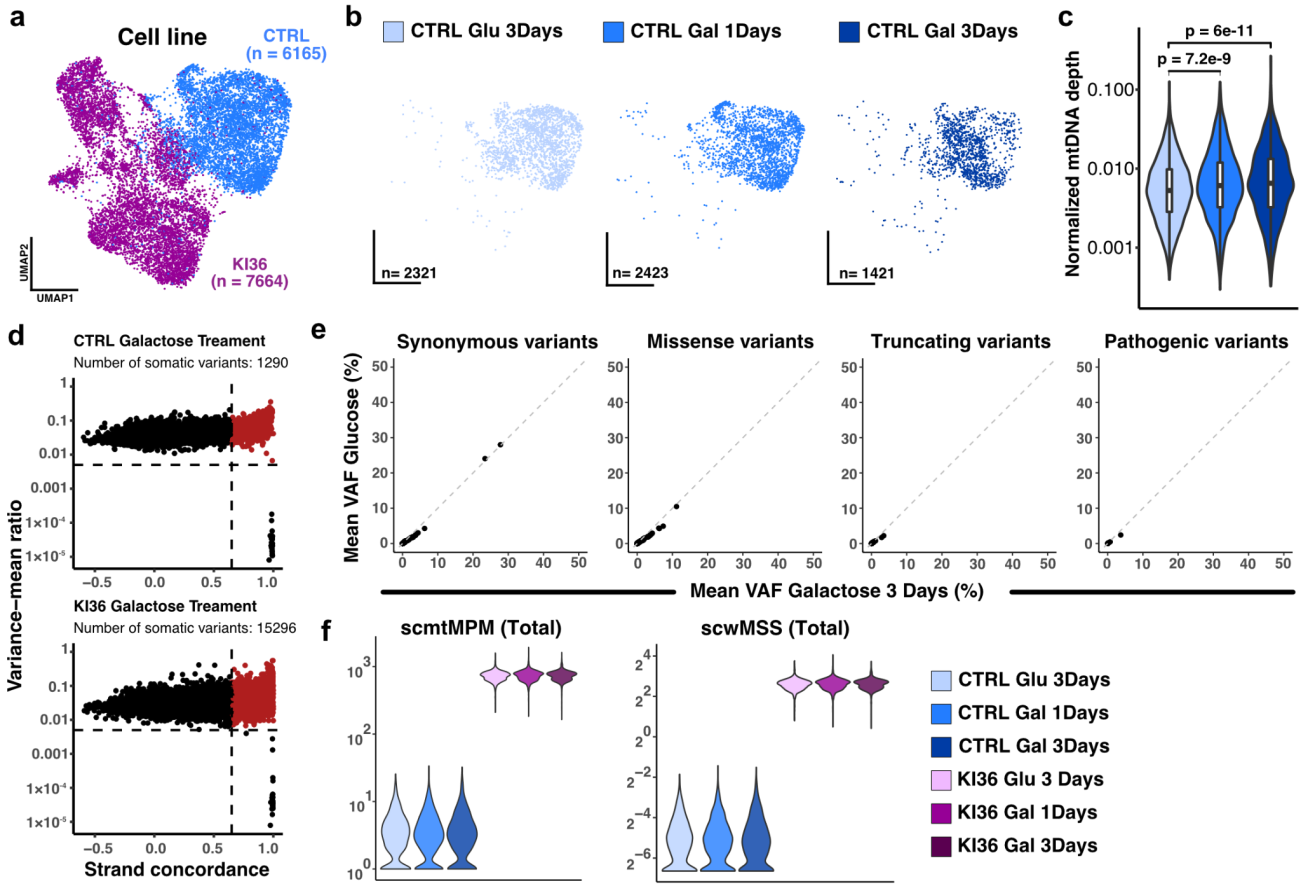
Supplementary Fig. 6 – Correlative distributions of somatic mtDNA mutational burden across the mitochondrial genome in *POLG*^{D274A} knock-in cell lines. (a) Pair-plot of scmtMPM for individual cells from the control line (CTRL, top), and the two hyper-mutated lines (K36, middle; and KIA2, bottom), stratified by functional categories. Each dot represents one cell. Kernel-density estimates are shown on the diagonal for each gene category. Scatter plots are shown in the lower left triangle, and the Pearson coefficient (r) is shown in the upper right triangle. **(b)** Heatmaps of Pearson correlation coefficients between scmtMPM values for each of the 37 mitochondrial genes in the same cells.



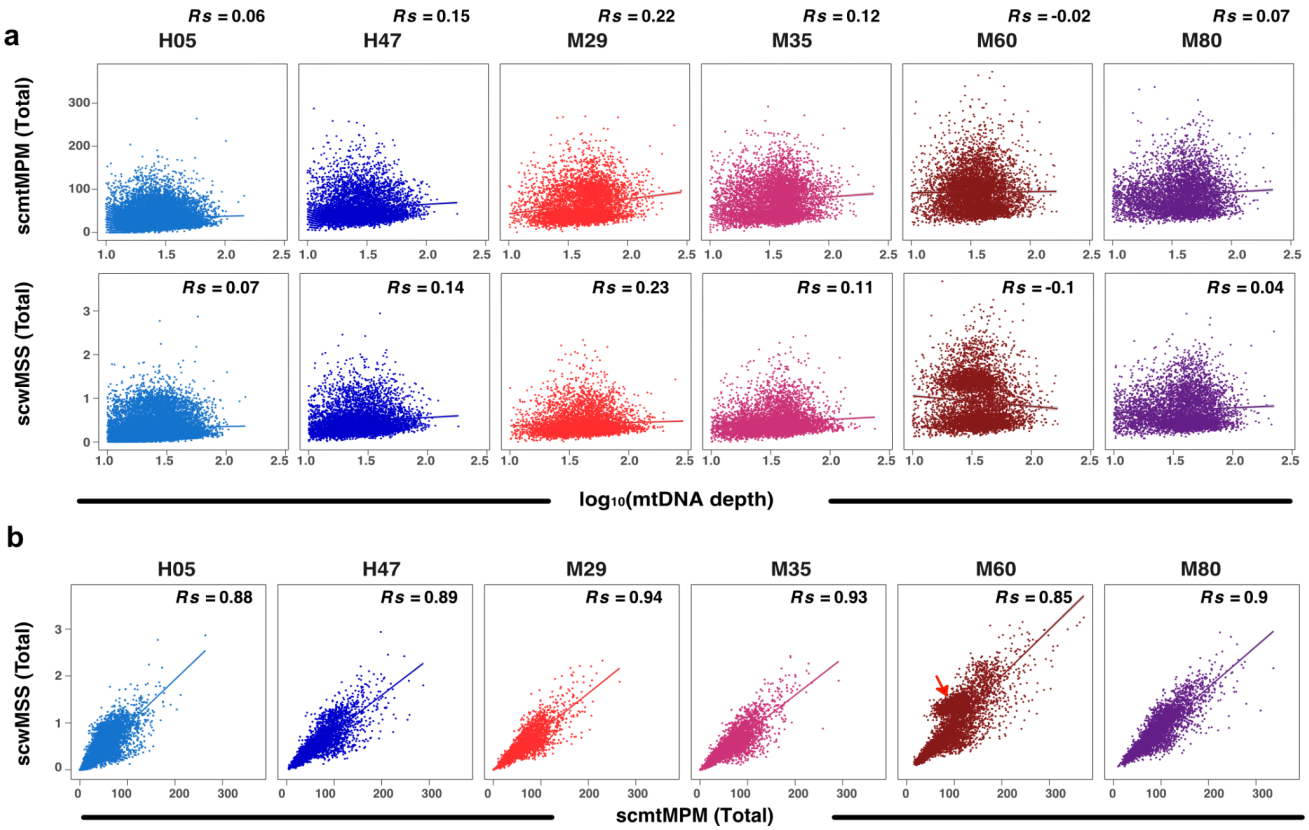
Supplementary Fig. 7 – Evaluation of sequencing depth on mtDNA variant detection in *POLG*^{D274A} cell lines. (a) Linear regression of total mtDNA variant detection rate as a function of sequencing depth up to <40M reads. Data points represent the number of total mtDNA variants detected. Regression lines with 95% confidence intervals and equations are shown. **(b)** Locally estimated scatterplot smoothing (LOESS) regression of total mtDNA variant counts across the broader range of sequencing depths (10–200M reads). Data points represent the number of total mtDNA variants detected, and LOESS curves with 95% confidence intervals are shown. **(c)** Violin plots of mtDNA variant counts in single cells at shallow sequencing depths (up to 40M reads). The linear regression was applied to the per-depth median number of detected variants. **(d)** Violin plots of mtDNA variant counts in single cells across the broader range (up to 200M). The LOESS models was applied to the per-depth median number of detected variants. **(e)** Stream plot of relative heteroplasmy across the downsampled sequencing depths, depicting the overall variant composition. Each band represents a single variant's relative abundance after normalizing to the summed pseudobulk VAFs of all variants across downsampled sequencing depths. **(f)** Scatter plots show pseudobulk VAFs of mtDNA variants or all detected variants in KIA2 when comparing datasets downsampled to 50M versus 200M mtDNA reads. Each point represents a single mtDNA variant. Insets zoom in on variants with VAF <1% to highlight concordance even at low allele frequencies. **(g)** Density plot of changes in estimated VAF (ΔVAF; 200M vs. 50M) for all non-zero-VAF mtDNA variants (7,100 variants) across individual KIA2 cells (n = 1,894). **(h)** Violin plots of scmtMPM and scwMSS in single cells across the fixed subsampled mtDNA sequencing depth (10–200M mtDNA reads).



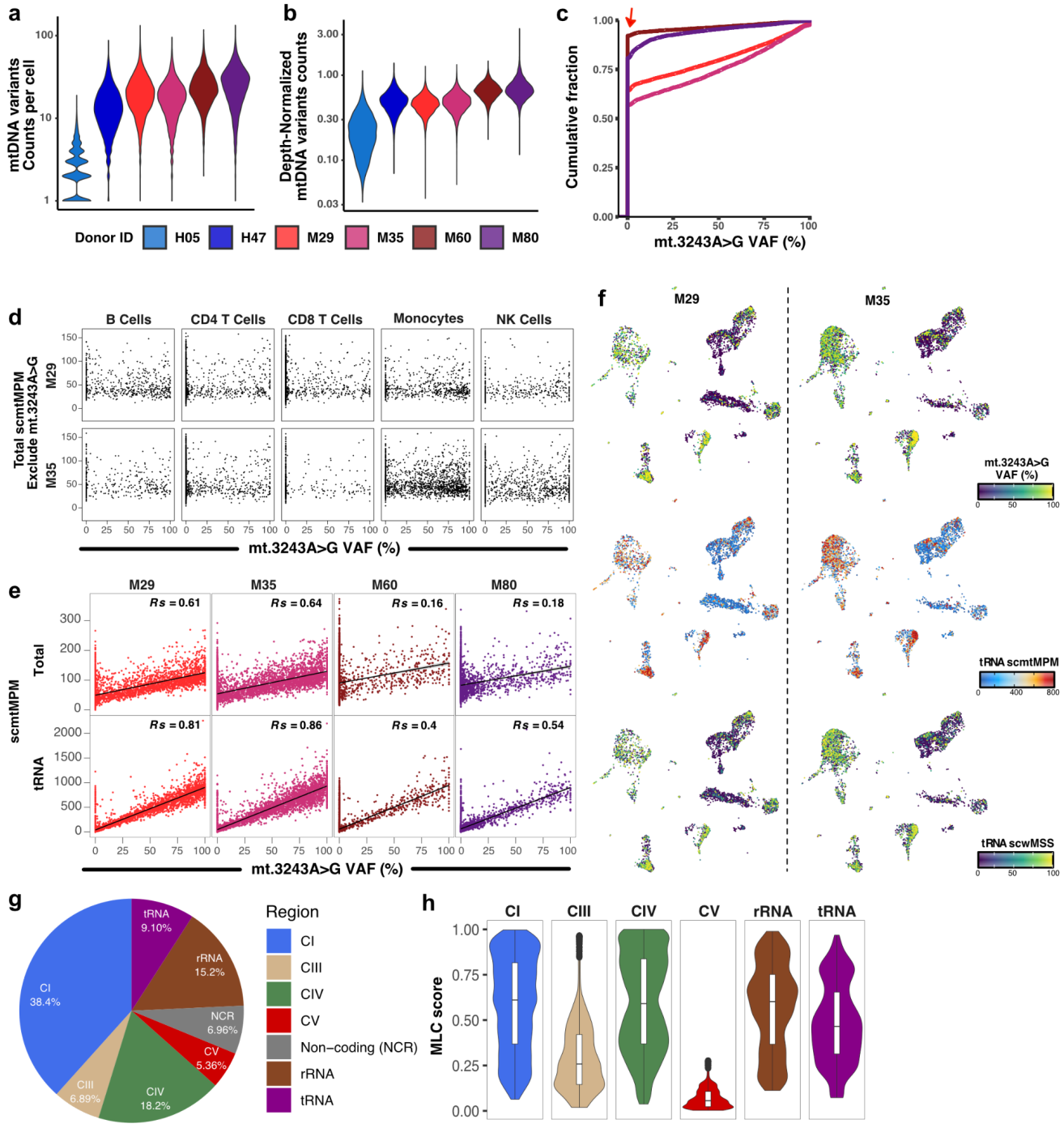
Supplementary Fig. 8 – Quality control metrics for mtscATAC-seq data of galactose-treated *POLG*^{D274A} HEK293 cell lines. (a) Growth curve of CTRL and KI36 lines in glucose- and galactose-containing medium. The fold-change relative to the number of seeded cells ($3x10^5$) is indicated. **(b)** Heatmap of centered log-ratio (CLR)–normalized HTO expression. **(c)** Ridge plots illustrating the HTO expression of the indicated cell line-treatment group after hashtag-derived oligo (HTO) demultiplexing. **(d)** Violin plot of nucleosome signal, percent reads in peaks, and unique ATAC read counts.



Supplementary Fig. 9 – Supporting information on the mtDNA mutational landscape of galactose-treated CTRL HEK293 cells. (a) UMAP projection of hashtag-demultiplexed CTRL and *POLG*^{D274A} *KI36* cell lines. n = cell numbers. **(b)** UMAP projection of CTRL cells colored by treatment condition. Glu, Glucose; Gal, Galactose. **(c)** Violin plots of single-cell mtDNA depth normalized to the number of unique ATAC read counts per condition in CTRL cells. Holm–Bonferroni adjusted p-values from the two-sided Wilcoxon test are shown, with significance at a type I error rate of 0.05. Boxes show the median (center line) and the interquartile range (25th–75th percentiles; box). Whiskers extend to the 5th and 95th percentiles. Outliers are displayed as single points. **(d)** High-confidence mtDNA variants were identified by filtering variants with a high variance-mean ratio (VMR) and high strand concordance in paired-end sequencing data. High-confidence heteroplasmic mtDNA variants for further analysis are colored red. **(e)** Scatter plots showing per-variant pseudobulk VAFs in CTRL cells treated with galactose for 3 days (y-axis) versus glucose for 3 days (x-axis), stratified by (functional) category. **(f)** Violin plots of scmtMPM (left) and scwMSS (right) of mtDNA for the indicated cell-treatment groups.

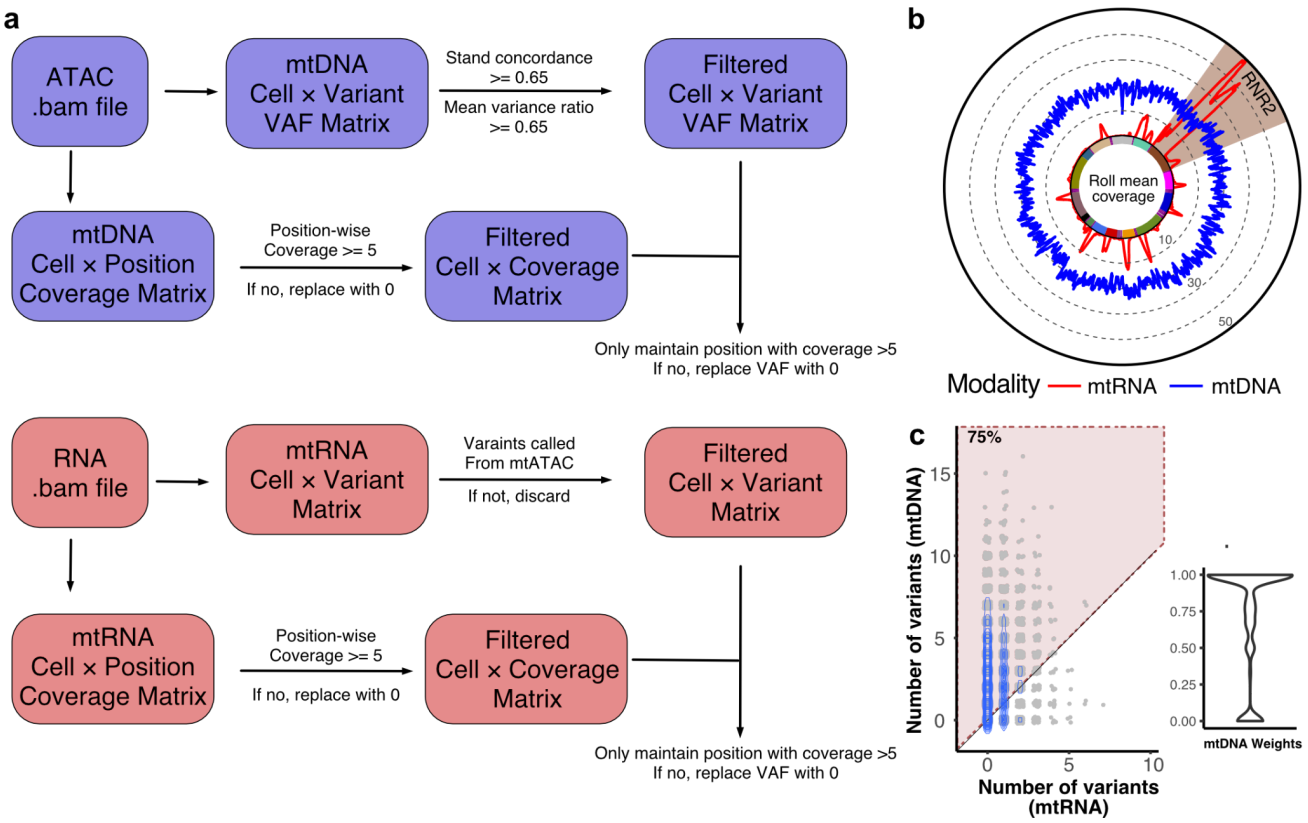


Supplementary Fig. 10 – Supporting information on mtDNA mutational burden quantification in primary human blood cells. (a) Correlations between mitochondrial DNA depth and mtDNA mutational burden quantification in individual donors. **(b)** Correlations between scmtMPM and scwMSS. Regression lines with the Spearman coefficient (R_s) are shown in the figure.



Supplementary Fig. 11 – Supporting information on scMMB quantification stratified by functional and gene groups in primary human blood cells. (a) Violin plots of the raw number of mtSNVs detected. **(b)** Violin plots showing the number of mtSNVs detected normalization to mtDNA depth. **(c)** Cumulative distribution plots of mt.3243A>G heteroplasmy stratified by MELAS patients. The red arrow highlights that more than 50% of the cells have no detectable pathogenic mutation in all four donors examined. **(d)** Cell-type resolved correlations between mt.3243A>G heteroplasmy and the total scmtMPM, with the pathogenic mt.3243A>G variant being deliberately excluded from the scmtMPM score. **(e)** Correlations between mt.3243A>G heteroplasmy and the total scmtMPM (top) and tRNA-specific scmtMPM (bottom). Regression lines with the Spearman coefficient (R_s) are shown in the figure. **(f)** Azimuth reference UMAP colored by mt.3243A>G heteroplasmy (top),

scmtMPM (middle), and scwMSS (bottom) for all tRNA genes for patient M29 (left) and M35 (right). For cell type annotation, compare to Fig. 4a. **(g)** Composition of human mitochondrial genome stratified by the length of OXPHOS complex and functional gene groups **(h)** Distribution of MLC score stratified by OXPHOS complex and functional gene groups (Median MLC [Q1, Q3], Complex I: 0.611 [0.368, 0.816]; Complex III: 0.258 [0.145, 0.420]; Complex IV: 0.591 [0.369, 0.837]; Complex V: 0.056 [0.024, 0.107]; rRNA: 0.602 [0.367, 0.751]; tRNA: 0.465 [0.314, 0.654]). Holm–Bonferroni adjusted p-values from the two-sided Wilcoxon test are shown, with significance at a type I error rate of 0.05. Boxes show the median (center line) and the interquartile range (25th–75th percentiles; box). Whiskers extend to the 5th and 95th percentiles. Outliers are displayed as single points.



Supplementary Fig. 12 – Comparison of variant calling from mtDNA- and mtRNA-derived sequence reads. (a) Bioinformatic workflow for cross-calling mtDNA and mtRNA variants from a 10x single-cell multiome PBMC dataset. **(b)** Rolling mean coverage of the mitochondrial genome from ATAC-seq and RNA-seq modalities. The brown-shaded area highlights the *MT-RNR2* gene locus, which exhibits the highest RNA coverage. The colors of the inner circle correspond to gene annotations of the mitochondrial genome (compare to Fig. 1e); the blue line: mtDNA coverage; and the red line: mtRNA coverage. **(c)** Jitter plot showing the number of confidently detected variants in mtDNA (y-axis) and their recovery from mtRNA (x-axis)**.** The violin plot shows the mtDNA-to-mtRNA weight ratio for variant detection, and 75% of cells have a higher mtDNA weight (>0.5).

Supplementary Table 1 | *Genetic and functional overview of complex I mtDNA variants*

Position	***Variant***	***Gene***	***Mutation type***	***Amino acid change***	***MLC***	***SIFT[1-2]/ AlphaMissense[3]***	***MITOMAP***
mt.10599	*G>A*	*MT-ND4*	*missense*	*A44T*	*0.747*	*likely benign*	*-*
mt.10270	*T>C*	*MT-ND3*	*missense*	*L71P*	*0.969*	*likely pathogenic*	*-*
mt.10398	*A>G*	*MT-ND3*	*missense*	*T114A*	*0.578*	*likely benign*	*pleiotropic[4-8]*

Footnotes

[1] Sorting Intolerant from Tolerant (SIFT)

[2] Sim, N.-L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40, W452–W457 (2012).

[3] Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381, eadg7492 (2023) doi:10.1126/science.adg7492.

[4] Smullen, M. et al. Modeling of mitochondrial genetic polymorphisms reveals induction of heteroplasmy by pleiotropic disease locus 10398A>G. Scientific Reports 13, 1–15 (2023).

[5] Tzeng, I. S. Role of mitochondrial DNA A10398G polymorphism on development of Parkinson's disease: A PRISMA-compliant meta-analysis. Journal of Clinical Laboratory Analysis 36, e24274 (2022).

[6] Juo, S. H. et al. A common mitochondrial polymorphism 10398A>G is associated metabolic syndrome in a Chinese population. Mitochondrion 10, 294–299 (2010).

[7] Pezzotti, A. et al. The mitochondrial A10398G polymorphism, interaction with alcohol consumption, and breast cancer risk. PloS one 4, e5356 (2009).

[8] van der Walt, J. M. et al. Mitochondrial polymorphisms significantly reduce the risk of Parkinson's disease. American Journal of Human Genetics 72, 804-811 (2003).

Supplementary Table 2 | BioLegend® TotalSeq™ hashtag antibodies and barcode assignments

Product Name	Cat No.	Barcode Sequence	Assignment
TotalSeq™-A0251 anti-human Hashtag 1 Antibody	394601	GTCAACTCTTTAGCG	CTRL
TotalSeq™-A0252 anti-human Hashtag 2 Antibody	394603	TGATGGCCTATTGGG	KI36
TotalSeq™-A0253 anti-human Hashtag 3 Antibody	394605	TTCCGCCTCTCTTTG	KIA2
TotalSeq™-B0251 anti-human Hashtag 1 Antibody	394631	GTCAACTCTTTAGCG	CTRL-Glc
TotalSeq™-B0252 anti-human Hashtag 2 Antibody	394633	TGATGGCCTATTGGG	CTRL-Gal (1 Day)
TotalSeq™-B0253 anti-human Hashtag 3 Antibody	394635	TTCCGCCTCTCTTTG	CTRL-Gal (3 Day)
TotalSeq™-B0254 anti-human Hashtag 4 Antibody	394637	AGTAAGTTCAGCGTA	KI36-Glc
TotalSeq™-B0255 anti-human Hashtag 5 Antibody	394639	AAGTATCGTTTCGCA	KI36-Gal (1 Day)
TotalSeq™-B0256 anti-human Hashtag 6 Antibody	394641	GGTTGCCAGATGTCA	KI36-Gal (3 Day)

Glc, Glucose; Gal, Galactose