

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

Received: 10 May 2025

Accepted: 26 February 2026

Published online: 16 March 2026

 Check for updates

Yu-Hsin Hsieh ^{1,2,3}, Pauline Kautz ^{1,2,3,4}, Lena Nitsch ^{1,2,3,5},
Ambre M. Giguelay ^{1,2,3}, Janet Liebold^{1,2,3}, Veronika Dimitrova ^{1,2,3,6},
Stephania Contreras Castillo ^{1,2,3}, Freya Jungen ^{1,2,3}, Gabor Zsurka ^{7,8},
Genevieve Trombly⁷, Markus Schuelke ^{1,9,10,11}, Wolfram S. Kunz ^{7,8},
Caleb A. Lareau ¹² & Leif S. Ludwig ^{1,2,3,13} 

Mitochondrial DNA (mtDNA) mutations occur more frequently than nuclear mutations and are associated with various diseases. While single-cell sequencing enables mtDNA variant heteroplasmy analysis, a holistic view of mtDNA mutational landscapes in individual cells has remained limited. Here, we leverage mitochondrial single-cell ATAC-seq and mtDNA-hypermutated *POLG*^{D274A} knock-in HEK293 cell lines to introduce two metrics—single-cell mtDNA mutations per million base pairs (scmtMPM) and heteroplasmy-weighted mitochondrial local constraint scores (scwMSS)—to capture cellular mutational loads and somatic mosaicism. We demonstrate that individual *POLG*^{D274A} cells exhibit complex mutational landscapes, with pathogenic mutations and truncating variants only present at subthreshold levels, indicative of their negative selection. In human healthy donors and mitochondrialriopathy patients, we identify constrained mutations in complex I, highlighting previously unrecognized mtDNA mutational landscape heterogeneity present on the single-cell level. Overall, scmtMPM and scwMSS provide a framework to investigate fundamental properties of mitochondrial genetics, disease, and somatic mosaicism.

Mitochondria are essential for cellular metabolism and contain a multicopy genome with hundreds to thousands of copies per cell^{1,2}. In humans, the 16.6 kb circular double-stranded mitochondrial DNA (mtDNA) encodes 13 protein subunits, contributing to four complexes

of the respiratory chain, alongside its own translational machinery, comprising two ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs). Notably, somatic mtDNA mutations occur more frequently than in their nuclear counterpart, with an up to >200× fold elevated rate in

¹Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany. ²Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. ³Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany. ⁴Technische Universität Berlin, Institute of Biotechnology, Berlin, Germany. ⁵Institute of Biology, Humboldt Universität zu Berlin, Berlin, Germany. ⁶Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany. ⁷Institute of Experimental Epileptology and Cognition Research, University Hospital Bonn, University of Bonn, Bonn, Germany. ⁸Department of Epileptology, University Hospital Bonn, University of Bonn, Bonn, Germany. ⁹Charité-Universitätsmedizin Berlin, NeuroCureCluster of Excellence, Berlin, Germany. ¹⁰Charité-Universitätsmedizin Berlin, Department of Neuropediatrics, Berlin, Germany. ¹¹Charité-Universitätsmedizin Berlin, NeuroCureClinical Research Center, Berlin, Germany. ¹²Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹³Charité-Universitätsmedizin Berlin, Cluster of Excellence ImmunoPrecept, Berlin, Germany. ✉e-mail: leif.ludwig@bih-charite.de

human blood³. Bulk-sequencing has demonstrated an age-related increase in somatic mtDNA variants^{4,5}, established multiple associations with cancers^{6–8}, and other human phenotypes¹. However, establishing genotype-phenotype relationships at the cellular level remains challenging due to heteroplasmy (the presence of wild-type and mutant mtDNA) with variable variant allele frequencies (VAFs), and cell-type- and state-dependent mtDNA copy number. Additionally, relaxed replication, vegetative segregation, and natural turnover of mitochondria collectively make cellular mtDNA VAFs more dynamic than for nuclear DNA variants².

The pathogenicity of mtDNA variants is best established in the context of maternally inherited mitochondrialopathies⁹. While typically multi-systemic, cell-type-specific effects may extend to red blood cell production, and selection against pathogenic mtDNA has been described in T cells^{10–13}. Defects in the murine mtDNA polymerase lead to a ‘mutator’ phenotype with increased levels of mtDNA variants that result in the premature onset of aging-related features^{14,15}. However, fundamental challenges remain in our understanding of the functional impact of mtDNA variants. While individual pathogenic single-nucleotide variants (SNVs) or large-scale deletions are readily definable, the additivity of constantly arising somatic variants is harder to account for. Depending on the specific gene affected and its VAFs, these variants may exert variable cell-type-specific effects. Traditionally, cells are thought to tolerate 60–80% of mutated mtDNA before reaching the biochemical threshold at which metabolic defects manifest themselves^{1,16}; however, recent genetic evidence suggests that phenotypes may already manifest at lower VAFs^{11,17}.

Recent advancements in single-cell sequencing have enabled massively parallel whole mitochondrial genome sequencing¹⁸, and have demonstrated an unprecedented level of variation in mtDNA variants¹⁹ and copy number²⁰ across thousands of cells. For example, mitochondrial single-cell ATAC-seq (mtscATAC-seq) enables the concomitant sequencing of mitochondrial genomes and chromatin accessibility^{19,21}. Moreover, a recently introduced mitochondrial genome constraint model provides a means of interpreting genetic variation underlying human phenotypes²² across genomic databases such as the Genome Aggregation Database (gnomAD). Here, we leverage these advances to characterize cell lines with proofreading-deficient mtDNA polymerases and to benchmark the sensitivity of these approaches for detecting and quantifying mitochondrial genetic variation. We demonstrate an even more substantial increase in mutational load than previously appreciated via bulk-sequencing and introduce a computational framework to quantify and interpret the genome-wide mtDNA mutational burden. We apply this framework to peripheral blood mononuclear cells (PBMCs) from human donors and mitochondrialopathy patients to aid the identification of pathogenic mtDNA variants. Our widely applicable framework opens avenues for single-cell-resolved study of mitochondrial genetics and somatic mosaicism in primary human cells and tissues.

Results

Elevated single-cell mtDNA burden in *POLG*^{D274A} cells

To assess the sensitivity of mtscATAC-seq to quantify increases in relative mitochondrial genome-wide mutational burden at the single-cell level, we harnessed exonuclease-deficient DNA polymerase γ (*POLG*^{D274A}) knock-in HEK293 cell lines with an elevated mutational rate and hypermutated mitochondrial genomes^{23–25}. Both KI36 and KIA2 lines showed metabolic impairment compared to CTRL (Supplementary Fig. 1a). The control and *POLG*^{D274A} cell lines were jointly profiled using mtscATAC-seq with cell hashing to distinguish individual lines and detect mtDNA variants alongside accessible chromatin profiles (Fig. 1a; Supplementary Fig. 1b, c; “Methods”). Following deep sequencing, quality control metrics were comparable among all three cell lines, which clustered distinctly, indicative of heterogeneous chromatin accessibility (Fig. 1b; Supplementary Fig. 1d). Of note,

single-cell mtDNA depth was 3–4-fold higher in both *POLG*^{D274A} knock-in lines (Fig. 1c; Supplementary Fig. 1d; median raw mtDNA depth CTRL: 45, KI36: 130, and KIA2: 214). Given that KIA2 cells also exhibited higher nuclear read counts (median single-cell ATAC-seq read counts CTRL: 18,398, KI36: 17,113, and KIA2: 41,700) that generally correlated with mtDNA depth across all three lines (Spearman correlation > 0.5), the mtDNA depth of individual cells was normalized to their respective nuclear ATAC read count (Fig. 1d; Supplementary Fig. 1d, e). However, the normalized single-cell mtDNA depth in both *POLG*^{D274A} knock-in lines remained significantly higher than in CTRL (Fig. 1e, f), suggesting higher absolute mtDNA content, likely reflecting a compensatory response in copy number to an elevated mutational load^{25,26}. Notably, in *POLG*^{D274A} cells, the pseudobulk rolling-mean mtDNA coverage appeared lower in the minor arc, the smaller segment of the genome between the replication origins of the heavy strand (oriH) and the light strand (oriL). This uneven distribution is consistent with the findings that *POLG* 3′–5′ exonuclease deficiency leads to the accumulation of linear mtDNA fragments in the major arc (Fig. 1g)^{23,24}.

Using the *mgatk*¹⁹ pipeline, 620 mtDNA variants were identified across CTRL cells. In the KI36 and KIA2 cells, a ~15-fold increase in mtDNA variants were identified, 9656 and 11407, respectively (Fig. 1h; Supplementary Data 1). While this higher number of mtDNA variants following disruption of *POLG* proofreading deficiency is expected, the overall magnitude is more pronounced than previously described, where only hundreds of somatic variants were identified by bulk-sequencing^{14,27,28}. Both control and *POLG*^{D274A} cells displayed a C > T transition-dominated mutational signature, indicating replication error as the primary driver (Fig. 1i)^{5,27,29}. Notably, relaxing the default thresholds in the *mgatk* variant calling resulted in a substantial increase in detected variants (Fig. 1h, blue dotted box; “Methods”), which down to a particular strand concordance still showed the expected mutational signature (Supplementary Fig. 1f). Presuming similar background levels compared to control, these findings suggest an even higher number of bona fide somatic mtDNA variants in *POLG*^{D274A} cells. Together, these observations demonstrate that mtscATAC-seq is a sensitive tool for detecting elevations in mtDNA mutational rates and the resulting increase in mtDNA variants.

Adaptive gene regulatory responses in *POLG*^{D274A} cells

To dissect nuclear–mitochondrial interactions in response to elevated mtDNA mutational burden, we performed differentially accessible gene (DAG) analysis of *POLG*^{D274A} vs. CTRL cells. For each *POLG*^{D274A} line, we identified 827 and 2319 upregulated DAGs, and 1198 and 3113 downregulated DAGs for KI36 and KIA2, respectively (Supplementary Fig. 2a; Supplementary Data 2, 3). Among these, 243 upregulated and 277 downregulated DAGs were shared between the KI36 and KIA2 lines. Upregulated genes included ones in extracellular matrix remodeling and mechanotransduction³⁰ (e.g., *CYR61*, *TGFBI*, *ADAMTS1*), protocadherins (*PCDH* gene family), and multiple zinc-finger (*ZNF*) transcription factors³¹, indicating altered cell–matrix interactions and epigenetic rewiring in *POLG*^{D274A} cells. Interestingly, mitochondrial genome maintenance exonuclease 1 (*MGME1*), a mitochondrial exonuclease essential for mtDNA maintenance and transcription associated with mitochondrial disease³², was among the downregulated genes in *POLG*^{D274A} cells (Supplementary Fig. 2b). *MGME1* deficiency^{23,33} has been shown to lead to accumulating linear mtDNA molecules, and *Mgme1*-knockout in mice elicited a strong inflammatory response via the cGAS–STING pathway³⁴. Additionally, several genes involved in cellular metabolism (e.g., *SLC25A31*, *SLC25A53*, *OTC*) and respiratory chain components/assembly³⁵ (e.g., *COX8A*, *TIMMDC1*, *SAMM50*) were downregulated in *POLG*^{D274A} lines, suggesting a global attenuation of mitochondrial biogenesis and OXPHOS function. Together, these findings demonstrate that mtscATAC-seq captures nuclear–mitochondrial interactions linked to mitochondrial dysfunction.

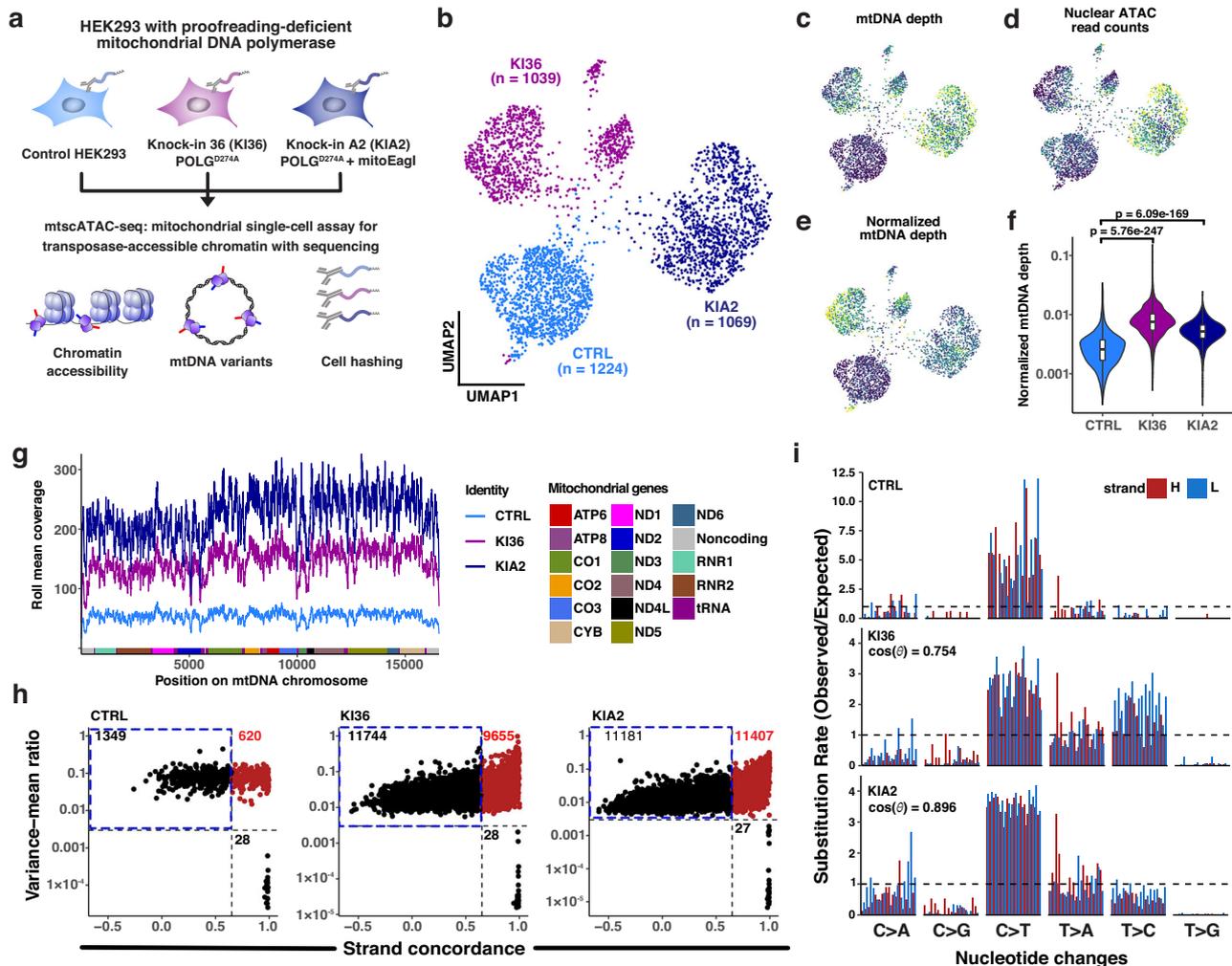


Fig. 1 | mtscATAC-seq-based profiling of single-cell mitochondrial DNA genotypes in *POLG*^{D274A} HEK293 cell lines. **a** Outline of the experimental design. CTRL and *POLG*^{D274A} HEK293 cell lines were characterized by mtscATAC-seq with cell hashing to jointly profile chromatin accessibility and mtDNA variants. **b** UMAP projection of hashtag-demultiplexed HEK293 cell lines. *n* = cell numbers. **c** UMAP projection of color-scaled raw mtDNA depth. **d** UMAP projection color-scaled by the number of unique ATAC read counts per cell. **e** UMAP projection of color-scaled ATAC-read-counts normalized by single-cell mtDNA depth. **f** Violin plot of the normalized single-cell mtDNA depth for cells passing QC (*n* = 1224, 1069, and 1039 for CTRL, KI36, and KIA2, respectively). 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 indicate the median (middle line), 25th to 75th percentiles (box), and

5th and 95th percentiles (whiskers). **g** Visualization of the mitochondrial genome position-wise pseudobulk coverage of the indicated cell lines, shown as a rolling-mean coverage at five-nucleotide resolution. **h** 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. The blue dotted box highlights the group of variants included when relaxing the default VMR threshold (CTRL: 1349, KI36: 11744, and KIA2: 11181). **i** Mutational signature of strand-specific substitutions (observed over expected) on a 96-trinucleotide context for de novo mtDNA variants identified by mgatk. Cosine similarity ($\cos(\theta)$) between mutant lines and the control line indicates the mitochondrial genome mutational signature to be non-distinctive from the control signature.

VAF-dependent mtDNA variants dynamics in *POLG*^{D274A} cells

Given the multi-copy nature of the mitochondrial genome, mtDNA variants may be present only in a fraction of the copies within a cell or population, a state known as heteroplasmy. As mtscATAC-seq yields a high mtDNA depth per cell (up to 400x and 800x in control and *POLG*^{D274A} knock-in cells, respectively), we determined the number of variants with a VAF exceeding a specified threshold. First, we detected a median of 35 high-confidence mtDNA variants per CTRL cell and 904 and 1487 in the two *POLG*^{D274A} cell lines regardless of their VAFs (Fig. 2a). The total number of mtDNA variants per cell primarily correlated with mtDNA sequencing depth, particularly at lower VAF levels (Supplementary Fig. 3). However, the mtDNA-depth-dependent increase in mtDNA variants detected per cell began to plateau for variants >3% VAF and higher and became mostly independent of depth >10% VAF (Supplementary Fig. 3). This aligns with the expectation that

the total detected number of low VAF variants will be a function of recovered mtDNA reads. In contrast, for higher VAF variants, their confident detection does not further benefit from deeper coverage. For further analysis, we opted to normalize the total number of mtDNA variants per cell by cellular mtDNA sequencing depth (Fig. 2b; “Methods”), after which the two *POLG*^{D274A} lines still showed an increase in the depth-normalized number of mtDNA variants per cell (median: CTRL: 0.769, KI36: 7.01, and KIA2: 6.97).

To refine the mutational analysis, we only considered ‘shared variants’ between the control and *POLG*^{D274A} lines, revealing 593 mtDNA parental variants. Notably, KIA2 showed a higher number of shared variants with the parental CTRL line (*n* = 586), compared to KI36 (*n* = 407), likely indicative of their more direct clonal relationship (Fig. 2c). In contrast, variants unique to KIA2 and KI36 cells were classified as “*POLG*^{D274A}-line specific” variants, which likely arose

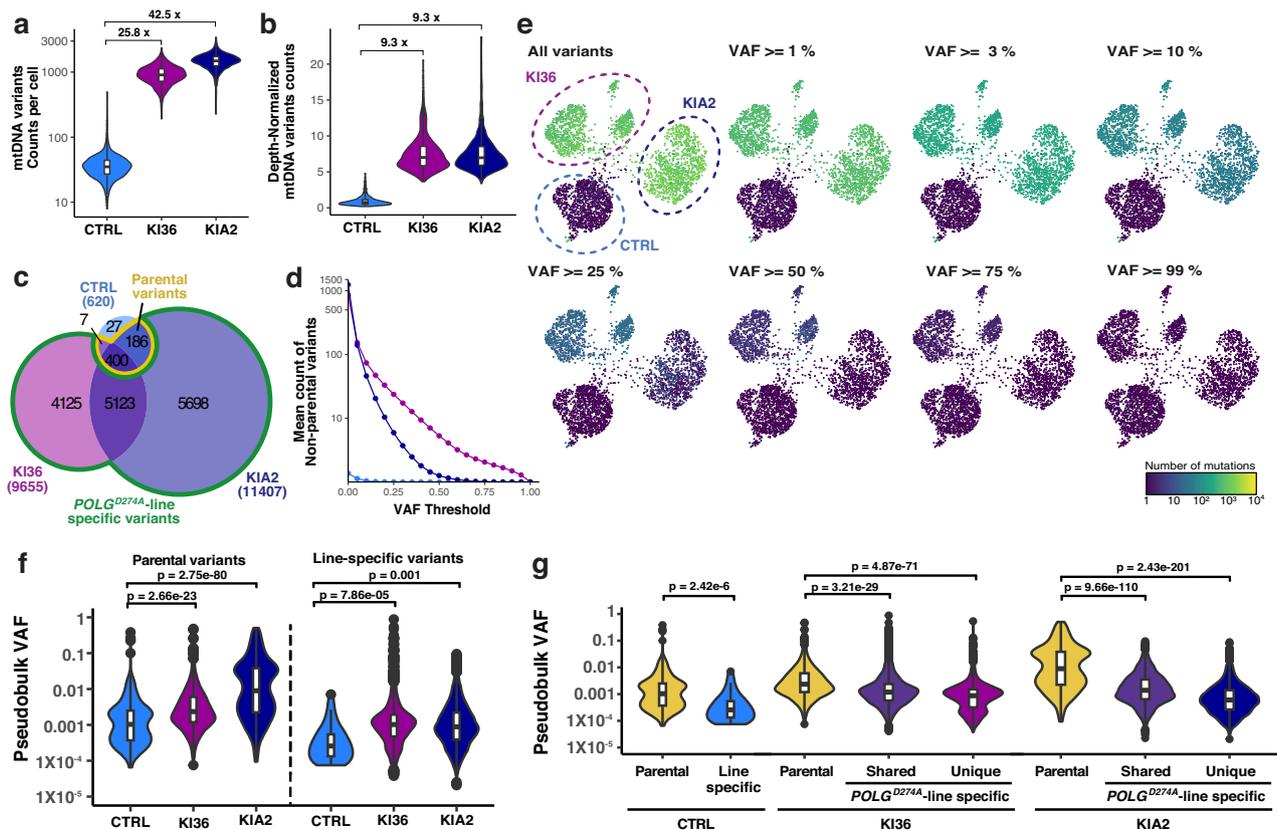


Fig. 2 | Heteroplasmy distribution and heterogeneity in *POLG*^{D274A} cell lines. **a** Violin plots of single cell raw mtDNA variant counts ($n = 1224, 1069, \text{ and } 1039$ for CTRL, KI36, and KIA2, respectively), with the fold-changes of the median relative to the control cell line being indicated. **b** Violin plots of single-cell mtDNA-depth-normalized mtDNA variant counts. Cell numbers were the same as described in panel (a). For panels (a, b), boxes show the median (center line) and the interquartile range (25th–75th percentiles; box). The fold-changes of the median relative to the control cell line were indicated. **c** Venn diagrams depicting shared and unique somatic mutations called by the mgatk pipeline. Parental variants, highlighted within the yellow circle, were defined as variants present in both *POLG*^{D274A} and control cell lines. **d** Line graph showing the mean count of non-parental somatic mtDNA variants (y-axis) at 5% incremental heteroplasmy thresholds. **e** UMAP projection of the color-scaled distribution of the number of non-parental somatic mtDNA variants above indicated variant allele frequency thresholds. **f** Comparison of mean pseudobulk heteroplasmy between parental ($n = 593, 407, \text{ and } 586$ for CTRL, KI36, and KIA2, respectively) and line-specific ($n = 27, 9248, \text{ and } 110821$ for CTRL, KI36, and KIA2, respectively) mtDNA variants stratified by cell lines. **g** Comparison of mean heteroplasmy across cell lines stratified by shared and line-specific (unique or shared between two *POLG*^{D274A} cell lines) mtDNA variants. The number of variants is as follows: parental, $n = 593, 407 \text{ and } 586$ for CTRL, KI36 and KIA2, respectively; CTRL-line-specific, $n = 27$; *POLG*^{D274A}-line-specific shared, $n = 5123$; *POLG*^{D274A}-line-specific unique, $n = 4125, 5698$ for KI36 and KIA2, respectively. Whiskers extend to the 5th and 95th percentiles. For panels (f, g), 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.

‘de novo’, though it is possible that these were already present at very low frequency in the parental line. Filtering cell line-specific variants by their VAFs notably showed a distinct distribution, with the number of low-frequency mtDNA variants, in particular, showing a 50- to 100-fold increase in *POLG*^{D274A} compared to control cells (Fig. 2d, e). This is consistent with constantly newly arising variants initially only being present in a single mtDNA copy, thus at low VAFs. Over time, relaxed replication and genetic drift can increase their prevalence across additional mtDNA copies. As such, the total number of detected mtDNA variants per cell progressively decreases when higher VAF thresholds are applied (Fig. 2d, e). Along these lines, parental variants displayed higher pseudobulk VAF levels than “line-specific” variants (Fig. 2f, g). These observations align with anticipated heteroplasmy dynamics of somatic events, with higher VAF levels requiring further genetic drift and vegetative segregation³⁶. Together, these analyses demonstrate that mtscATAC-seq sensitively detects de novo somatic mtDNA variants in single cells, even at low allelic frequencies.

To further examine the VAF distribution of the most variable variants, we categorized them into synonymous, missense, truncating, and pathogenic mtDNA variants previously reported in MITOMAP³⁷

(Fig. 3a–d; “Methods”). Synonymous and missense mtDNA variants showed broad VAF distributions, with a substantial fraction of variants reaching higher VAF (25–50%) in single cells. Collectively, these variants exhibited a wide range of Shannon entropy values, reflecting diverse heteroplasmy VAF patterns (Fig. 3b, c). In contrast, both pathogenic and truncating mtDNA variants exhibited highly skewed VAF distributions, with nearly all remaining at zero or low allelic frequencies and well below thresholds typically associated with mitochondrial dysfunction of pathogenic variants (e.g., 60–80%)¹⁶ (Fig. 3b, c). Moreover, out of the 130 reported pathogenic variants in MITOMAP (Supplementary Data 4), 43 and 41 variants were detected in KI36 and KIA2, respectively. These deleterious variants also exhibited consistently lower entropy values, suggesting restricted heteroplasmy dynamics at sub-threshold levels, indicative of their negative selection during continuous culture. Pairwise assessments of VAF distributions of truncating and pathogenic variants further revealed that many pathogenic variant pairs showed a dense cluster of cells with zero VAFs for both variants and appear only infrequently concomitantly detected in the same cells (Supplementary Fig. 4a). In contrast, truncating variants were more frequently observed together in individual cells,

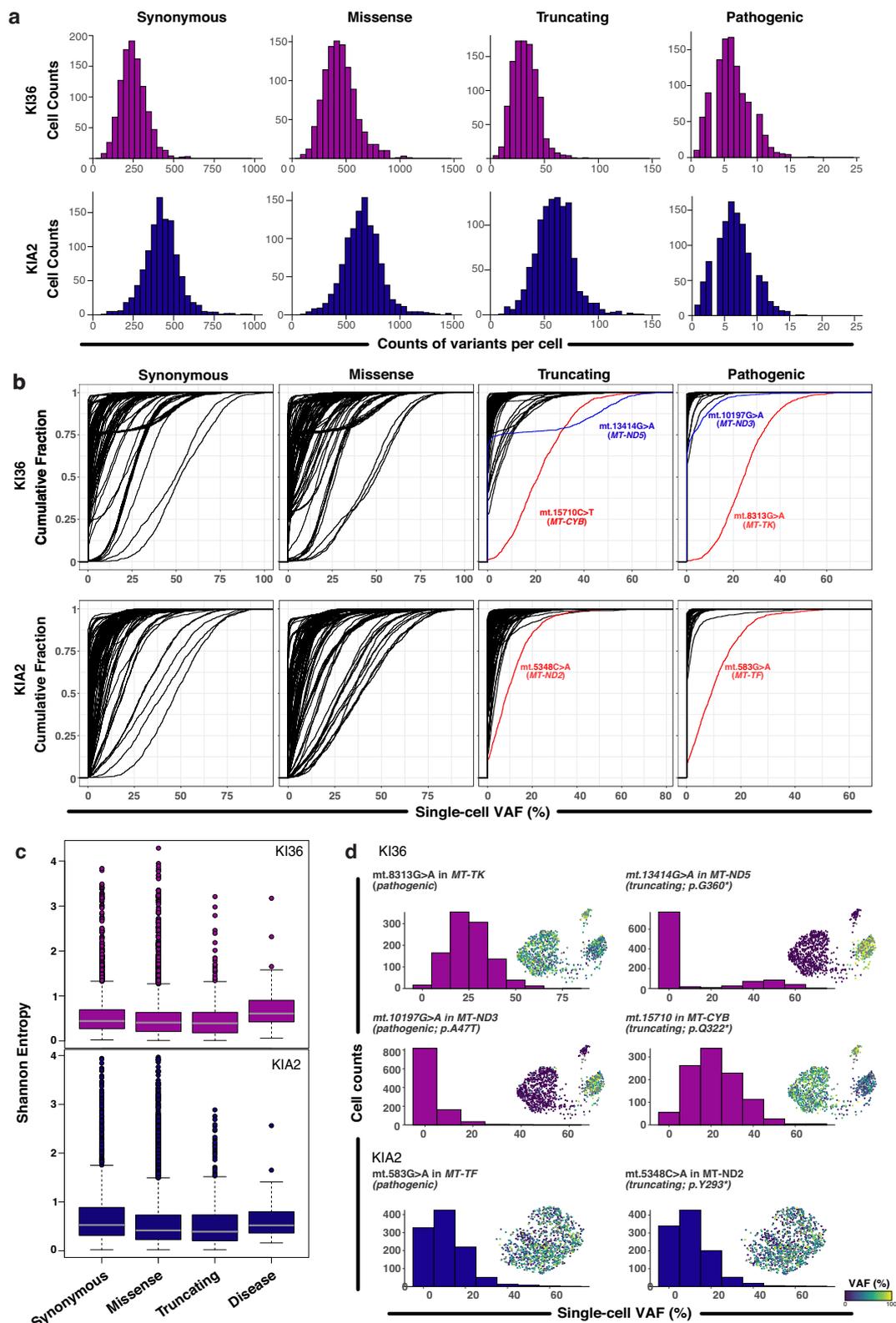


Fig. 3 | Restricted heteroplasmy dynamics of deleterious mtDNA variants in *POLG*^{D274A} cell lines. **a** Histogram showing the numbers of detected variants per cell, stratified by indicated mutation type (synonymous, missense, truncating, pathogenic). **b** Cumulative distribution plots of variant allele frequencies (VAFs) across single cells in the KI36 and KIA2 line, stratified by the indicated mutation type. Each curve represents a single variant, with select variants being highlighted. **c** Boxplots showing the distribution of Shannon entropy values for individual

variants, stratified by the indicated mutation type. 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. Variant counts (synonymous, missense, truncating, pathogenic): KI36, 2236, 4684, 339, and 43; KIA2, 2634, 5073, 463, and 41. **d** UMAPs and histograms showing the VAF distribution of mtDNA variants highlighted in (b).

including at higher but mostly <60% VAF, resulting in a greater number of double-positive cells (Supplementary Fig. 4b). In summary, *POLG*^{D274A} cells show complex mtDNA variant patterns, with truncating and pathogenic mtDNA variants being relatively depleted at higher VAF levels. These variants remain below classical biochemical thresholds in individual cells, suggesting that a complex and widespread mutational burden, as opposed to a single or few deleterious variants, may underlie *POLG*-mutation associated mitochondrial defects (Supplementary Fig. 1a). While mtscATAC-seq does not directly establish genotype-phenotype associations at the metabolic level, these findings provide important insights into *POLG* mutator genetics and motivate the systematic investigation of how a cumulative mutational burden, rather than single variants, impact mitochondrial, metabolic, and cellular function.

Single-cell mtDNA mutational burden in *POLG*^{D274A} cells

To date, most single-cell resolved mtDNA genotyping approaches have primarily focused on individual mtDNA variants rather than considering the complete variant call set per cell, which may further be highly variable across a population of cells^{11,19}. Specifically, individual *POLG*^{D274A} cells may carry hundreds of mtDNA variants at variable VAFs, and variable pathogenicity distributed throughout the mitochondrial genome, affecting different genes and respiratory complexes (Figs. 3, 4a). Motivated by the heterogeneous VAF distributions of mtDNA variants, we aimed to establish a framework to quantify and infer the functional consequences of the elevated total mtDNA mutational burden of individual *POLG*^{D274A} cells. Specifically, we introduced two metrics for evaluating the overall single-cell mtDNA mutational burden (scMMB). First, the number of mutations per million base pairs (MPM) is widely recognized for quantifying the mutational burden in the nuclear genomes of cancers^{38–40}. This metric has been adapted for assessing the mitochondrial genome mutational burden in bulk exome sequencing datasets⁶. By normalizing the number of mutations to sequencing depth, MPM further provides a standardized measure that facilitates comparisons across samples and conditions. Here, we adapt the principle of MPM quantification to single-cell mitochondrial genome sequencing data, which we termed single-cell mtDNA mutations per million base pairs (Fig. 4b, scmtMPM). Because mtDNA depth may vary from cell to cell, scmtMPM provides a normalized assessment of the overall mutational frequency of the mitochondrial genome and its heterogeneity in single cells.

In addition to scmtMPM, a recently introduced constraint model built on large-scale genomic data from gnomAD^{22,41} compares the observed genetic variations with those expected under neutral evolution to assess the local constraints of the mitochondrial genome. Specifically, a mtDNA local constraint (MLC) score enables the quantification of the local tolerance to base substitutions. The MLC score ranges from 0 to 1, with 1 indicating the most constrained positions and, consequently, the most likely to have a deleterious impact when mutated. Furthermore, the sum of MLC scores (MSS) across a sequenced whole genome enables the inference of the overall implications of mtDNA genome-wide variation, which has further been shown to be significantly associated with human phenotypes and the risk of disease progression^{22,41,42}. Because individual cells may exhibit unique VAF distributions of mtDNA variants² (Fig. 4a), we further account for their VAFs rather than only considering the presence or absence of a variant. For example, a heteroplasmic variant present at 80% VAF is expected to exert a more substantial phenotypic effect compared to when present at 5% VAF. As such, we propose a single-cell Heteroplasmy-Weighted MSS (scwMSS) metric, which weights the contributions of each variant MLC by its VAF to allow for a refined evaluation of the potential functional impact (Fig. 4b).

For simplicity, we first evaluated scmtMPM and scwMSS scores by functional groups of genes, i.e., encoding for complex I, III, IV, and V, rRNAs, and tRNAs, analogous to prior bulk-sequencing based studies,

for example, having described distinct mutational burden across complexes during tumor evolution⁶. As expected, *POLG*^{D274A} cells showed substantially elevated scmtMPM and scwMSS scores, which were independent of mtDNA sequencing depth (Fig. 4c, d; Supplementary Fig. 5a–d; median total scmtMPM, CTRL: 0, KI36: 2084, and KIA2: 1796; median total scwMSS, CTRL: 0, KI36: 16.8, and KIA2: 14.7). The presence of cluster-specific mtDNA variants thereby suggests independent subclonal trajectories underlying the acquisition of somatic mitochondrial genetic mosaicism (Fig. 4e).

Generally, the control cells exhibited an overall low mtDNA mutational burden across functional groups, with a few control cells showing elevated scores for complex I and IV (Supplementary Fig. 5b, c). In contrast, *POLG*^{D274A} cells showed a narrower distribution of scmtMPM and scwMSS scores, which were generally higher in the KIA2 derivative (Fig. 4c, d; Supplementary Fig. 5a–c). Overall, scmtMPM and scwMSS scores were well correlated, though this correlation was more apparent in *POLG*^{D274A} cells (Supplementary Fig. 5e), suggesting that both measures provide a means of quantifying the cellular genome-wide mtDNA mutational burden.

Further, we assessed correlations in the scmtMPM scores across functional groups and genes across all individual cells (Supplementary Fig. 6a, b). In the control line, correlations were overall weak ($r < 0.1$), and only a few cells displayed a high burden in more than one respiratory chain complex, suggesting the sporadic independent acquisition of mutations across functional groups and genes. However, in the hypermutated cells, the scmtMPM scores exhibited an elevated scmtMPM score and increased positive correlation (median $r = 0.33$, interquartile range 0.25–0.55) across functional groups and genes, consistent with a continuous and genome-wide somatic mutational process uniformly affecting the entire mitochondrial genome (Supplementary Fig. 6a, b). Collectively, these data highlight the utility of scmtMPM and scwMSS scores to quantify the genome-wide mtDNA mutational burden at the level of individual cells, complexes, and their interrelationships.

Depth-dependent mtDNA variant discovery and saturation

To evaluate how mtDNA sequencing depth impacts variant detection, we downsampled mtDNA reads to defined depths and quantified the number of detected variants in pseudobulk and in single cells. As expected, the number of total detected variants increased with depth. At shallow depths (<40 M reads; Supplementary Fig. 7a), the relationship was approximately linear, with a linear regression model yielding slopes of 17.9, 134.9, and 141 variants per million reads for CTRL, KI36, and KIA2, respectively. This suggests an 8-fold higher discovery rate in KI36 and KIA2 relative to CTRL. At higher sequencing depths (>50 M reads, Supplementary Fig. 7b), the rate of discovery plateaued in *POLG*^{D274A} lines. Next, we evaluated the impact of sequencing depth on mtDNA variant detection at the level of single-cells and observed a similar depth-dependent trend as in pseudobulk (Supplementary Fig. 7a, c, d). Linear regression models on the median number of single-cell mtDNA variants yielded slopes of 1.25, 12.27, and 16.77 variants per million reads for CTRL, KI36, and KIA2, respectively, indicating >10-fold higher discovery rates in KI36 and KIA2 relative to CTRL, consistent with the pseudobulk analysis.

Despite the expected depth-dependent effects on variant number, their heteroplasmy estimates were remarkably stable. Pseudobulk VAFs and variant composition remained concordant across depths (Supplementary Fig. 7e, f). Single-cell mtDNA heteroplasmy quantifications were highly consistent between 50 M (corresponding to ~25,000 mtDNA reads per cell) and 200 M (~100,000 mtDNA reads per cell) total mitochondrial reads, with most per-variant changes in VAF (Δ VAF) values within $\pm 5\%$ and a global Δ VAF distribution tightly centered at zero (Supplementary Fig. 7g). Likewise, the scmtMPM and scwMSS metrics remained relatively stable across a broad depth range and showed minimal bias even at 20 M total mtDNA reads (~10,000

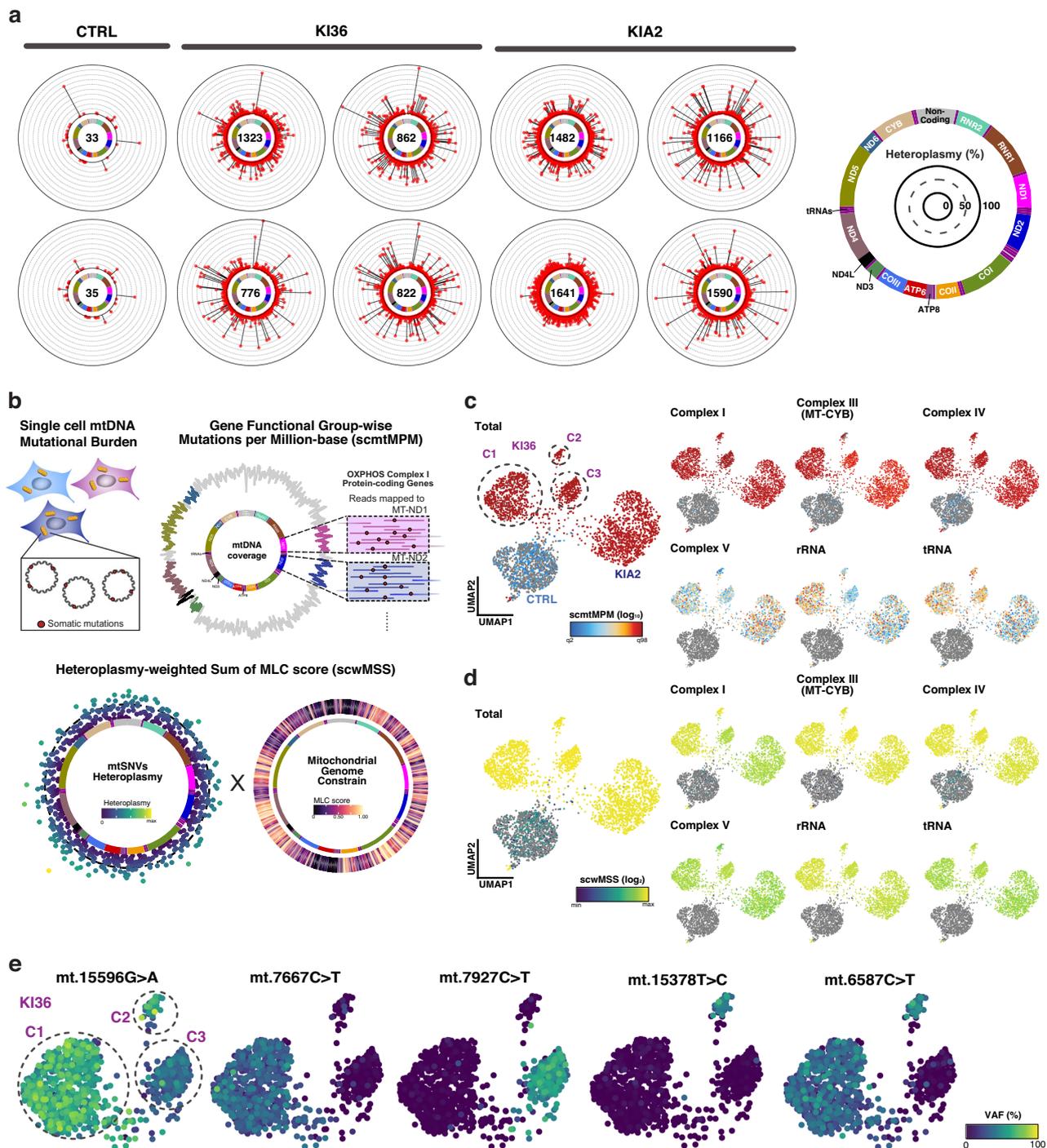


Fig. 4 | Quantification of single-cell mtDNA mutational burden via scmtMPM and scwMSS. a Examples of mtDNA SNV heteroplasmy distribution in randomly selected single cells around the mitochondrial genome. The number of confidently detected variants is indicated in the middle. Each red dot represents a single mutation at the indicated variant allele frequency levels. Mitochondrial genes are annotated, as shown on the right. **b** Schematic of single-cell mtDNA mutational burden quantitative metrics: Mutations per million bases (scmtMPM) and

Heteroplasmy-weighted sum of mitochondrial local constraint score (scwMSS). **c** UMAP projections of scmtMPM for indicated respiratory chain complexes, tRNA, and rRNA genes in *POLG*^{D274A} HEK293 cells. **d** UMAP projections scwMSS for indicated respiratory chain complexes, tRNA, and rRNA genes in *POLG*^{D274A} HEK293 cells. **e** UMAP projections of representative clonal somatic mtDNA variants and their VAFs (color scale) within KI36 cell line-specific subclusters C1–C3.

reads per cell; Supplementary Fig. 7h), in contrast to the strongly depth-dependent raw mutation counts. Together, these analyses indicate that while sequencing depth is an important technical factor impacting variant discovery, mtDNA heteroplasmy and mutational burden remain stable once moderate coverage is achieved, and depth is appropriately accounted for.

Galactose amplifies mtDNA copy number without clonal selection

We next examined how *POLG*^{D274A} cells respond to metabolic stress imposed by galactose^{43–47}. Specifically, galactose is selectively lethal to cells carrying pathogenic mtDNA variants⁴⁶ and selects against cells with higher large-fragment mtDNA deletion heteroplasmy⁴⁷. Notably,

CTRL cells showed a transient delay in cell growth but proliferated normally by day 3. In contrast, *POLG*^{D274A} cells showed substantial cell death, with KIA2 cells not tolerating galactose treatment and only a fraction of KI36 cells surviving, indicating a pronounced selective pressure (Supplementary Fig. 8a).

To test whether the selection dynamics occurred at the level of individual variants or the overall mtDNA mutational burden, we performed mtscATAC-seq of CTRL and KI36 cells cultured in 10 mM glucose or galactose for 1 and 3 days, yielding a total of 6165 and 7664 high-quality CTRL and KI36 cells, respectively (Fig. 5a; Supplementary Fig. 9b–d). In line with our baseline characterization of the *POLG*^{D274A} lines (Fig. 1b), CTRL and KI36 clustered separately (Fig. 5a; Supplementary Fig. 9a, b). Notably, galactose-treated KI36 cells clustered more distinctly in response to galactose, likely reflecting a compensatory response (Fig. 5b). At the level of mtDNA variants, we did not detect notable shifts in pseudo-bulk heteroplasmy levels in either CTRL or KI36, nor did we observe alterations in the overall distribution of single-cell VAFs under galactose compared to glucose conditions (Fig. 5d, e; Supplementary Fig. 9d–f). Moreover, the mutational burden remained comparable between glucose and galactose treatment (Supplementary Fig. 9f). However, there was a significant increase in mtDNA copy number in both CTRL and KI36, with KI36 showing a more pronounced increase (Fig. 5c; Supplementary Fig. 9c; fold change of normalized median mtDNA depth for Gal Day 3 vs. Glu Day 3: 1.3× in CTRL and 2.3× in KI36), indicating that mtDNA amplification represents the primary mode of compensation in this setting^{47–49}.

To investigate correlating changes in chromatin accessibility, we focused on two large subclones marked by specific somatic mtDNA variants and compared their epigenetic state in glucose vs. galactose treatment (Fig. 5f). Differentially accessible gene (DAG) analysis revealed 112 and 63 DAGs in C1 and C3, respectively (Supplementary Data 5, 6). Among these, 10 DAGs were upregulated in galactose treated cells in both C1 and C3 and involved genes in mitochondrial/ER stress and inflammatory signaling pathways (e.g., *CCL20*, *TRIB3*, *INHBE*, *ARHGAP2*)^{50–52}, and four non-coding RNAs (i.e., *DLGAPI-AS2*, *MIR6738*, *MIR6864*, and *MIR6865*), likely reflective of metabolic rewiring and adaptation to OXPHOS-enforcing conditions in galactose (Fig. 5g). Interestingly, *GDF15*, a well-known biomarker of mitochondrial diseases^{33–35}, was also identified as DAG in C1, showing a gradual increase in chromatin accessibility under continuous galactose treatment (Fig. 5h). As strongly pathogenic mtDNA variants were largely absent in this context, galactose appeared to have a predominantly uniform effect across the population rather than selecting for a favorable mtDNA variant profile, with elevation in mtDNA copy number being the primary mode of adaptation.

Heterogeneous mtDNA somatic mosaicism in human blood cells

Maternally inherited mtSNVs and mtDNA somatic variants acquired during aging shape the mutational landscape and have been implicated in human phenotypes, including aging, metabolic disease, neuromuscular disease, cancers, and congenital mitochondrial diseases^{1,2,7,9,56}. Moreover, the co-occurrence and mutual exclusivity of mtSNVs suggest complex interactions among mtDNA variants^{19,56}. However, the dynamics of mtSNVs and the impact of pathogenic mtDNA variants on the mtDNA mutational landscape remain largely unexplored. To address these questions, we first evaluated the applicability of scMMB scores to primary human cells. We analyzed PBMCs from two healthy individuals (5- and 47-year-old) and four patients (29-, 35-, 60-, and 80-year-old) with mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS) syndrome (Fig. 6a). Across all individuals, scmtMPM and scwMSS scores were well correlated and independent of mtDNA depth (Supplementary Fig. 10). Additionally, an age-related increase in the number of detected mtDNA variants was observed across healthy donors and patients (Supplementary Fig. 11a, b), consistent with prior

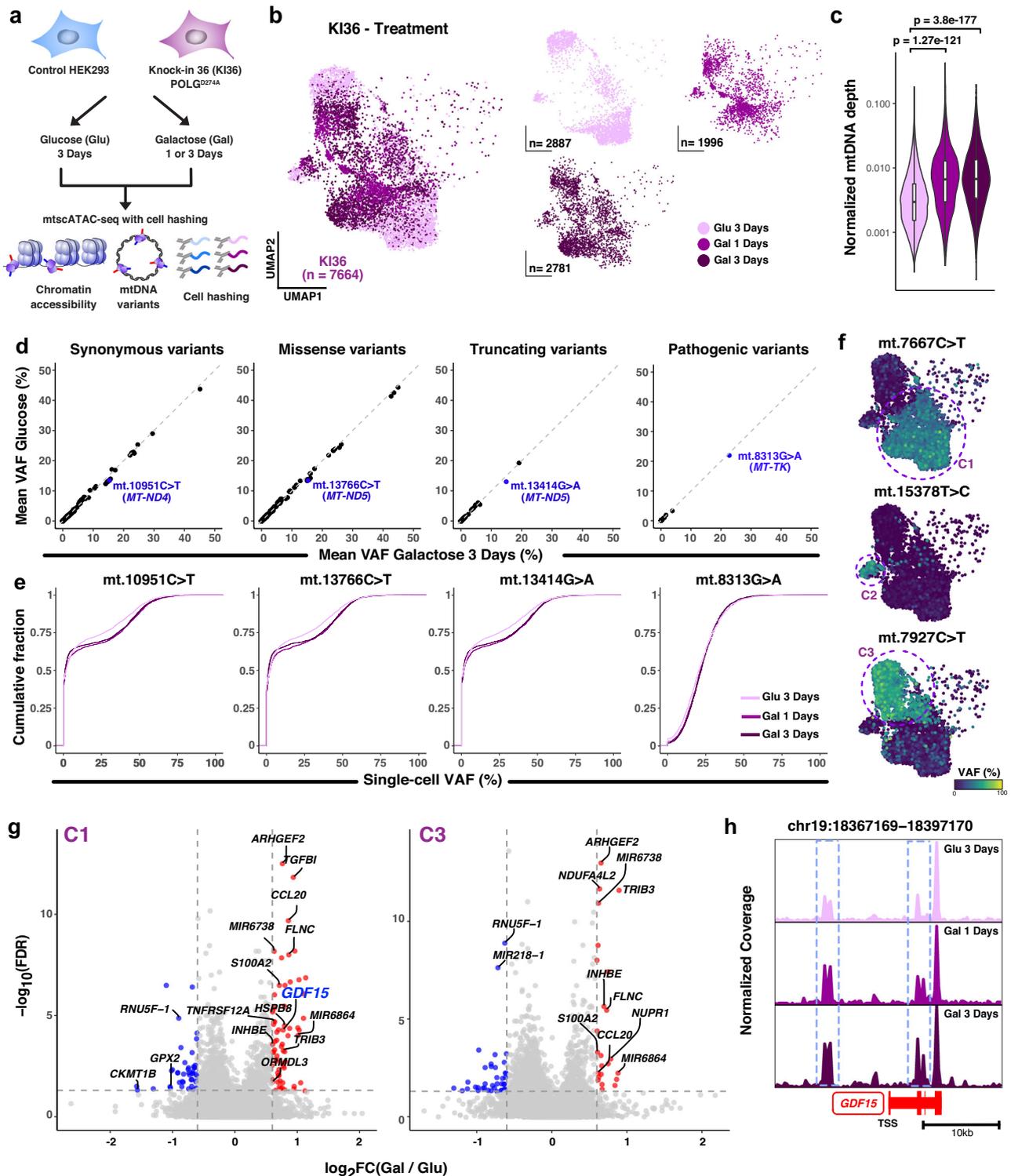
observations³. However, a more heterogeneous distribution of scmtMPM and scwMSS scores in primary cells, compared with cell lines, was observed in a cell-type- and donor-specific manner (Fig. 6b, c; Supplementary Fig. 5).

Specifically, in the pediatric donor, groups of cells with a relatively higher mutational burden were observed across different immune cell types (Fig. 6c; top left, arrows), suggesting the non-uniform acquisition and distribution of somatic variants across cell types. As expected, the immune cells of patients with pathogenic mtDNA variants showed higher scmtMPM and scwMSS scores and a more substantial bimodal distribution (Fig. 6b; red violin plots). As the pathogenic mt.3243A > G mutation undergoes purifying selection in specific T cell subsets and the hematopoietic system throughout the lifespan of patients (Supplementary Fig. 11c)^{10,11,13}, we also reevaluated scMMB metrics after excluding the pathogenic variant. Notably, the removal of the mt.3243A > G mutation abolished the bimodal distribution in the younger patients (29- and 35-year-old). However, this reduction was not observed in the older patients (60- and 80-year-old), suggesting that while the mt.3243A > G mutation may be gradually lost, new mutations emerge with age contributing to elevated scmtMPM and scwMSS scores (Fig. 6b). Additionally, a significant fraction of cells with no detectable mt.3243A > G heteroplasmy, showed relatively higher scmtMPM scores (Fig. 6d, e; arrows) compared to cells even with low mt.3243A > G VAFs, suggesting the presence of a highly pathogenic congenital mtDNA variant to potentially exert selective pressures to accumulate additional somatic mtDNA variants. Notably, reevaluation after ‘removing’ mt.3243A > G showed that the scmtMPM metric remains low in cells with detectable levels of the variant, regardless of mt.3243A > G VAFs, mtDNA depth, and cell type (Fig. 6e, f; Supplementary Fig. 11d). Together, scmtMPM and scwMSS scores enable the single-cell resolved evaluation of patterns of pathogenic mtDNA variants and demonstrate their impact on somatic mosaicism in primary human cells.

Increased complex-specific mtDNA mutational burden across lineages

To further investigate the nature of elevated scMMB scores, we evaluated whether they are attributable to a single highly constrained variant or to the cumulative effect of multiple mtDNA mutations. For example, a single highly constrained variant would over proportionally contribute to elevated scMMB scores. To assess this, we deliberately removed the variant with the highest VAF heteroplasmy, after which scMMB scores decreased significantly, suggesting that a single somatic variant profoundly affected scMMB scores (Fig. 7a; yellow violin plots). Next, we stratified scores based on gene and functional groups to further resolve the underlying heterogeneity of scmtMPM and scwMSS in cells with pathogenic mtDNA (Fig. 7b). As expected, MELAS patient cells showed higher scores for tRNA genes, and cell-type-specific analysis further revealed tRNA scores and pathogenic mt.3243A > G variant heteroplasmy to be highly correlated (Supplementary Fig. 11e, f). Interestingly, distinct distributions of respective scores were observed among all functional groups. While scmtMPM for complex I, complex IV, and mt-rRNA genes were generally elevated, a more bimodal distribution was observed for complex III and V genes (Fig. 7b). Moreover, the contribution to scwMSS scores for complex III and V variants was generally lower (Fig. 7b; lower panel), likely due to the smaller number and shorter length of their genes, as well as their overall relatively lower MLC scores compared to other mtDNA gene groups (Supplementary Fig. 11g, h)⁴². These findings indicate distinct mutation dynamics, as also previously intimated across cancers, where complex I subunits were enriched for pathogenic mutations, whereas complex V mutations were broadly depleted for all non-synonymous mutations⁶.

Upon further evaluation of elevated scMMB scores across distinct functional groups, we followed up on the unexpected observation of



high scmtMPM and scwMSS scores in complex I for the pediatric donor (H05) and two MELAS patients (M60 and M80). Specifically, individual gene-level analysis highlighted aberrant increases in scMMB for *MT-ND4L* in H05 and *MT-ND3* in both M60 and M80 (Fig. 7b, c). To identify the underlying mtDNA variants, we ranked all detected variants by their pseudo-bulk heteroplasmy. Notably, in each instance, only a single variant exceeded 1% VAF: mt.10599G > A in *MT-ND4* in the H05 (mean VAF = 8%), mt.10270T > C in *MT-ND3* in M60 (mean VAF = 25%), and mt.10398A > G in *MT-ND3* in M80 (mean VAF = 5.8%), respectively (Fig. 7c, d; Supplementary Table 1). The mt.10599G > A variant in H05

showed a broad and uniform representation among immune cells, including T, NK cells, and monocytes, marking 22% of the entire population of cells (Fig. 7d). Given the low likelihood of such a pronounced clonal expansion at the donor's young age, this suggests the mt.10599G > A variant to have already been present at higher VAF in the zygote, upon which it further distributed to multiple multipotent hematopoietic stem cells during development as recently suggested by the whole genome-sequencing based evaluation of cellular phylogenies of hematopoietic colonies^{3,57}. Moreover, we observed pronounced enrichments of cells with high mt.10559G > A VAFs in both

Fig. 5 | Galactose-induced metabolic stress leads to a global cellular response without clonal selection in *POLG*^{D274A} cells. **a** Schematic overview of the experiment. CTRL and *POLG*^{D274A} KI36 cells were cultured in 10 mM glucose for 3 days or 10 mM galactose for 1 or 3 days, followed by mtscATAC-seq with cell hashing for joint profiling. **b** UMAP embeddings of KI36 single cells colored by treatment condition. Glu, Glucose; Gal, Galactose. *n* = cell numbers. **c** Violin plots of single-cell mtDNA depth normalized to the number of unique ATAC read counts per condition in *POLG*^{D274A} KI36 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** Scatter plots showing per-variant pseudobulk VAFs in KI36 cells treated

with galactose for 3 days (y-axis) vs. glucose for 3 days (x-axis), per indicated (functional) category. **e** Cumulative distribution plots of single-cell VAFs for the most shifted variants highlighted in (e), stratified by treatment condition. **f** UMAP projections of KI36 cells colored by the VAF of subclone-specific somatic mtDNA variants as identified in Fig. 4e. **g** Volcano plots of differentially accessible genes (DAGs) in subclones C1 (left) and C3 (right) comparing galactose vs. glucose treatment (3 days). Statistically significant upregulated (FDR ≤ 0.05 with log₂FC ≥ 0.6) DAGs in galactose treatment are highlighted in red, and downregulated (FDR ≤ 0.05 with log₂FC ≤ -0.6) DAGs are highlighted in blue. The Wilcoxon pairwise test method was utilized to identify statistically significant DAGs. **h** Pseudobulk chromatin accessibility track plot at the *GDF15* locus. The blue dashed boxes highlight select differentially accessible peaks.

CD4⁺ and CD8⁺ memory T cells, indicating the mtDNA variant tracks with clonal expansion and memory formation (Fig. 7e). In contrast, the mt.10270T > C variant in M60 exhibited a narrow and exclusive distribution marking 76% of CD8⁺ T effector memory cells (Fig. 7d). In concordance with a very high MLC score of 0.969, SIFT and Alpha-Missense predict mt.10270T > C to be deleterious, which likely explains why this variant has not been reported in ClinVar due to a strong purifying selection at the population level²². However, the strong enrichment of the mt.10270T > C mutant allele in memory CD8⁺ T cells suggests substantial clonal expansion and lineage-specific tolerance toward a pathogenic variant (Fig. 7e). This specific population also notably deviated from the regression line of the scmtMPM and scwMSS (Supplementary Fig. 10b; red arrow), indicating that while both scores are generally correlated, scwMSS integrates the pathogenicity of variants that are not fully reflected by the overall mutational burden alone. The mt.10398A > G variant in M80 is present in various cell types, with 25% of CD8⁺ naive T cells marked (Fig. 7d), again suggesting the occurrence of likely clonal events associated with individual mtDNA mutations. This variant has been reported to be pleiotropic and associated with distinct disease phenotypes in Mitochondrial, such as Leigh syndrome, breast cancer, metabolic syndrome, and Parkinson's disease⁵⁸. While this variant appeared to have medium-to-low VAFs in most immune cell populations, we observed enrichment for cells with high mt.10398A > G VAFs in monocytes and CD8⁺ T cell effector/memory populations (Fig. 7e), again suggesting a potential lineage-specific preference for individual mtDNA mutations. In summary, the evaluation of cellular scMMB scores readily identifies functionally relevant complex- and gene-specific variants, providing insight into their distribution, potential pathogenicity, and implications for cell lineage and mitochondrial function, and motivating further functional investigation.

Discussion

The comprehensive analysis of mtDNA genetic variation from population-scale databases, such as mitochondrial gnomAD⁵⁹ and mitoMAP³⁷, has broadened our understanding of the essentiality and local constraint across the genes of the mitochondrial genome^{6,22,60}. However, the bulk nature of the applied sequencing assays likely vastly underestimate the full spectrum of genetic variation, with recent advances in single-cell mtDNA genotyping providing avenues to reveal cellular somatic mosaicism and cell-type-specific responses toward mtDNA mutations^{10,11,19,36}. However, analytical frameworks for interpreting mitochondrial genetic variation at the single-cell level are lacking, limiting our understanding of mtDNA variant burden and dynamics. Here, we applied single-cell mtDNA genotyping using mtscATAC-seq to quantify mtDNA mutational burden across individual cells via adapting bulk mtDNA burden and constraint metrics for single-cell analysis^{6,22,41,42} (Fig. 4). Both scmtMPM and scwMSS are mtDNA depth-normalized and account for cell-specific VAF variation, enabling more precise and standardized assessment of mtDNA mutational burden at the single-cell level (Supplementary Figs. 5, 7, 10).

By leveraging exonuclease-deficient *POLG*^{D274A} HEK293 cells, we benchmarked this analytic framework to quantitatively and qualitatively estimate the cellular mtDNA mutational burden (Figs. 2–4). We revealed a substantially higher number of somatic mtDNA variants than bulk approaches and provided insights into their heterogeneous distribution at unprecedented resolution. Notably, unlike synonymous and missense variants, pathogenic and truncating variants were observed predominantly at subthreshold levels, indicating their negative selection. Further, these findings suggest *POLG*^{D274A} phenotypes to be driven by the cumulative mutational burden as compared to the impact of individual or a few deleterious variants as observed in mitochondrialopathies (Fig. 3). Additionally, we systematically evaluated the impact of sequencing depth on mtDNA variant detection and mutational burden quantification (Supplementary Fig. 7). Although higher sequencing depth increased discovery, mutational burden metrics stabilized once moderate per-cell coverage was achieved, with ~25,000 mtDNA reads/cell optimal for mutation counts and ~10,000 sufficient for robust burden estimation.

We further treated *POLG*^{D274A} cells with galactose, exerting a strong metabolic pressure driving reliance on OXPHOS (Fig. 5). Despite adaptive responses, such as mtDNA amplification and concomitant chromatin accessibility changes in genes related to mitochondrial stress, the pseudo-bulk heteroplasmy and single-cell VAF distributions of individual mtDNA variants remained largely stable without clear evidence of (clonal) selection. However, the uniformly high mutational burden and scarcity of pathogenic mtDNA variants in the KI36 line may drive these results, and a system with more heterogeneous and pathogenic mtDNA variant profiles is likely to have revealed distinct variant dynamics. Moving forward, more sensitive sequencing approaches, including duplex sequencing⁶¹ and their adaptation to single-cell omics, will further enhance the resolution of low-frequency heteroplasmic variants.

In the context of immunity, we further resolve mitochondrial genetic heterogeneity in humans (Figs. 6, 7). In patients with MELAS, we identified several likely functional mtDNA mutations that had previously eluded clinical diagnosis. Together, this work underscores the utility of single-cell mtDNA sequencing to provide insights into mitochondrial genome stability and disease associations. Specifically, recent studies have demonstrated that mtDNA heteroplasmy may be clinically informative^{6,7,17,41,42,62}. In population-scale bulk-sequencing data, MSS scores were associated i) with increased cancer-specific mortality⁴¹, ii) the presence of clonal hematopoiesis, and iii) a higher risk of developing myeloid neoplasms⁴². Additionally, in metastatic melanoma⁶², tumor lesions harboring mtDNA mutations with >50% VAF showed a 2.6-fold greater response to immune checkpoint blockade. Collectively, our framework motivates the prospective integration of single-cell mtDNA mutational burden analysis, experimental base-editing^{36,63–65}, alongside other clinical indicators to assess risk and treatment responses in cancer, as well as probing the biology of mitochondrial dysfunction in aging^{14,15}.

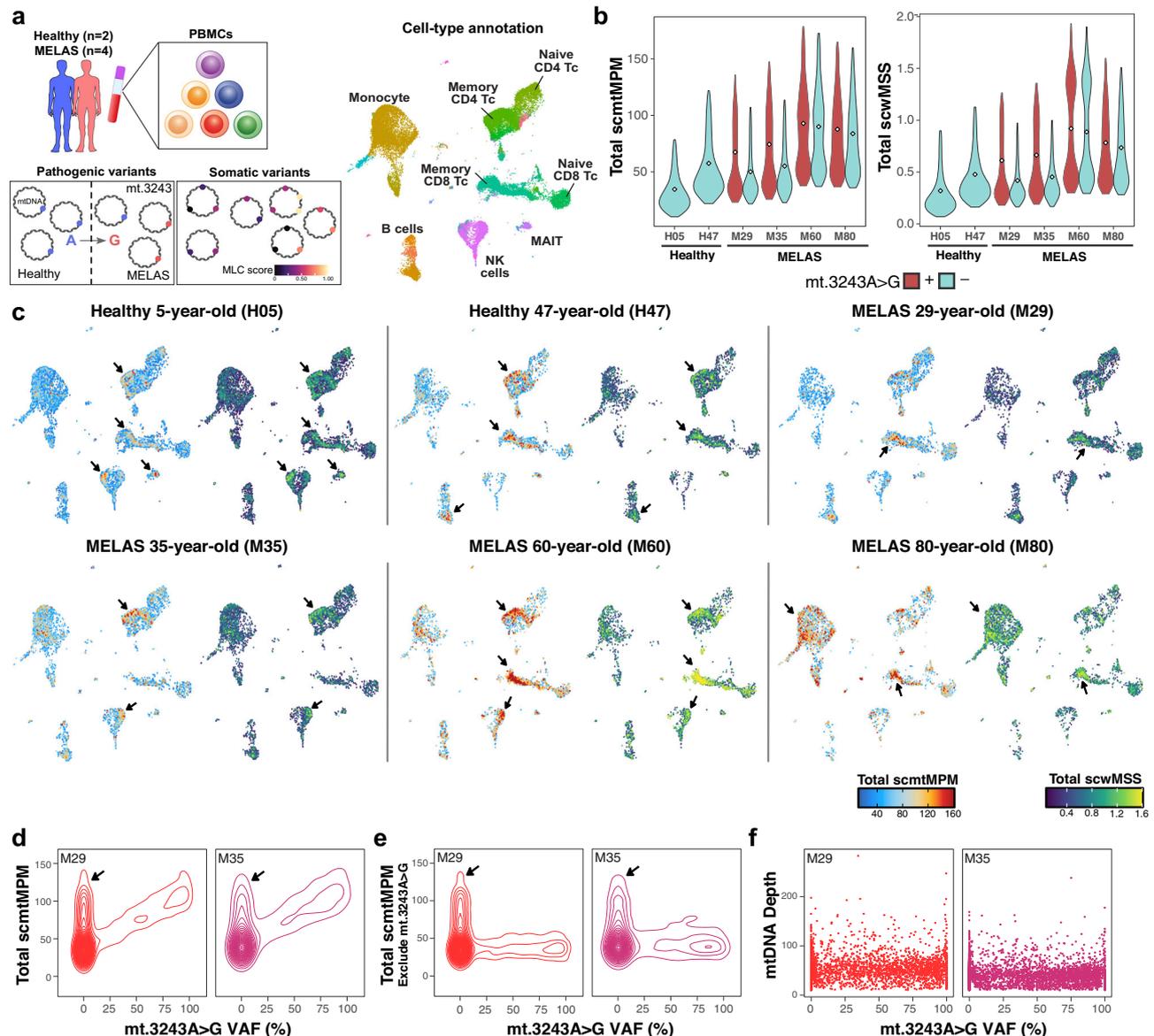


Fig. 6 | Quantification of mtDNA mutational burden in human peripheral mononuclear blood cells in healthy individuals and MELAS patients.

a Schematic representation of the available datasets and analysis design. The human icon by Servier ([smart.servier.com](https://www.servier.com)) is licensed under CC-BY 3.0 Unported.

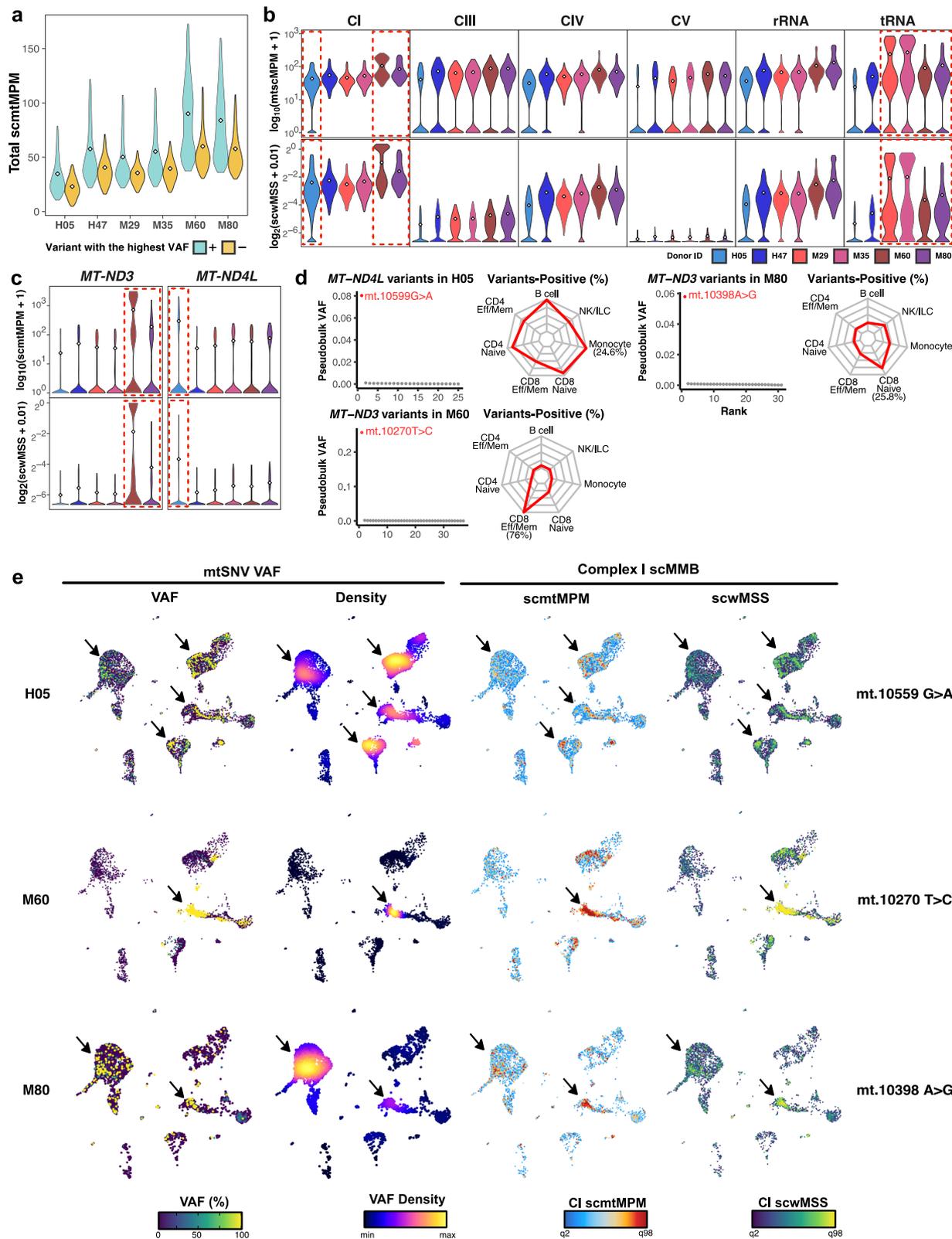
b Violin plots of mitochondrial genome-wide mutation burden metrics +/- the pathogenic variant mt.3243A > G. The white dots indicate the pseudobulk value. **c** Azimuth reference UMAPs colored by total scmtMPM (left panel) and scwMSS (right panel) after excluding the pathogenic mt.3243A > G variant in the *MT-TL1* gene for six indicated donors. “H” indicates a healthy donor, and “M” indicates a

donor with MELAS. The two-digit number represents the donors’ ages. **d** Contour plots illustrate the relationship between the mt.3243A > G heteroplasmy (x-axis) and total scmtMPM (y-axis) in patients M29 and M35. Contour lines indicate data density, with bins set to 50. Peaks in density at low/absent heteroplasmy are marked with arrows. **e** Contour plots showing the relationship between cells with mt.3243A > G heteroplasmy (x-axis) and their total scmtMPM (y-axis) in samples M29 and M35, with the pathogenic mt.3243A > G variant being excluded from the scmtMPM score. **f** Correlations between mt.3243A > G heteroplasmy and the mtDNA depth.

While scRNA-seq remains a popular method for single-cell profiling^{41,42,62}, variant calling from mtRNA transcripts is subject to RNA editing and high false-positive rates, potentially attributable to transcriptional polymerase error. Moreover, sequence coverage is limited in 5’ or 3’ based sequencing techniques^{18,66}. To assess this, we analyzed DOGMA-seq libraries that co-capture mtDNA variation alongside 3’-primed mRNA libraries from the same cell^{11,67}. As expected, mtRNA-derived sequence reads exhibited a significantly more uneven and 3’-enriched coverage compared to mtDNA reads (Supplementary Fig. 12a, b). Accordingly, we observed a significant disparity in the degree of identified variants between mtDNA and mtRNA, with more detectable mtDNA variants in 75% of the cells

(Supplementary Fig. 12c). Given the possible substantial contribution of individual mtDNA variants to cellular scMMB scores that may not be recoverable in mtRNA and the high false-positive rate⁶⁶, we warrant caution surrounding the sole transcriptomic-focused analysis of functionally relevant mtDNA variation.

Given the heteroplasmic nature of mtDNA, it is widely accepted that a non-synonymous/pathogenic variant must reach a biochemical threshold (typically around 60–80% VAF) to cause metabolic defects in cells⁶. However, with the advent of deeper sequencing techniques, large-scale population studies have identified certain regions of mtDNA that lack even low-VAF variants, suggesting that the threshold may be lower than previously expected^{2,68}. In fact, we previously



demonstrated purifying selection in patients' blood cells, where more than 50% of cells were depleted entirely of pathogenic mtDNA variants (Supplementary Fig. 11c; red arrow)^{10,11}. In line with these findings, the presence of a pathogenic mtDNA variant, even at low VAF, appears to impede the accumulation of additional somatic mtDNA variants (Fig. 6d, e), supporting the idea that low-VAF mtDNA variants can have biological impacts on the mtDNA mutational

landscape. While additional validation will be required, a recent study has demonstrated that the co-occurrence of 'incompatible' common variants contributes to severe mitochondrial disease⁵⁶, further supporting the notion of pronounced genetic interactions. Together, these observations underscore the complex interplay among mtDNA variants and highlight the importance of employing genomic methods to detect the subtle yet notable effects of mtDNA

Fig. 7 | Individual somatic mtDNA variants drive cellular scMMB scores and show cell-type-specific distributions. **a** Violin plots of mitochondrial genome-wide mutational burden +/- the mtDNA variant with the highest VAF independent of the pathogenic variant mt.3243A > G. The white dots indicate the pseudobulk value. **b** Violin plots of total scmtMPM and scwMSS of human PBMCs stratified by the indicated mitochondrial gene functional groups. Dotted boxes highlight individuals with highly complex I and tRNA scMMB metrics. **c** Violin plot of scmtMPM and scwMSS for the indicated mitochondrial complex I genes. **d** Identification and cell

type biases of mtDNA variants contributing to elevated complex I mutational burden scores. Ranked mean heteroplasmy of mtDNA variants in the indicated mitochondrial genes and individuals. The radar plots depict the percentage (%) of the cells that carry the mtDNA variant in the indicated immune cell subsets. **e** UMAP projection colored by the heteroplasmy and density of the identified mtSNVs, and by the scMMB metrics, scmtMPM, and scwMSS scores, of OXPHOS complex I for the indicated individuals H05, M60, and M80. Compare to Fig. 6a for cell type annotation.

alterations, complementing traditional metabolic assessments of cellular function.

In conclusion, studies of mitochondrial genetics have come with unique challenges, given the variation in mtDNA copy number and distinct modes of replication and segregation, which have hampered the quantification of somatic mitochondrial mosaicism. Here, we benchmark mtscATAC-seq's capability to sensitively detect and quantify genome-wide mtDNA variants in single cells. We introduce a quantitative framework to assess mitochondrial genome-wide mutational burden and constraint. Their application to primary immune cells showcased their utility in sensitively detecting pathogenic mtDNA variants and variations in their cellular distribution. Together, these approaches provide a robust framework to systematically investigate fundamental aspects of mitochondrial genetics in single cells.

Methods

Our research complies with all relevant ethical regulations, including ethics approval and guidance from Charité—Universitätsmedizin Berlin, the Berlin Institute of Health at Charité, and the Max Delbrück Center for Molecular Medicine (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany.

Healthy donor and patient samples

Primary human peripheral blood samples were collected after informed consent and approval was granted by the local ethics commission at Charité—Universitätsmedizin Berlin, Germany (EA1/192/20, EA2/144/15, EA2/284/25). PBMCs were isolated using Ficoll Paque Plus (SigmaAldrich, GE17-1440) by density gradient centrifugation using SepMate tubes (StemCell Technologies) and were stored in vapor-phase liquid nitrogen after cryopreservation with FBS supplemented with 10% dimethyl sulfoxide until the day of the single-cell sequencing experiment.

Cell culture

POLG^{D274A} knock-in cell lines, with (Knock-in clone A2, KIA2) or without (Knock-in clone 36, KI36) the introduction of a tetracycline-inducible mitochondrial restriction exonuclease mitoEagI, were previously described²³. The knock-in line KI36 was derived from the parental control cell line (CTRL), with the KIA2 line being derived from more distantly related HEK293T cells. Briefly, control and KI36 HEK293 cell lines were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM; Gibco, 41966052) supplemented with 10% [v/v] FBS Standard, 50 µg/L Uridine, and 1 U/ml Pen-Strep. KIA2 HEK293T cell lines were supplemented additionally with 50 µg/ml Hygromycin B, and 15 µg/ml Blasticidin S Hydrochloride without the induction of mitoEagI. When the cell culture reached 90% confluency, cells were detached with Trypsin-EDTA (Gibco, 25200056) and split at 1:10 or cryopreserved. Cells were cultured for up to 2 weeks and underwent up to three passages before single-cell sequencing to limit the accumulation of additional mtDNA mutations.

Galactose treatment

Control and *POLG*^{D274A} knock-in cell lines were seeded in high-glucose DMEM and incubated at 37 °C in a 5% CO₂ incubator overnight to allow cell attachment. The next day, cells were rinsed with PBS, and the

medium was replaced with glucose-free DMEM (Gibco, 11966025) supplemented with 1 mM sodium pyruvate (Gibco, 11360070) and with either 10 mM glucose (3 days) or 10 mM galactose (1 day or 3 days) (ThermoFisher, J15995.A1), as indicated. On the day of single-cell sequencing, cells were washed with PBS to remove dead cells, detached with trypsin, and then washed with PBS. Each cell line-treatment combination was labeled with dedicated TotalSeq-B hashtag antibodies, washed, and pooled at equal ratios. The pooled cells were subsequently processed using the mtscATAC-seq protocol as described below.

Mitochondrial enzyme activity measurements

The activities of the mitochondrial enzymes NADH:CoQ₁ oxidoreductase (complex I, CI), cytochrome c oxidase (complex IV, CIV), and citrate synthase (CS) were measured as previously described⁶⁹. Briefly, cells were homogenized by sonication. Complex I activity was measured as the rotenone-sensitive rate of NADH oxidation in a medium containing 50 mM KCl, 1 mM EDTA, 10 mM Tris-HCl (pH 7.4), 2 mM KCN, 100 µM CoQ₁, and 150 µM NADH. Total NADH oxidation was recorded for 2 min before adding 20 µM rotenone to determine the rotenone-insensitive rate. Complex I activity was defined as the rate difference between total and rotenone-insensitive NADH oxidation. Complex IV activity was measured in 0.1 M potassium phosphate buffer (pH 7.4) containing 0.02% laurylmaltoside, and 200 µM ferrocytochrome c (reduced cytochrome c), and ferrocytochrome c oxidation was monitored at 510/535 nm. All measurements were performed at 30 °C using a dual-wavelength spectrophotometer (Aminco DW-2000, SLM Instruments, Rochester, NY, USA).

Cell hashing and FACS sorting

Cryopreserved cell lines were thawed in RPMI with 10% FBS by serial dilution. TotalSeqTM-A anti-human hashtag antibodies were purchased from BioLegend. Two million cells per cell line were resuspended in cell staining buffer (BioLegend, 420201) with Human TruStain FcX (BioLegend, 422301) for 10 min on ice. Cell lines were then stained with 0.25 µg of individual hashtag antibodies (see Supplementary Table 2 for barcode assignment) in the cell staining buffer. After staining, cells were washed three times with PBS/BSA. All three HEK293 cell lines were pooled at approximately equal ratios. For live/dead cell discrimination, Sytox Blue (ThermoFisher, S34857) was used at a 1:1000 dilution in PBS/BSA/EDTA. Cell sorting was conducted using the BD Bioscience FACS ARIA III cell sorter with a 100-µm nozzle at the MDC/BIH Genomics Platform.

mtscATAC-seq with cell hashing

mtscATAC-seq libraries were generated as previously described and according to the manufacturer's instructions (CG000209-Rev F) with the Chromium Controller and Chromium Next GEM Single Cell ATAC Reagent Kits v1.1 (PN-1000176) with minor modification and sample multiplexing with TotalSeqTM-A hashtag antibodies. Briefly, sorted cells were fixed in 1% formaldehyde (FA; ThermoFisher) in PBS for 10 min at room temperature, then quenched with glycine solution to a final concentration of 0.125 M before washing twice with PBS/BSA by centrifugation at 400 g for 5 min at 4 °C. Cells were subsequently permeabilized with lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl,

3 mM MgCl₂, 0.1% NP40, 1% BSA) for 3 min (primary cells) or 5 min (cell lines) on ice, followed by adding 1 ml of chilled wash buffer (10 mM Tris-HCL pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 1% BSA) and inversion before centrifugation at 500 g, 5 min, 4 °C. The supernatant was discarded, cells were diluted in 1x Diluted Nuclei buffer (10x Genomics), counted using Trypan Blue and a Countess III FL Automated Cell Counter, and processed for Tn5-based tagmentation. After tagmentation, the cells were loaded onto a Chromium controller to generate single-cell Gel Bead-In-Emulsions (GEMs), followed by linear PCR as described in the protocol, using a C1000 Touch Thermal cycler with a 96-Deep Well Reaction Module (Bio-Rad). To enable the amplification of the TotalSeq-A hashtag antibody, 0.5 μl of 1 μM bridge oligo A (BOA) was added to the barcoding mix. Subsequent ATAC and cell-hashing library generation were conducted as described in the user manual and as also outlined online at <https://cite-seq.com/asapseq/>, respectively.

Sequencing and demultiplexing

All libraries were sequenced using the Illumina NovaSeq6000 sequencing platforms at the Genomic Core Facility of BIH/MDC (FacilityID = 1565, The CoreMarketplace: MDC & BIH Technology Platform Genomics). mtscATAC-seq libraries were sequenced with a 2 × 100 paired-end read configuration. Raw-sequencing data were demultiplexed using CellRanger-ATAC ‘mkfastq’.

scATAC-seq analyses

Raw fastq files/sequencing reads were aligned to the mtDNA blacklist modified hg38 reference genome using CellRanger-ATAC count (v2.1.0). Peak-cell matrices and fragments files were further pre-processed using the Seurat v.4 and Signac v.1.10.R packages. Doublet detection was performed using AMULET v.1.1 with default parameters (<https://github.com/UcarLab/AMULET>)⁷⁰. For quality control, cells meeting all the following criteria were retained for downstream analysis: >1,000 but <25,000 unique nuclear fragments and a TSS enrichment score >3. Normalization and dimensionality reduction were performed using latent semantic indexing (LSI). The cell hashtag tag abundances were quantified using ASAP_to_kite.py (https://github.com/caleblareau/asap_to_kite), followed by the kallisto, bustools, and kite pipelines. Cell hash tag counts were then normalized and scaled using the centered log-ratio (CLR) transformation and the ScaleData function implemented in Seurat. For PBMC, cell-type annotations and UMAP coordinates were transferred and mapped from the Azimuth CITE-seq reference dataset using public 10x Genomics Multiome (RNA- and ATAC-seq) PBMC data as a cross-modality bridge via the Seurat Dictionary Learning bridge integration (v4.0.3) method⁷¹. The differentially accessible gene (DAG) analysis of galactose-treated cells was performed on the gene score matrix calculated by the scATAC-seq analysis package ArchR⁷² (v1.0.3) using the using the build-in ‘getMarkerFeatures’ with the default Wilcoxon pairwise test method.

mtDNA variants calling

Mitochondrial DNA variants were identified from the ATAC-seq modality using mgatk (v0.7.0) (<https://github.com/caleblareau/mgatk>). Cells with mtDNA coverage <5 were excluded. mtDNA variant allele frequencies were determined as described in ‘variant calling’ (https://github.com/caleblareau/mtscATACpaper_reproducibility/), returning a SummarizedExperiment object, containing two assay matrices, ‘allele_frequency’ and ‘coverage’. High-quality mtDNA variants were identified by filtering variants with a strand concordance > 0.65, mean-variance ratio > 0.01 in the PBMC dataset, mean-variance ratio > 0.0031 in the HEK293T dataset, and n_conf_cells_detected ≥ 1, unless indicated otherwise. In addition, four mtDNA variants that are known to be sensitive to sequencing artifacts due to their homopolymer sequence context, including mt.301A>C, mt.310T>C, mt.316G>C, and mt.3302A>C, were excluded from downstream analysis⁵⁹. For the “relaxed” filtering, we additionally examined mtDNA

variants with a strand concordance between 0.45 and 0.65. Notably, these variants retained the expected replication-error mutational signature, as demonstrated in Supplementary Fig. 1f. However, all other analyses presented throughout the manuscript exclusively utilized variants identified via the stricter filtering criteria. The list of pathogenic mtDNA variants was obtained from the MITOMAP database (https://www.mitomap.org/foswiki/bin/view/MITOMAP/Confirmed_Mutations). Only variants annotated as “Confirmed (Cfrm)”, “Confirmed-Pathogenic (Cfrm [P])”, or “Confirmed-Likely Pathogenic (Cfrm [LP])” were retained. The resulting list of confirmed pathogenic variants used for annotation is provided in Supplementary Data 4.

Sequencing downsampling analysis

Position-sorted BAM files generated from the ‘cellranger-atac’ pipeline were processed with samtools (v1.22). Reads aligning to the mitochondrial chromosome (chrM) were extracted, and cell barcode lists obtained from hashtag-based demultiplexing were used to separate individual cell line groups (CTRL, KI36, and KIA2). To normalize sequencing depth across samples, BAM files were randomly downsampled using samtools view -s (seed = 42) to fixed read depths of 10 M, 20 M, 30 M, 40 M, 50 M, 100 M, 150 M, and 200 M reads. The resulting downsampled BAM files were subsequently processed with mgatk (v0.7.0) for mtDNA variant calling. To visualize depth-dependent changes in variant composition, we normalized the pseudobulk VAFs of each mtDNA variant to the sum heteroplasmy of all variants at each downsampled depth (relative heteroplasmy). For the ΔVAF density plots, we first filtered mtDNA variants detected in both the 50 M and 200 M KIA2 datasets, and computed per-cell, per-position ΔVAF as VAF_{200M} - VAF_{50M}, excluding variants with zero VAF at both depths in that cell. The resulting ΔVAF values were visualized using kernel density estimation to summarize the distribution of single-cell VAF changes between 50M and 200 M total mtDNA reads.

Single-cell mitochondrial DNA mutational burden (scMMB)

We adapted metrics previously used to assess mtDNA mutational burden [6] and local constraint^{22,41,42}. These were modified into single-cell mtDNA MPM (scmtMPM) and single-cell Heteroplasmy-Weighted MSS (scwMSS). The mtDNA mutational burden in single cells was calculated at the levels of individual genes, respiratory chain complexes (I, III, IV, and V), and the whole mtDNA genome, as previously conducted in whole-exome sequencing data⁶. Mutation counts were calculated by first multiplying the ‘allele frequency’ and ‘coverage’ matrices, then summing across the respective mitochondrial genomic regions. The total genomic length sequenced was calculated by summing up the coverage matrix of the respective mitochondrial genomic regions of individual cells. Note that the scmtMPM score is normalized to mtDNA-aligned coverage (i.e., the total number of bases sequenced within the defined mtDNA region per cell). mtDNA mutation per million base pairs (MB) was calculated as the total somatic mtDNA mutation count divided by the total genomic length sequenced (in MB) for the respective genomic regions of a single cell. The weighted mitochondrial local constraint (MLC) score sum (MSS)²² was calculated by multiplying the position-wise MLC score by its single-cell heteroplasmy and summing the weighted scores for the respective genomic regions within a single cell.

Comparison of variant calling from mtDNA- and mtRNA-derived sequence reads

To assess the utility of popular RNA transcriptomic workflows for mtRNA variant calling, we quantified VAFs from both modalities from a 10x Multiome/DOGMA-seq library of PBMCs from the same healthy pediatric donor¹¹ using a cross-calling and filtering workflow (Supplementary Fig. 12a). First, variants were primarily called from mtDNA using mgatk to exclude artifactual variants that have previously been shown to be enriched in RNA-seq data. Second, variants identified from mtDNA

were cross-called in mtRNA reads. To account for uneven mtRNA transcript coverage, we further filtered the cell-by-position mtDNA/mtRNA coverage matrices and retained only positions with at least five reads per cell. We then filtered and re-quantified mutation counts and VAFs matrices with the filtered coverage matrices and compared the number of confidently detected variants in both mtDNA and mtRNA (Supplementary Fig. 7c). Modality weight was calculated as the number of detectable mtDNA variants divided by the sum of detectable variants in mtDNA and mtRNA (plus 0.01 for computational reasons).

Statistics and reproducibility

No statistical method was used to predetermine the sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessments. Two-sided Wilcoxon rank-sum tests were performed, and *P*-values were adjusted for multiple comparisons using the Holm–Bonferroni procedure, with significance defined at an FDR of 5%. All statistical analyses were conducted in R, and custom code used to replicate the analyses is available as part of the code availability.

Reporting summary

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

Data availability

The HEK293 cell line data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession code [GSE291877](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291877). The mtscATAC-seq dataset from a healthy pediatric donor (H05) and an additional MELAS patient (M80), along with the 10x Genomics multiomic dataset from the same healthy pediatric donor (H05), are available through the European Genome-phenome Archive (EGA) under accession number [EGAD50000001489](https://www.ebi.ac.uk/ega/egadataset/EGAD50000001489). Due to the sensitive nature of patient data, access is restricted and requires a formal application via the EGA website. Raw.fastq files for publicly accessible datasets were downloaded from the GEO and the Genotypes and Phenotypes database (dbGaP). The healthy adult dataset was downloaded from GEO under the accession number [GSE142745](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142745). Datasets from MELAS patients (M29, M35, M60) were available on dbGAP under the study accession number [phs002217.v1.p1](https://www.ncbi.nlm.nih.gov/beta/study/phs002217.v1.p1) [<https://dbgap.ncbi.nlm.nih.gov/beta/study/phs002217.v1.p1/#study>].

Code availability

Custom code to reproduce all analyses supporting this paper is available at <https://github.com/lsludwig-lab/scmtMMB>.

References

- Stewart, J. B. & Chinnery, P. F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* **16**, 530–542 (2015).
- Stewart, J. B. & Chinnery, P. F. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nat. Rev. Genet.* **22**, 106–118 (2021).
- Campbell, P. et al. Mitochondrial mutation, drift and selection during human development and ageing. *Res. Sq.* <https://doi.org/10.21203/rs.3.rs-3083262/v1> (2023).
- Sanchez-Contreras, M. et al. The multi-tissue landscape of somatic mtDNA mutations indicates tissue-specific accumulation and removal in aging. *Elife* **12**, e83395 (2023).
- Kennedy, S. R., Salk, J. J., Schmitt, M. W. & Loeb, L. A. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet.* **9**, e1003794 (2013).
- Gorelick, A. N. et al. Respiratory complex and tissue lineage drive recurrent mutations in tumour mtDNA. *Nat. Metab.* **3**, 558–570 (2021).
- Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* **52**, 342–352 (2020).
- Smith, A. L. et al. Age-associated mitochondrial DNA mutations cause metabolic remodelling that contributes to accelerated intestinal tumorigenesis. *Nat. Cancer* **1**, 976–989 (2020).
- Gorman, G. S. et al. Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, 16080 (2016).
- Walker, M. A. et al. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N. Engl. J. Med.* **383**, 1556–1563 (2020).
- Lareau, C. A. et al. Single-cell multi-omics of mitochondrial DNA disorders reveals dynamics of purifying selection across human immune cells. *Nat. Genet.* **55**, 1198–1209 (2023).
- Zhang, J. et al. Antigen receptor stimulation induces purifying selection against pathogenic mitochondrial tRNA mutations. *JCI Insight* **8**, e167656 (2023).
- Franklin, I. G. et al. T cell differentiation drives the negative selection of pathogenic mitochondrial DNA variants. *Life Sci. Alliance* **6**, e202302271 (2023).
- Trifunovic, A. et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417–423 (2004).
- Kujoth, G. C. et al. Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309**, 481–484 (2005).
- Rosignol, R. et al. Mitochondrial threshold effects. *Biochem. J.* **370**, 751–762 (2003).
- Boscenco, S. et al. Functionally dominant hotspot mutations of mitochondrial ribosomal RNA genes in cancer. *Nat. Genet.* **57**, 2705–2714 (2025).
- Nitsch, L., Lareau, C. A. & Ludwig, L. S. Mitochondrial genetics through the lens of single-cell multi-omics. *Nat. Genet.* **56**, 1355–1365 (2024).
- Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2020).
- Kim, M. et al. Single-cell mtDNA dynamics in tumors are driven by the coregulation of nuclear and mitochondrial genomes. *Nat. Genet.* **56**, 889–899 (2024).
- Lareau, C. A. et al. Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility. *Nat. Protoc.* **18**, 1416–1440 (2023).
- Lake, N. J. et al. Quantifying constraint in the human mitochondrial genome. *Nature* **635**, 390–397 (2024).
- Peeva, V. et al. Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nat. Commun.* **9**, 1727 (2018).
- Nissanka, N., Bacman, S. R., Plastini, M. J. & Moraes, C. T. The mitochondrial DNA polymerase gamma degrades linear DNA fragments, precluding the formation of deletions. *Nat. Commun.* **9**, 1–9 (2018).
- Hedberg-Oldfors, C. et al. Deep sequencing of mitochondrial DNA and characterization of a novel POLG mutation in a patient with arPEO. *Neurol. Genet.* **6**, e391 (2020).
- Filigrana, R. et al. Modulation of mtDNA copy number ameliorates the pathological consequences of a heteroplasmic mtDNA mutation in the mouse. *Sci. Adv.* **5**, eaav9824 (2019).
- Maclaime, K. D., Stebbings, K. A., Llano, D. A. & Havird, J. C. The mtDNA mutation spectrum in the PolG mutator mouse reveals germline and somatic selection. *BMC Genom. Data* **22**, 52 (2021).
- Ni, T. et al. MitoRCA-seq reveals unbalanced cytosine to thymine transition in Polg mutant mice. *Sci. Rep.* **5**, 1–13 (2015).
- Liu, M. H. et al. DNA mismatch and damage patterns revealed by single-molecule sequencing. *Nature* **630**, 752–761 (2024).

30. Romani, P. et al. Mitochondrial fission links ECM mechanotransduction to metabolic redox homeostasis and metastatic chemotherapy resistance. *Nat. Cell Biol.* **24**, 168–180 (2022).
31. Cassandri, M. et al. Zinc-finger proteins in health and disease. *Cell Death Discov.* **3**, 17071 (2017).
32. Kornblum, C. et al. Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nat. Genet.* **45**, 214–219 (2013).
33. Matic, S. et al. Mice lacking the mitochondrial exonuclease MGME1 accumulate mtDNA deletions without developing progeria. *Nat. Commun.* **9**, 1202 (2018).
34. Bahat, A. et al. Ribonucleotide incorporation into mitochondrial DNA drives inflammation. *Nature* **647**, 726–734 (2025).
35. Vercellino, I. & Sazanov, L. A. The assembly, regulation and function of the mitochondrial respiratory chain. *Nat. Rev. Mol. Cell Biol.* **23**, 141–161 (2021).
36. Kotrys, A. V. et al. Single-cell analysis reveals context-dependent, cell-level selection of mtDNA. *Nature* **629**, 458–466 (2024).
37. Lott, M. T. et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr. Protoc. Bioinforma.* **44**, 23.1–1.23.26 (2013).
38. Muyas, F. et al. De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nat. Biotechnol.* **42**, 758–767 (2023).
39. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
40. Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
41. Hong, Y. S. et al. Deleterious heteroplasmic mitochondrial mutations are associated with an increased risk of overall and cancer-specific mortality. *Nat. Commun.* **14**, 6113 (2023).
42. Hong, Y. S. et al. Mitochondrial heteroplasmy improves risk prediction for myeloid neoplasms. *Nat. Commun.* **15**, 1–15 (2024).
43. Arroyo, J. D. et al. A genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.* **24**, 875–885 (2016).
44. Kugapreethan, R. et al. Complex II assembly drives metabolic adaptation to OXPHOS dysfunction. *Sci. Adv.* **11**, eadr6012 (2025).
45. Balsa, E. et al. Defective NADPH production in mitochondrial disease complex I causes inflammation and cell death. *Nat. Commun.* **11**, 2714 (2020).
46. Fu, Y. et al. Engineering mtDNA deletions by reconstituting end joining in human mitochondria. *Cell* **188**, 2778–2793.e21 (2025).
47. Giordano, C. et al. Efficient mitochondrial biogenesis drives incomplete penetrance in Leber’s hereditary optic neuropathy. *Brain* **137**, 335–353 (2014).
48. Zhong, Y. et al. Mitochondrial transcription factor A overexpression and base excision repair deficiency in the inner ear of rats with D-galactose-induced aging. *FEBS J.* **278**, 2500–2510 (2011).
49. Gross, G. et al. Improved CAR-T cell activity associated with increased mitochondrial function primed by galactose. *Leukemia* **38**, 1534–1540 (2024).
50. Hoffmann, R. F. et al. Mitochondrial dysfunction increases pro-inflammatory cytokine production and impairs repair and corticosteroid responsiveness in lung epithelium. *Sci. Rep.* **9**, 15047 (2019).
51. Zhu, Y. et al. ARHGEF2/EDN1 pathway participates in ER stress-related drug resistance of hepatocellular carcinoma by promoting angiogenesis and malignant proliferation. *Cell Death Dis.* **13**, 652 (2022).
52. Koh, H.-J. et al. Tribbles 3 mediates endoplasmic reticulum stress-induced insulin resistance in skeletal muscle. *Nat. Commun.* **4**, 1871 (2013).
53. Chen, J. et al. Skeletal muscle mitochondrial dysfunction is associated with increased Gdf15 expression and circulating GDF15 levels in aged mice. *Sci. Rep.* **15**, 8101 (2025).
54. Lehtonen, J. M. et al. Diagnostic value of serum biomarkers FGF21 and GDF15 compared to the muscle sample in mitochondrial disease. *J. Inher. Metab. Dis.* **44**, 469–480 (2021).
55. Suomalainen, A. et al. FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: a diagnostic study. *Lancet Neurol.* **10**, 806–818 (2011).
56. Schaefer, P. M. et al. Combination of common mtDNA variants results in mitochondrial dysfunction and connective tissue dysregulation. *Proc. Natl. Acad. Sci.* **119**, e2212417119 (2022).
57. An, J. et al. Mitochondrial DNA mosaicism in normal human somatic cells. *Nat. Genet.* **56**, 1665–1677 (2024).
58. Smullen, M. et al. Modeling of mitochondrial genetic polymorphisms reveals induction of heteroplasmy by pleiotropic disease locus 10398A > G. *Sci. Rep.* **13**, 1–15 (2023).
59. Laricchia, K. M. et al. Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome Res* **32**, 569–582 (2022).
60. Bolze, A. et al. A catalog of homoplasmic and heteroplasmic mitochondrial DNA variants in humans. Preprinted at *bioRxiv* <https://www.biorxiv.org/content/10.1101/798264v3> (2020).
61. Lawson, A. R. J. et al. Somatic mutation and selection at the population scale. *Nature* **647**, 411–420 (2025).
62. Mahmood, M. et al. Mitochondrial DNA mutations drive aerobic glycolysis to enhance checkpoint blockade response in melanoma. *Nat. Cancer* **5**, 659–672 (2024).
63. Yi, Z. et al. Strand-selective base editing of human mitochondrial DNA using mitoBEs. *Nat. Biotechnol.* **42**, 498–509 (2023).
64. Cho, S.-I. et al. Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell* **185**, 1764–1776.e12 (2022).
65. Chen, L. et al. Efficient mitochondrial A-to-G base editors for the generation of mitochondrial disease models. *Nat. Biotechnol.* **1**, 12 (2025).
66. Ludwig, L. S. et al. Lineage tracing in humans is enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).
67. Mimitou, E. P. et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021).
68. Wei, W. et al. Germline selection shapes human mitochondrial DNA diversity. *Science* **364**, eaau6520 (2019).
69. Debska-Vielhaber, G. et al. Impairment of mitochondrial oxidative phosphorylation in skin fibroblasts of SALS and FALS patients is rescued by in vitro treatment with ROS scavengers. *Exp. Neurol.* **339**, 113620 (2021).
70. Thibodeau, A. et al. AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* **22**, 1–19 (2021).
71. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2023).
72. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

Acknowledgments

We thank the MDC/BIH Genomics Platform, Berlin (FacilityID=1565, The CoreMarketplace: MDC & BIH Technology Platform Genomics) for technical support and sequencing. The computation has been performed on the HPC research cluster of the Berlin Institute of Health. We are grateful to the members of the Kunz, Lareau, and Ludwig labs for valuable discussions. This work was supported by NIH K99/R00 HG012579 (CAL), UM1 HG012076 (CAL, LSL), the Hector Fellow Academy (YHH, PK, LN, LSL), the Paul Ehrlich Foundation (LSL), the EMBO Young Investigator Program (LSL), and grants by the Deutsche Forschungsgemeinschaft (DFG, LU 2336/3-1, LU 2336/6-1, STA 1586/5-1, TRR241, SFB1588, Heinz Maier-Leibnitz Award to LSL; KU 911/22-1 to

WSK; ZS 99/4-1 to GZ.). L.S.L. was further supported by the DFG under Germany's Excellence Strategy – EXC 3118/1 – project number 533770413; and by the DFG Emmy Noether (LU 2336/2-1) and Heisenberg Program (LU 2336/10-1).

Author contributions

Y.H.H. and L.S.L. conceived and designed the project. Y.H.H. led, designed, and performed experiments with J.L., V.D., and F.J. YHH led all analyses with input from P.K., A.G., L.N., S.C.C., M.S., C.A.L., and L.S.L. G.T., G.Z., and W.S.K. contributed to HEK293-*Polg*^{D274A} cell lines and biological perspectives. P.K. and L.N. contributed to healthy sample data generation. C.A.L. and L.S.L. supervised various aspects of this work. YHH and LSL wrote the paper with input from all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The Broad Institute has filed for patents relating to the use of technologies described in this paper, where CAL and LSL are named inventors (US patent applications 17/251,451 and 17/928,696). CAL is a consultant to Cartography Biosciences. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-70399-y>.

Correspondence and requests for materials should be addressed to Leif S. Ludwig.

Peer review information *Nature Communications* thanks Wei Wei and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026