# Single-cell multi-omics resolved analysis of mitochondrial genome-wide mutational burden, constraint, and mosaicism

**Leif Ludwig**

leif.ludwig@mdc-berlin.de

Berlin Institute of Health / Berlin Institute for Medical Systems Biology https://orcid.org/0000-0002-2916-2164

**Yu-Hsin Hsieh**

Berlin Institute of Health / Berlin Institute of Medical Systems Biology

**Pauline Kautz**

Berlin Institute of Health at Charité – Universitätsmedizin Berlin

**Lena Nitsch**

Berlin Institute of Health / Berlin Institute of Medical Systems Biology

**Ambre Giguelay**

Berlin Institute of Health at Charité - Universitätsmedizin Berlin https://orcid.org/0000-0003-0417-1109

**Janet Liebold**

Berlin Institute of Health at Charité – Universitätsmedizin Berlin

**Gábor Zsurka**

University of Bonn https://orcid.org/0000-0002-6379-849X

**Genevieve Trombly**

University Bonn Medical Center

**Markus Schuelke**

Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin; German Center for Child and Adolescent Health (DZKJ https://orcid.org/0000-0003-2824-3891

**Wolfram Kunz**

https://orcid.org/0000-0003-1113-3493

**Caleb Lareau**

Memorial Sloan Kettering Cancer Center https://orcid.org/0000-0003-4179-4807

---

**Technical Report**

**Keywords:**

**Posted Date:** May 21st, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6539733/v1



**Additional Declarations: Yes** there is potential Competing Interest. 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.

# Single-cell multi-omics resolved analysis of mitochondrial genome-wide mutational burden, constraint, and mosaicism

Yu-Hsin Hsieh[1,2,3], Pauline Kautz[2,3,4], Lena Nitsch[2,3,5], Ambre M. Giguelay[2,3], Janet Liebold[2,3], Gabor Zsurka[6,7], Genevieve Trombly[6], Markus Schülke[8,9,10], Wolfram S. Kunz[6,7], Caleb Lareau[11], Leif S. Ludwig[2,3,+]

1. Charité Universitätsmedizin Berlin, Germany
2. Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany
3. Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany
4. Technische Universität Berlin, Institute of Biotechnology, Berlin, Germany
5. Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany
6. Institute of Experimental Epileptology and Cognition Research, Medical Faculty, University of Bonn, Bonn, Germany
7. Department of Epileptology, University Hospital Bonn, Bonn, Germany
8. Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health (BIH), NeuroCureCluster of Excellence, Berlin, Germany
9. Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health (BIH), Department of Neuropediatrics, Berlin, Germany
10. Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health (BIH), NeuroCureClinical Research Center, Berlin, Germany
11. Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[+]To whom correspondence should be addressed: leif.ludwig@bih-charite.de

**Abstract**

Mitochondria contain their own circular, multi-copy genome, encoding essential components of the mitochondrial respiratory chain vital for cellular metabolism. Mitochondrial DNA (mtDNA) mutations occur more frequently than nuclear mutations and are associated with various diseases. While single-cell sequencing has allowed analysis of mtDNA variant heteroplasmy, a holistic view of mtDNA mutational landscapes in individual cells has remained limited, prohibiting the investigation of fundamental questions in mitochondrial genetics. Here, we leverage mitochondrial single-cell ATAC-seq (mtscATAC-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 are characterized by complex mutational landscapes, with pathogenic mutations and truncating variants only present at subthreshold levels, indicative of their negative selection. In human healthy donors and mitochondriopathy 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 novel framework to investigate fundamental properties of mitochondrial genetics, disease, and somatic mosaicism in human tissues.

## Introduction

Mitochondria are a central hub of cellular metabolism and contain a multi-copy genome with hundreds to thousands of copies per cell[1]. In humans, the 16.6 kb circular double-stranded mitochondrial DNA (mtDNA) encodes 13 protein subunits contributing to four of the 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 considerably more frequently than in their nuclear counterpart, with an up to >200× fold elevated rate in human blood[2]. While bulk-sequencing efforts have demonstrated the age-related increase of somatic mtDNA variants[3–6], established multiple associations surrounding cancers[7–9] and other human phenotypes[10,11], the unique features of mitochondrial genetics have made it challenging to establish genotype-phenotype relationships at the cellular level. These features include the highly variable variant allele frequencies (VAFs) of mtDNA variants known as heteroplasmy (the presence of two or more mtDNA variants per cell), variable mtDNA copy number as a function of cell type and state, relaxed (cell-cycle independent) replication, vegetative segregation, and natural turnover of mitochondria, which collectively make cellular mtDNA VAFs significantly more dynamic than for nuclear DNA variants[12].

The pathogenicity of mtDNA variants is best established in the context of a group of maternally inherited mitochondrial disorders, collectively known as mitochondriopathies[12,13]. While typically multi-systemic, given the ubiquity of mitochondria, mitochondriopathies are clinically heterogeneous and may present with particularly organ-specific phenotypes, often affecting the central nervous and muscular systems. However, cell-type-specific effects may extend to defects in red blood cell production and selection against pathogenic mtDNA, which has recently been described across primary human T cell populations[14–17]. Defects in the mitochondrial DNA polymerase of mice notably lead to an mtDNA 'mutator' phenotype with increased levels of mtDNA variants that result in the premature onset of aging-related phenotypes[18,19]. However, fundamental challenges remain in our understanding of the functional impact of mtDNA genetic variants. While individual pathogenic single-nucleotide variants (SNVs) or large-scale deletions are readily definable, the additivity of constantly arising somatic mtDNA variants across the breadth of the mitochondrial genome is more challenging to quantify and account for. Depending on the cellular context, the specific gene affected, and their VAFs, these variants may exert variable phenotypic effects. Generally, it is believed that cells can tolerate 60–80% of mutated mitochondrial genomes before reaching the biochemical threshold at which metabolic defects become detectable[12,20,21]; however, recent genetic evidence suggests that phenotypic effects may already exert themselves at lower VAFs[15].

Recent advancements in single-cell multi-omic sequencing have enabled new avenues in massively parallel whole mitochondrial genome sequencing, along with measures of cellular states, and have demonstrated an unprecedented level of variation in mtDNA variants[22] and copy number[23] across thousands of cells. For example, mitochondrial single-cell ATAC-seq (mtscATAC-seq) leverages the 10x Genomics platform and enables the concomitant sequencing of mitochondrial genomes and chromatin accessibility in a massively parallel manner[22,24]. Cells are thereby fixed to retain mitochondria and their genomes within,

which are subsequently tagmented by the Tn5 transposase, after which mtDNA and accessible chromatin fragments are PCR-amplified following microfluidic droplet compartmentalization. Moreover, a recently introduced mitochondrial genome constraint model provides a novel means for interpreting genetic variation underlying human phenotypes[25] across genomic databases such as gnomAD. Here, we leverage these advances to characterize cell lines with proofreading-deficient mitochondrial DNA polymerases and benchmark the sensitivity of these approaches to detect and quantify mitochondrial genome-wide genetic variation. We demonstrate an even more substantial mutational increase than previously appreciated via bulk sequencing approaches and introduce a computational framework to quantify and interpret genome-wide mtDNA mutational burden. We apply this framework to peripheral blood mononuclear cells (PBMCs) from healthy human donors and mitochondriopathy patients to aid the identification of pathogenic mtDNA variants in a cell-type-specific manner. Our widely applicable framework opens new avenues for the single-cell resolved study of mitochondrial genetics and somatic mosaicism in primary human cells and tissues.

## Results

### Single-cell mtDNA genotyping reveals an elevated number of single-nucleotide variants in *POLG*$^{D274A}$ cells

To assess the sensitivity of mtscATAC-seq to quantify increases in relative mitochondrial genome-wide mutation 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[26–28]. 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 (**Methods**). Both KI36 and KIA2 lines showed metabolic impairment compared to CTRL, in particular with respect to complex I and IV activities (**Extended Data Fig. 1a**). The control and *POLG*$^{D274A}$ cell lines were jointly profiled using mtscATAC-seq with cell hashing using oligonucleotide-conjugated antibodies to distinguish individual lines and detect mtDNA variants alongside accessible chromatin profiles (**Fig. 1a** and **Extended Data Fig. 1b**; **Methods**). The mtscATAC-seq quality control metrics were comparable among all three cell lines, which clustered distinctly, indicative of heterogeneous chromatin accessibility features (**Fig. 1b,c** and **Extended Data Fig. 1c**). Of note, the median mtDNA depth was 3-4-fold higher in both *POLG*$^{D274A}$ knock-in lines (**Fig. 1c** and **Extended Data Fig. 1c**; median mtDNA depth CTRL: 10.8, KI36: 27.8, and KIA2: 43.9). Given that KIA2 cells also exhibited higher nuclear read counts (median ATAC-seq read counts CTRL: 3917.5, KI36: 3748, and KIA2: 7959) that correlated with mtDNA depth across all three cell lines (Spearman correlation > 0.5), mtDNA depth was normalized to the nuclear ATAC read count (**Fig. 1c,d** and **Extended Data Fig. 1c,d**). However, the normalized mtDNA depth of both *POLG*$^{D274A}$ knock-in lines remained significantly higher than the control, suggesting a higher absolute mtDNA content, likely reflecting a compensatory response in mtDNA copy number to an elevated mutational load[28–30]. Notably, in *POLG*$^{D274A}$ cells, mtDNA depth across the mitochondrial genome appeared lower in the minor arc, the smaller segment of the genome between the replication origin of the heavy strand (oriH) and the replication origin of the light strand (oriL). This uneven

distribution is consistent with the previous finding that POLG 3′-5′ exonuclease deficiency leads to the accumulation of linear mtDNA fragments in the major arc (**Fig. 1e**)[26,27].

To detect mtDNA variants genome-wide, we used the single-cell mtDNA variant-calling pipeline, *mgatk*[22], and identified 455 variants across control cells using default parameters (variance mean ratio (VMR) >0.01, strand concordance >0.65). In the derived KI36 and KIA2 *POLG*$^{D274A}$ knock-in cells, a 5-10x-fold higher number of total mtDNA variants was identified, 2484 and 4651, respectively (**Fig. 1f**). While this is in concordance with the expected higher number of mtDNA mutations following disruption of *POLG* proofreading deficiency, the overall magnitude is more pronounced than previously described, where only hundreds of somatic variants were identified by bulk-sequencing-based approaches[18,31,32]. As expected, control and *POLG*$^{D274A}$ cells displayed a C>T transition-dominated mutational signature, indicating replication error as the primary driver underlying mtDNA mutations (**Fig. 1g**)[3,31,33,34]. Notably, relaxing the default thresholds in the mgatk variant calling resulted in a substantial increase in variants in *POLG*$^{D274A}$ compared to control cells (**Fig. 1f**, blue dotted box; **Methods**), which down to a particular strand concordance still showed the expected mutational signature (**Extended Data Fig. 1e**). Presuming similar background levels in control and knock-in cells, these findings suggest an even higher number of *bona fide* somatic mtDNA variants in *POLG*$^{D274A}$ cells. Together, these observations demonstrate mtscATAC-seq to be a sensitive tool to detect elevations in mtDNA mutational rates and the resulting increase in mtDNA variants across the mitochondrial genome.

**Heteroplasmy analysis of *POLG*$^{D274A}$ cells reveals a variant allele frequency-dependent increase of mtDNA variants and their dynamics**

Given the multi-copy nature of the mitochondrial genome, mtDNA variants may be present in only a fraction of the copies of a cell or population, a state known as heteroplasmy. As mtscATAC-seq yields a high mtDNA depth per cell (up to 30-40x in *POLG*$^{D274A}$ cells), we determined the number of variants with a VAF that exceeds a specified threshold. First, we detected a median of 9 mtDNA variants per cell in the control and 180 and 472 in the two *POLG*$^{D274A}$ cell lines regardless of their VAFs (**Fig. 2a**). The total number of detected mtDNA variants per cell primarily correlates with mtDNA sequencing depth, particularly at lower VAF levels (**Extended Data Fig. 2**). However, the increases started to plateau for variants >3% VAF and higher mtDNA depth and became mostly independent of mtDNA depth for variants >10% VAF (**Extended Data Fig. 2**). This observation is in line with the expectation that the total detected number of low VAF variants will be a function of recovered mtDNA depth. In contrast, for higher VAF variants and at a specified mtDNA depth, their confident detection does not further benefit from deeper mtDNA coverage. For further analysis, we opted to normalize the total number of mtDNA variants per cell by its mtDNA sequencing depth (**Fig. 2b; Methods**), after which the two *POLG*$^{D274A}$ cell lines still showed an increase in the depth normalized number of mtDNA variants per cell compared to the control line (median: CTRL: 0.769, KI36:6.38, and KIA2: 10.9).

To further refine the mutational analysis, we only considered 'shared variants' between the control and *POLG*$^{D274A}$ lines, revealing 388 mtDNA variants. Notably, KIA2 showed a higher number of shared variants with the parental CTRL line (n=227), compared to KI36 (n=52), likely indicative of their direct clonal relationship (**Fig. 2c**). In contrast, variants that were unique to KIA2 and KI36 cells were classified as likely '*de novo*' mtDNA variants. Filtering *de novo* variants by their VAFs notably showed a distinct distribution, with the number of, in particular, low-frequency mtDNA variants showing a 50- to 100-fold increase in *POLG*$^{D274A}$ compared to control cells (**Fig. 2d,e**). This aligns with the notion that constantly newly arising mtDNA variants will initially be present in a single mtDNA copy, thus at low VAFs, with relaxed replication and genetic drift of mtDNA variants eventually leading to their increased prevalence across mtDNA copies over time. As such, the total number of mtDNA variants per cell progressively decreased at higher VAF thresholds (**Fig. 2d,e**). Along these lines, 'shared' variants displayed higher pseudobulk VAF levels than likely '*de novo*' variants (**Fig. 2f,g**). These observations align with anticipated heteroplasmy dynamics of somatic events, with higher VAF levels requiring further genetic drift and vegetative segregation[35]. Together, these analyses demonstrate that mtscATAC-seq and mgatk can sensitively detect *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 MITOMAP[36] (**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 across cells (**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 that are typically associated with mitochondrial dysfunction of individual pathogenic variants (e.g., 60-80%)[20,21] (**Fig. 3b,c**). Moreover, out of the 130 reported pathogenic variants in MITOMAP, 12 and 18 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 (**Extended Data Fig. 3a**). In contrast, truncating variants were more frequently observed together in individual cells, including at higher but mostly <60% VAF, resulting in a greater number of double-positive cells (**Extended Data Fig. 3b**). 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 further do not exceed classical biochemical thresholds in individual cells, suggesting the cellular phenotype to be driven by the cumulative mutational burden as opposed to the phenotypic impact of individual or few deleterious variants.

**Quantitative analysis of single-cell genome-wide mtDNA mutational burden in *POLG*$^{D274A}$ cells**

To date, most single-cell resolved mtDNA genotyping approaches primarily focused on individual mtDNA variants compared to considering the complete variant call set per cell, which may further be highly variable across a population of cells[15,22]. 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 complexes of the respiratory chain (**Fig. 3** and **4a**). Motivated by the heterogeneous VAF distributions of mtDNA variants in *POLG*$^{D274A}$ lines, 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[37–39]. This metric has been adapted for assessing the mitochondrial genome mutational burden in bulk exome sequencing datasets[7]. 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). As mtDNA depth may also vary from cell to cell, scmtMPM provides a normalized assessment of the overall mutational frequency of the mitochondrial genome and its heterogeneity at the level of single cells.

In addition to scmtMPM, a recently introduced constraint model built on large-scale genomic data from the Genome Aggregation Database (gnomAD)[25,40] compares the observed genetic variation 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 across the mitochondrial genome. The MLC score ranges from 0 to 1, with a score of 1 marking the most constrained positions and, consequently, 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[25,40,41]. As individual cells may exhibit unique VAF distributions of mtDNA variants (**Fig. 4a**), we further take their VAFs into account, 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 evolution[7]. As expected, *POLG*$^{D274A}$ cells showed substantially elevated scmtMPM and scwMSS scores which were

independent of mtDNA sequencing depth (**Fig. 4c,d** and **Extended Data Fig. 4a-d;** median total scmtMPM, CTRL: 8.87, KI36: 1193, and KIA2: 1586; median total scwMSS, CTRL: 0.089, KI36: 9.37, and KIA2: 13.1). Notably, the KI36 derivative showed a bimodal distribution in respective scores (**Extended Data Fig. 4a;** red arrow), corresponding to distinct subclusters (**Fig. 4c;** C1-C3). 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 line cells exhibited an overall low mtDNA mutational burden across functional groups (**Extended Data Fig. 4b,c**). However, a subset of control cells showed a relatively high mutational burden in rRNA genes, indicative of this space potentially being more tolerant to mutations, given its noncoding characteristics (**Fig. 4c,d** and **Extended Data Fig. 4c**). Though this was less pronounced, a few control cells also showed elevated scores for complex I and IV (**Extended Data Fig. 4b**). In contrast, *POLG*$^{D274A}$ cells showed a narrower distribution of scmtMPM and scwMSS scores, which were generally higher in the KIA2 derivative (**Fig. 4b,c** and **Extended Data Fig. 4a**). Overall, scmtMPM and scwMSS scores were well correlated, though this correlation was more apparent in *POLG*$^{D274A}$ cells (**Extended Data Fig. 4e**), suggesting both measures to 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 (**Extended Data Fig. 5a,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 (**Extended Data Fig. 5a,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.

**Single-cell MMB analysis reveals heterogeneity and somatic mosaicism of human immune cells**

Maternally inherited mtSNVs and sporadic mtDNA mutations acquired during aging together shape the mutational landscape and have been implicated in human phenotypes, such as aging, metabolic disease, neuromuscular disease, cancers, and congenital mitochondrial diseases[8,10,12,13,42,43]. Moreover, the co-occurrence and mutual exclusiveness of mtSNVs suggest complex interactions between mtDNA variants[22,43]. 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 evaluate 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. 5a**). Across all individuals, scmtMPM and scwMSS scores were well correlated and independent of mtDNA depth (**Extended Data Fig. 6a,b**).

Additionally, an age-related increase in the number of detected mtDNA variants was observed across healthy donors and patients (**Extended Data Fig. 7a,b**), consistent with prior observations of mtDNA heteroplasmy quantifications in hematopoietic cells[2]. However, a more heterogeneous distribution of scmtMPM and scwMSS scores in primary cells compared to cell lines was notable, including in a cell-type- and donor-specific manner (**Fig. 5b,c** and **Extended Data Fig. 4**).

Specifically, in the 5-year-old healthy pediatric donor, groups of cells with a relatively higher mutational burden were observed across different immune cell types (**Fig. 5c**; top left, arrows), suggesting the non-uniform acquisition and distribution of somatic mutations 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. 5b**; red violin plots). As the pathogenic m.3243A>G mutation has been described to undergo purifying selection in select T cell subsets and the hematopoietic system throughout the lifespan of patients (**Extended Data Fig. 7c**)[14,15], we also reevaluated scMMB metrics after excluding the pathogenic variant. Notably, the removal of the m.3243A>G mutation abolished the bimodal distribution in the two younger patients (29 and 35 years old). However, this reduction was not observed in the older patients (60 and 80 years old), suggesting that while the m.3243A>G mutation may be gradually lost, new functionally relevant mutations emerge with age contributing to elevated scmtMPM and scwMSS scores (**Fig. 5b**). Additionally, a significant fraction of cells with no detectable m.3243A>G heteroplasmy, showed relatively higher scmtMPM scores (**Fig. 5d,e**; arrows) compared to cells even with low m.3243A>G VAFs, suggesting the presence of a highly pathogenic congenital mtDNA variant to exert selective pressures to accumulate additional somatic mtDNA mutations. Notably, reevaluation after 'removing' m.3243A>G showed that the scmtMPM metric remains low in cells with detectable levels of the variant, regardless of m.3243A>G VAFs and cell type (**Fig. 5e** and **Extended Data Fig. 7d**). Additionally, we did not observe differences in mtDNA depth across cells with varying m.3243A>G VAFs (**Fig. 5f**). 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.

**Analysis of elevated mtDNA mutational burden reveals potential pathogenic variants and lineage-specific implications in complex I**

To further investigate the nature of elevated scMMB scores, we evaluated whether these are attributable to a single highly constrained or the cumulative effect of multiple mtDNA variants. For example, a single highly constrained variant would overproportionally contribute to elevated scMMB scores. To assess this, we deliberately removed the variant with the highest VAF heteroplasmy, upon which scMMB scores significantly decreased, suggesting that a single somatic variant profoundly impacted scMMB scores (**Fig. 6a**; yellow violin plots). Next, we stratified scores based on gene and functional groups to further resolve the underlying heterogeneity of scmtMPM and scwMSS in primary cells with pathogenic mtDNA (**Fig. 6b**). As expected, MELAS-derived patient cells showed higher scores for tRNA genes, and cell-type-specific analysis further revealed tRNA scores and pathogenic m.3243A>G variant heteroplasmy to be highly correlated

(**Extended Data Fig. 7e,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. 6b**). Moreover, the contribution to scwMSS scores for complex III and V variants was generally lower (**Fig. 6b**; 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 (**Extended Data Fig. 7g,h**)[41]. These findings indicate distinct mutation dynamics, as also previously intimated across human cancers where complex I subunits were enriched for pathogenic mutations, whereas complex V mutations were broadly depleted for all non-synonymous mutations[7].

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 healthy 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. 6b,c**). To identify the underlying mtDNA variants, we ranked all detected variants in respective genes by their pseudo-bulk heteroplasmy. Notably, in each instance, only a single variant exceeded 1% VAF: m.10599G>A in *MT-ND4* in the H05 (mean VAF = 8%), m.10270T>C in *MT-ND3* in M60 (mean VAF = 25%) and m.10398A>G in *MT-ND3* in M80 (mean VAF = 5.8%), respectively (**Fig. 6c,d** and **Table S1**). The m.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. 6d**). Given the low likelihood of such a pronounced clonal expansion at the donor's young age, this suggests the m.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[2,44]. Moreover, we observed pronounced enrichments of cells with high m.10559G>A VAFs in both $CD4^+$ and $CD8^+$ memory T cells, indicating the mtDNA variant tracks with clonal expansion and memory formation (**Fig. 6e**). In contrast, the m.10270T>C variant in M60 exhibited a narrow and exclusive distribution marking 76% of $CD8^+$ T effector memory cells (**Fig. 6d**). In concordance with a very high MLC score of 0.969, SIFT and AlphaMissense predict m.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 level[25]. However, the strong enrichment of the m.10270T>C mutant allele in memory $CD8^+$ T cells suggests substantial clonal expansion and lineage-specific tolerance toward a pathogenic variant (**Fig. 6e**). This specific population also notably deviated from the regression line of the mtscMPM and scwMSS (**Extended Data Fig. 6b**; 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 m.10398A>G variant in M80 is present in various cell types, with 25% of the $CD8^+$ naive T cells being marked (**Fig. 6d**), 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 MitoMAP, such as Leigh syndrome, breast cancer, metabolic syndrome, and Parkinson's disease[45–48]. While this variant

appeared to have medium-low VAFs in the majority of the immune cell population, we observed enrichments of cells with high m.10398A>G VAFs in monocytes and $CD8^+$ T cell effector/memory populations (**Fig. 6e**), again suggesting a potential lineage-specific preference toward individual mtDNA mutations. In summary, the evaluation of cellular scMMB scores readily enables the identification of 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 in both healthy and disease contexts, such as mitochondrial gnomAD[49] and mitoMAP[36], has significantly broadened our understanding of the essentiality and local constraint across the genes of the mitochondrial genome[7,25,50]. However, the bulk nature of the assay likely vastly underestimates the frequency of pathogenic or functional mtDNA variants present only at low VAF levels at the population level, which may otherwise be impacting a small population of cells. Specifically, individual adult stem cells and their progeny may, for example, harbor their own heteroplasmic or even homoplasmic variants following the acquisition of a somatic event[22,51], which are not readily detectable at the bulk level. Recent advances in single-cell mtDNA genotyping have started to uncover the full spectrum of somatic mosaicism and revealed heterogeneous and cell-type-specific responses toward mtDNA mutations[14,15,22,35]. However, analytical frameworks for studying and interpreting mitochondrial genetic variation at the single-cell level are lacking, still limiting our understanding of mtDNA variant dynamics and somatic mosaicism. Here, we applied single-cell mtDNA genotyping using mtscATAC-seq to analyze the mtDNA mutational burden across individual cells. Inspired by recent studies in quantifying mtDNA mutational burden in cancer exome sequencing[7] using MPM and mtDNA constraint model-derived MLC/MSS[7,25,40,41], we adapted these metrics to introduce a general analytic framework to quantify the mitochondrial genome-wide mutational burden in single cells. Both scmtMPM and scwMSS are mtDNA depth-normalized and take the variation of VAFs of mutations in single cells into account, thus providing a more precise and standardized approach to studying the mtDNA mutation burden at the single-cell level, even at lower sequencing depth (**Extended Data Fig. 4** and **Extended Data Fig. 6**).

By leveraging exonuclease-deficient *POLG*$^{D274A}$ HEK293 cells, we benchmarked this analytic framework to quantitatively and qualitatively estimate the mutational burden across the mitochondrial genome (**Figs. 2-4**). We observed an elevated mutation rate and notable increases in mtDNA *de novo* variants in knock-in cell lines (KI36 and KIA2) compared to the parental control line. While these findings are not unexpected, the single-cell-based mtDNA sequencing suggests a substantially higher number of somatic mtDNA mutations as compared to bulk approaches and reveals insights into their heterogeneous distribution at unprecedented resolution. Notably, we observed pathogenic and truncating mutations only present at subthreshold variant levels compared to synonymous and missense variants, indicating their negative selection. Further, these findings clearly 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 classical mitochondriopathies (**Fig. 3**).

In the context of human blood and immune cells, we further demonstrate how scmtMPM and scwMSS provides a new avenue to resolve mitochondrial genetic heterogeneity, dynamics, and mosaicism in humans (**Figs. 5** and **6**). In patients with MELAS, we notably identified several likely functional mtDNA mutations that had previously eluded clinical diagnosis, highlighting the unbiased nature of this approach. Furthermore, single-cell sequencing analyses revealed lineage-specific mtDNA variant distributions, reminiscent of previously described selection and clonal expansion in primary human immune cells. Together, this work underscores the utility of single-cell mtDNA sequencing approaches to provide insights into mitochondrial genome stability, potential disease associations, and somatic mosaicism and dynamics.

While RNA-seq-based transcriptional profiling approaches remain the most popular single-cell sequencing technique, variant calling from mtRNA transcripts has been shown to be subject to RNA-editing and high false positive rates, potentially attributable to transcriptional polymerase error. Moreover, sequence coverage is particularly limited in 5' or 3' based sequencing techniques[1,51]. To quantitatively assess this, we analyzed 10x Genomics Multiome-based DOGMA-seq libraries that co-capture mtDNA variation from the ATAC-seq modality of the library alongside 3'-primed mRNA libraries from the same cell[15,52]. As expected, mtRNA-derived sequence reads exhibited a significantly more uneven and 3'-enriched coverage compared to mtDNA reads (**Extended Data Fig. 8a,b**). Accordingly, we observed a significant disparity in the degree of identified mitochondrial genetic variation between mtDNA and mtRNA, with more detectable mtDNA variants in 75% of the cells (**Extended Data Fig. 8c**). 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 rate[51], we warrant caution surrounding the sole transcriptomic-focused analysis of functionally relevant mitochondrial genetic variation.

Given the heteroplasmic nature of mtDNA variants, it is widely accepted that a non-synonymous/pathogenic mtDNA variant must reach a biochemical threshold (typically around 60–80% VAF) to cause metabolic defects in cells [20,21]. However, with the advent of deeper sequencing techniques, large-scale population studies have found that certain regions of mtDNA appear to lack even low-VAF variants, suggesting that the phenotypic threshold may be lower than previously expected[13,53]. In fact, we previously demonstrated purifying selection against mtDNA pathogenic variants and deletions in patients' peripheral blood and immune cells, where more than 50% of cells were depleted entirely of pathogenic mtDNA variants (**Extended Data Fig. 7c**; red arrow)[14,15]. In line with these findings, we further identified that the presence of a pathogenic mtDNA variant, even at low VAF, appears to impede the accumulation of additional mtDNA somatic mutations (**Fig. 5d,e**), supporting the idea that low-VAF mtDNA mutations can have biological impacts on the mtDNA mutational landscape. In line with these observations, a recent study has demonstrated that the co-occurrence of 'incompatible' common variants contributes to severe mitochondrial

disease[43], further supporting the notion of pronounced genetic interactions. Additionally, we showed that a single high-VAF mtDNA variant can dominate the mtDNA mutational landscape (**Fig. 6a**). Together, these observations underscore the complex interplay of mtDNA variants and highlight the importance of employing genomic methods to detect the subtle yet significant effects of mtDNA alterations on the mutational landscape, complementing traditional metabolic assessments of cellular function. Future work applying the presented experimental and analytic framework to longitudinally collected primary blood or tissue samples from mitochondrial disorders will deepen our understanding of the dynamics of mtDNA mutation accumulation and selection, as well as their contribution to disease progression and phenotype variability. Further developments of multimodal single-cell genomic assays with integrative transcriptomic and metabolic readout will be required to comprehensively evaluate mtDNA genotype-phenotype relationships and functionally assess the cellular consequences of an elevated mtDNA mutational burden, providing deeper insights into how mitochondrial genome somatic mosaicism impacts cellular phenotype and disease progression.

In conclusion, the studies of mitochondrial genetics have come with unique challenges, given the variation in mtDNA copy number and distinct modes of replication and segregation that have hampered the quantification of somatic mitochondrial mosaicism. Here, we use *POLG*$^{D274A}$ cells to benchmark mtscATAC-seq's capability to detect and quantify genome-wide mtDNA variants in single cells sensitively. We introduce a quantitative framework to assess mitochondrial genome-wide mutational burden and constraint. Their application to primary human immune cells, including patients with mitochondriopathies, showcased their utility in sensitively detecting pathogenic mtDNA variants and variations in their cellular distribution. Together, these approaches provide a robust and scalable method to systematically investigate fundamental aspects of mitochondrial genetics at the level of individual cells.

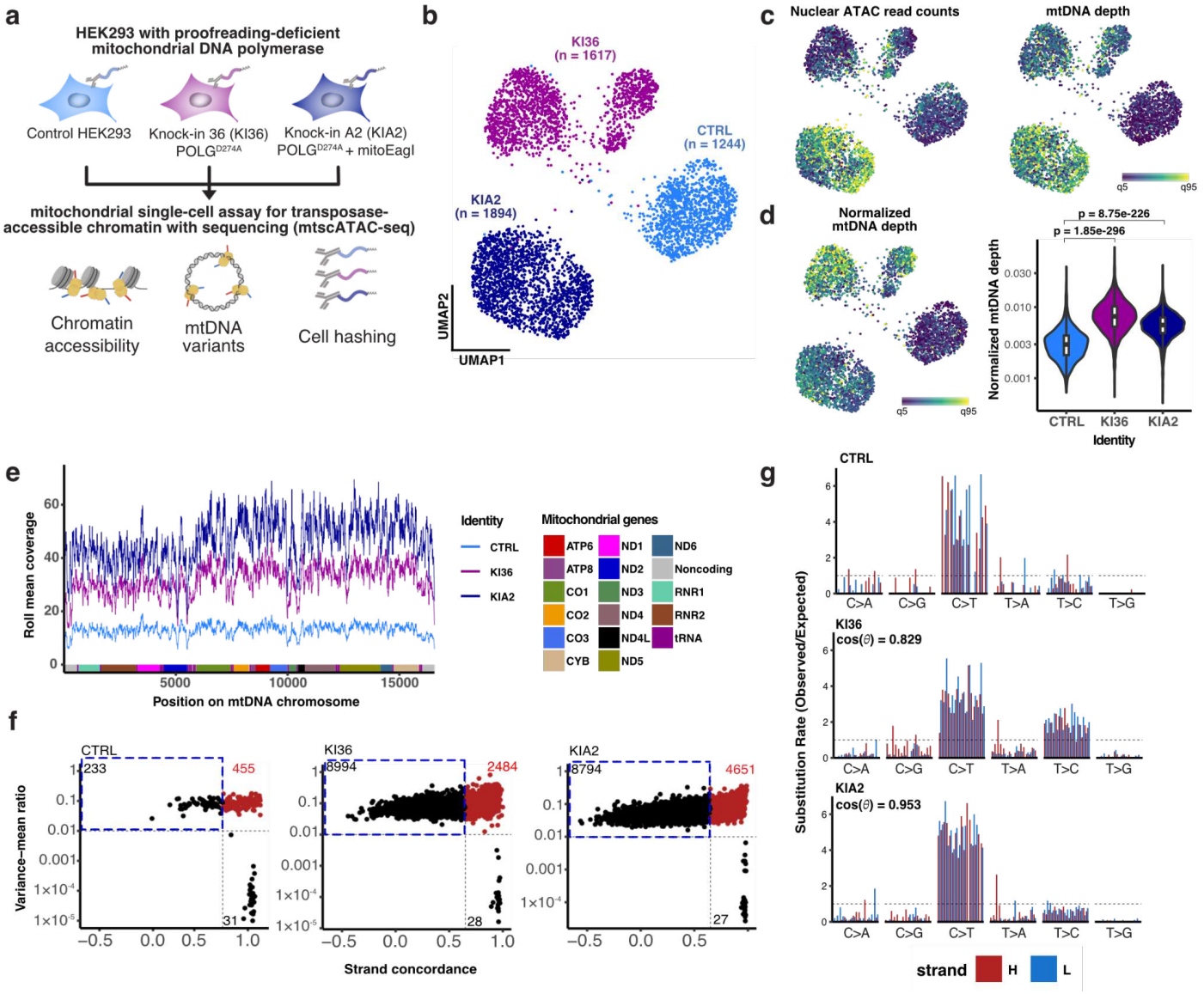



**Figure 1 – mtscATAC-seq-based profiling of single-cell mitochondrial DNA genotypes in *POLG*$^{D274A}$ HEK293 cell lines. (a)** Outline of the experimental design. **(b)** UMAP projection of hashtag-demultiplexed HEK293 cell lines. n = cell numbers **(c)** UMAP projection color-scaled by the number of per cell unique ATAC read counts (left) and mtDNA depth (right). **(d)** UMAP projection (left) and violin plot (right) of mtDNA depth normalized to the number of unique ATAC read counts. **(e)** Visualization of the mitochondrial genome position-wise coverage of the indicated cell lines. The rolling mean coverage at five nucleotide resolutions is shown. **(f)** 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: 233, KI36: 8994, and KIA2: 8794). **(g)** 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.

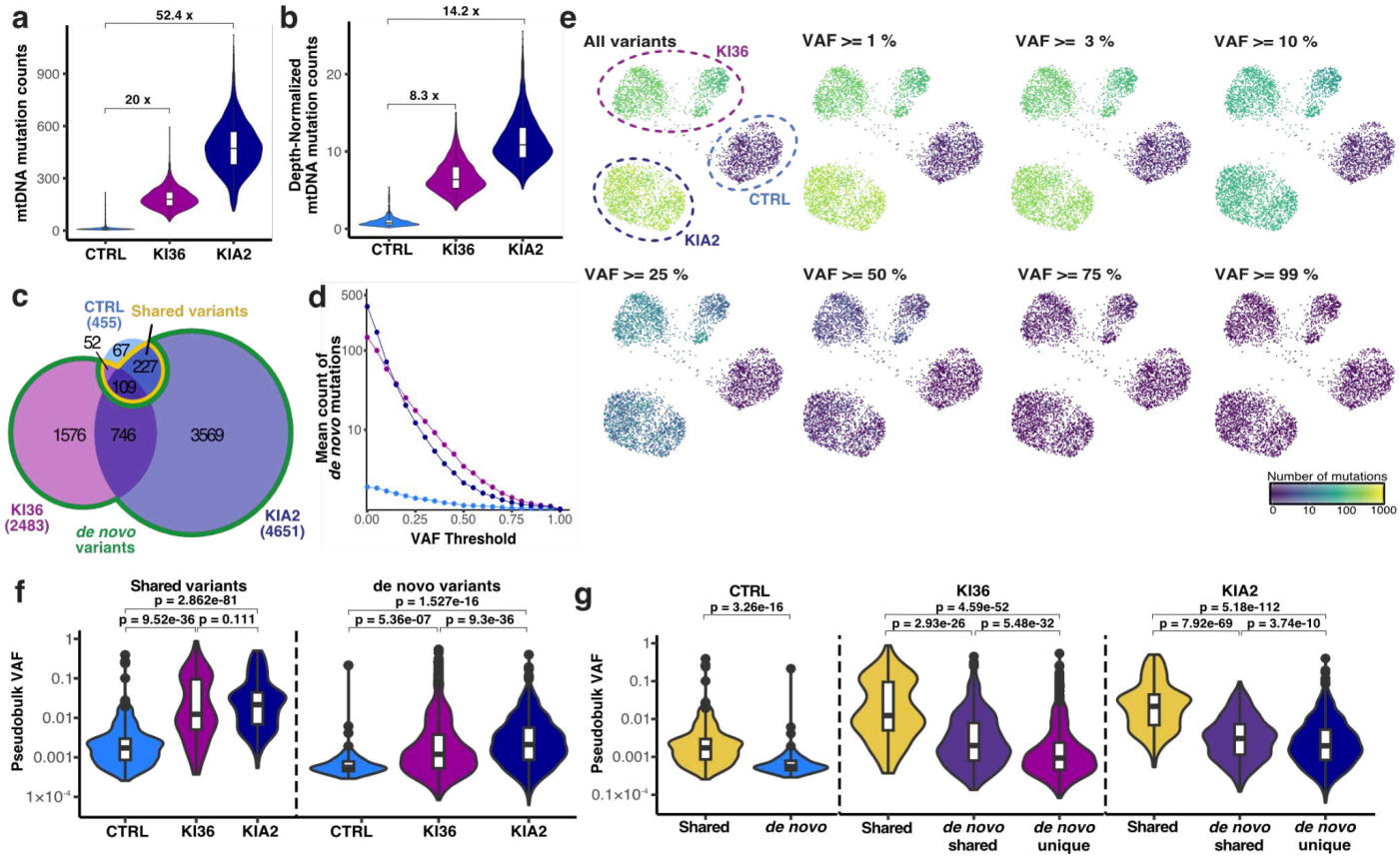



**Figure 2 – Heteroplasmy distribution and heterogeneity in *POLG*$^{D274A}$ cell lines. (a,b)** Violin plots of raw (a) and mtDNA-depth-normalized (b) mtDNA variant counts. Fold-changes of the median relative to the control cell line are shown. **(c)** Venn diagrams depicting shared and unique somatic mutations called by the mgatk pipeline. Parental variants, highlighted within the yellow circle, were defined as variants that were present in both *POLG*$^{D274A}$ and control cell lines. **(d)** Line graph showing the mean count of *de novo* somatic mtDNA variants (y-axis) at 5% incremental heteroplasmy thresholds. **(e)** UMAP projection of the color-scaled distribution of the number of *de novo* somatic mtDNA variants above indicated variant allele frequency thresholds. UMAP as in Fig. 1b. **(f)** Comparison of mean pseudobulk heteroplasmy between parental and *de novo* mtDNA variants stratified by cell lines. **(g)** Comparison of mean heteroplasmy across cell lines stratified by shared and *de novo* (unique or shared between two POLG$^{D274A}$ cell lines) mtDNA variants. Holm–Bonferroni adjusted p-values from the Wilcoxon test are shown, with significance at a type I error rate of 0.05.

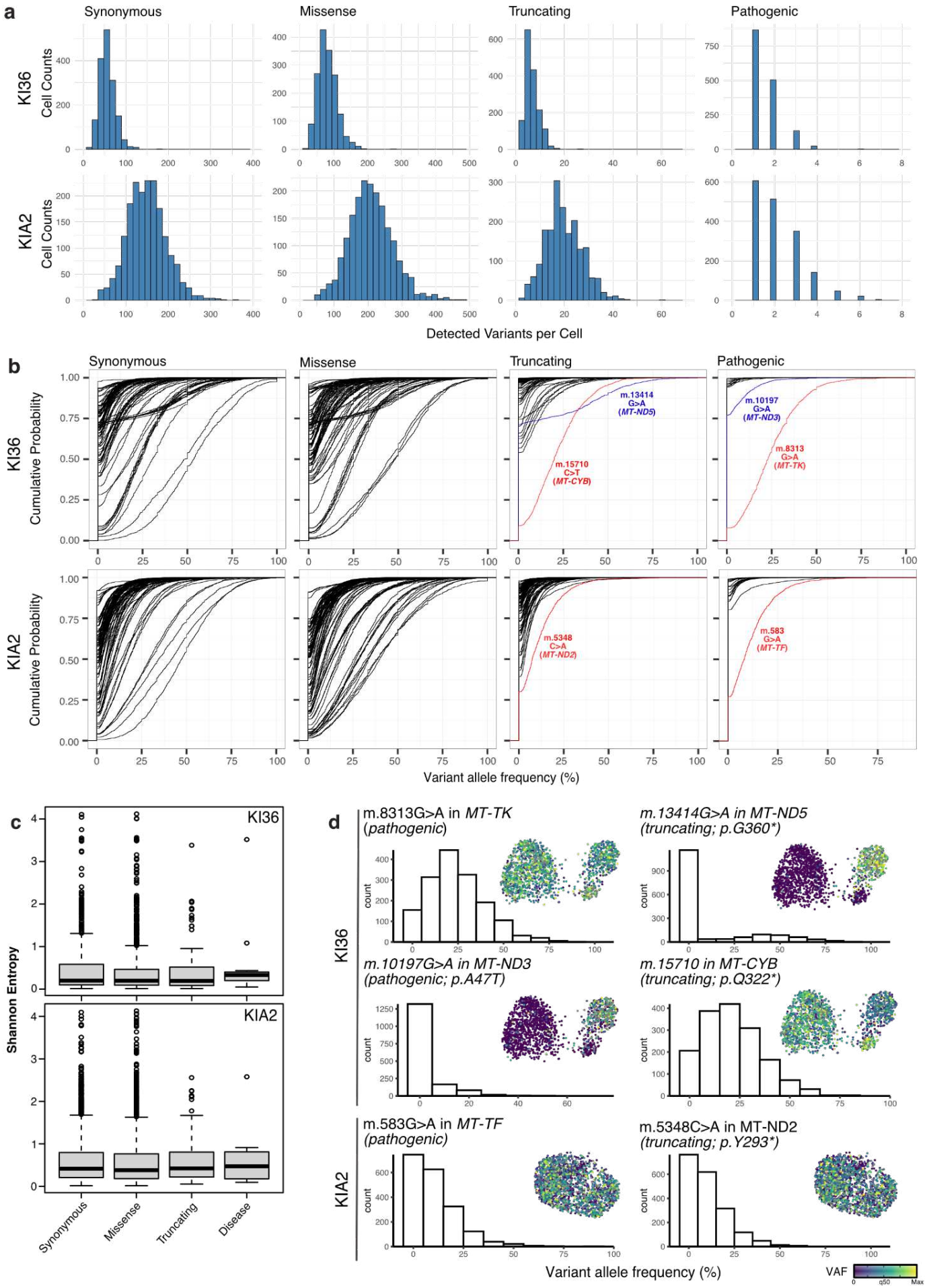

a
Synonymous
Missense
Truncating
Pathogenic
KI36
KIA2
Cell Counts
Detected Variants per Cell
b
Cumulative Probability
m.13414 G>A (MT-ND5)
m.15710 C>T (MT-CYB)
m.10197 G>A (MT-ND3)
m.8313 G>A (MT-TK)
m.5348 C>A (MT-ND2)
m.583 G>A (MT-TF)
Variant allele frequency (%)
c
Shannon Entropy
Synonymous
Missense
Truncating
Disease
d
m.8313G>A in MT-TK (pathogenic)
m.13414G>A in MT-ND5 (truncating; p.G360*)
m.10197G>A in MT-ND3 (pathogenic; p.A47T)
m.15710 in MT-CYB (truncating; p.Q322*)
m.583G>A in MT-TF (pathogenic)
m.5348C>A in MT-ND2 (truncating; p.Y293*)
count
Variant allele frequency (%)
VAF

**Figure 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 (upper) and KIA2 (lower) line, stratified by 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 indicated mutation type. **(d)** UMAPs and histograms showing the VAF distribution of mtDNA variants highlighted in (b).

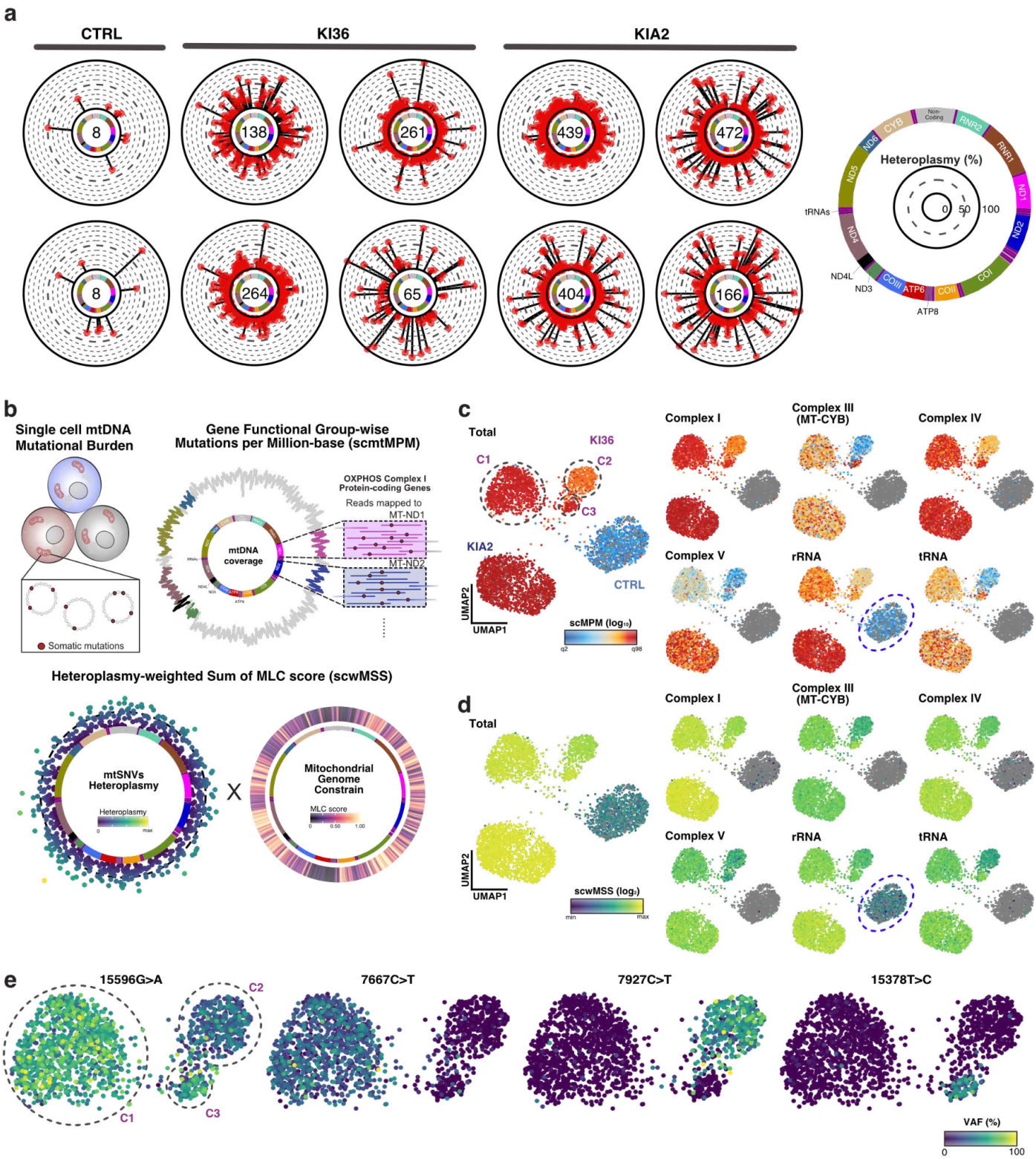



**Figure 4 – Quantification of single-cell mtDNA mutational burden via scmtMPM and scwMSS. (a)** Examples of mtDNA SNV heteroplasmy distribution in randomly selected single cells around the mitochondrial genome. The numbers of confidently detected variants are indicated in the middle. Each red dot represents 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; upper) and Heteroplasmy-weighted sum of mitochondrial local constraint score (scwMSS; lower). **(c,d)** UMAP projections as in Fig. 1b of scmtMPM (c) and scwMSS (d) for

indicated respiratory chain complexes, tRNA, and rRNA genes in *POLG*$^{D274A}$ HEK293 cells. The blue dotted circle highlights elevated rRNA scores in the control cell line. (**e)** UMAP projections of heteroplasmy of representative clonal somatic mtDNA variants within KI36 cell line-specific subclusters C1-C3.

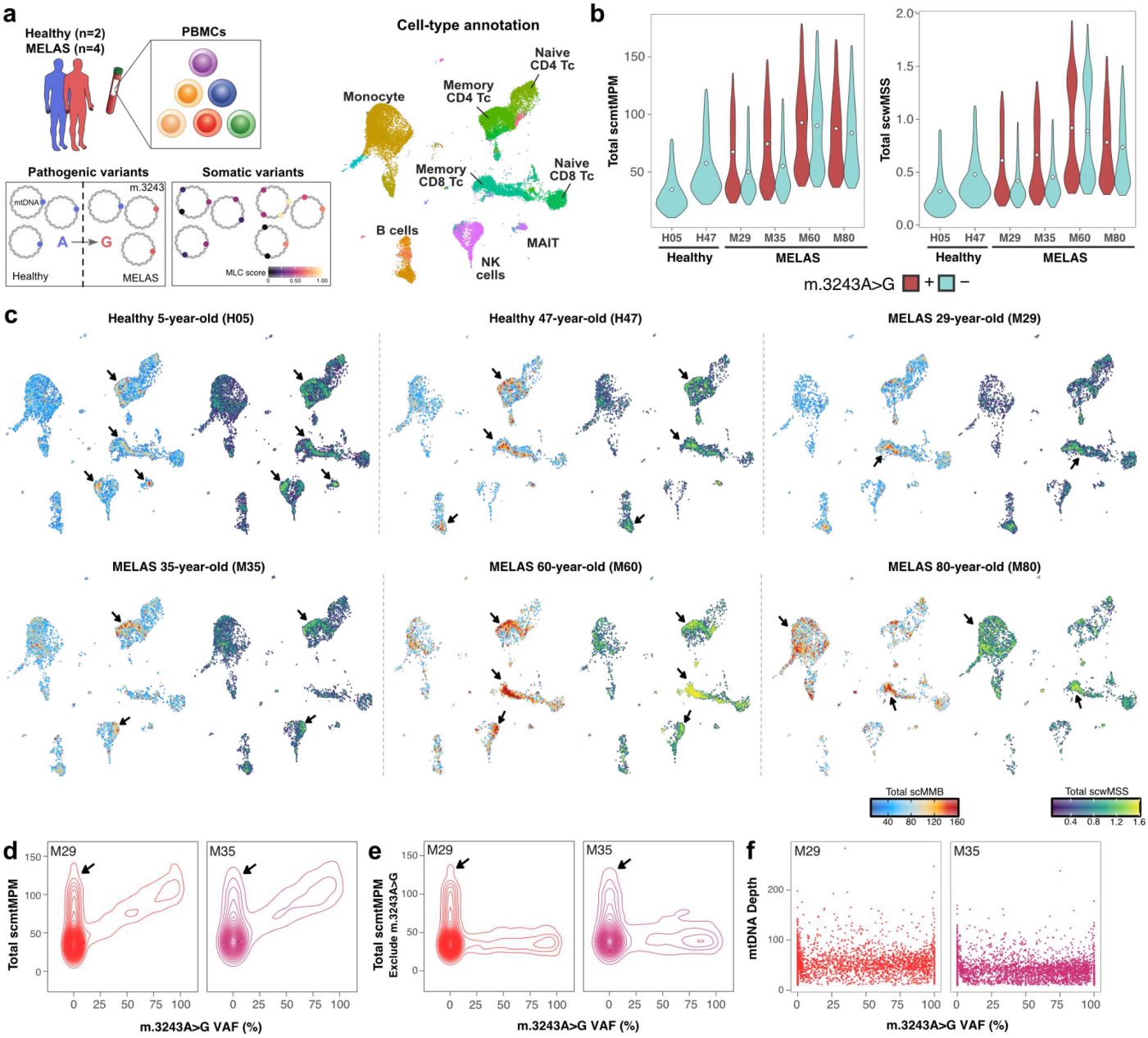



**Figure 5 – Quantification of mtDNA mutational burden in human peripheral mononuclear blood cells in healthy individuals and MELAS patients. (a)** Schematic representations of the available datasets and analysis design. **(b)** Violin plots of mitochondrial genome-wide mutation burden metrics +/- the pathogenic variant m.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 m.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 age of the donors. **(d)** Contour plots illustrate the relationship between the m.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 m.3243A>G heteroplasmy (x-axis) and their total scmtMPM (y-axis) in samples M29 and M35, with the pathogenic m.3243A>G variant being excluded from the scmtMPM score. **(f)** Correlations between m.3243A>G heteroplasmy and the mtDNA depth

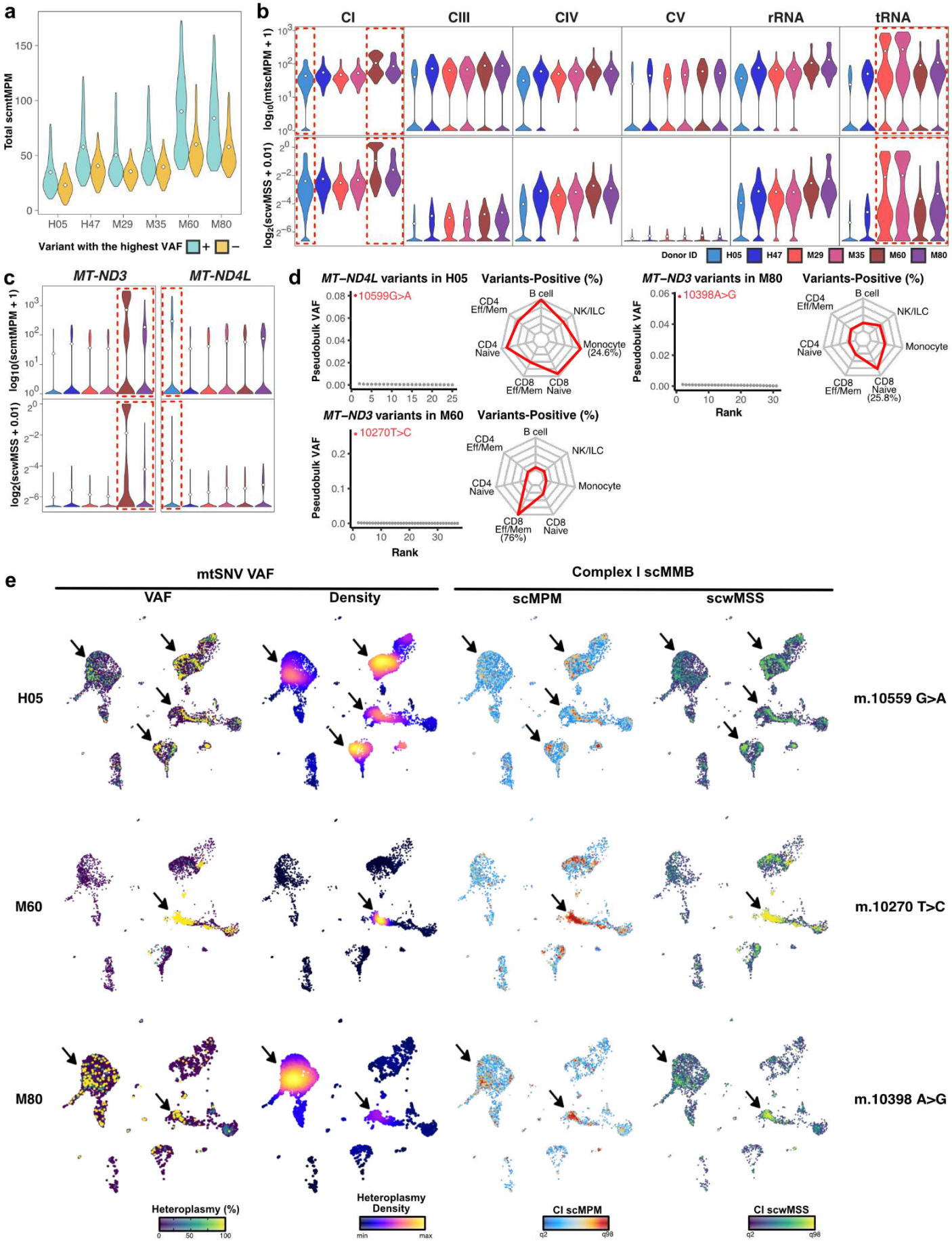
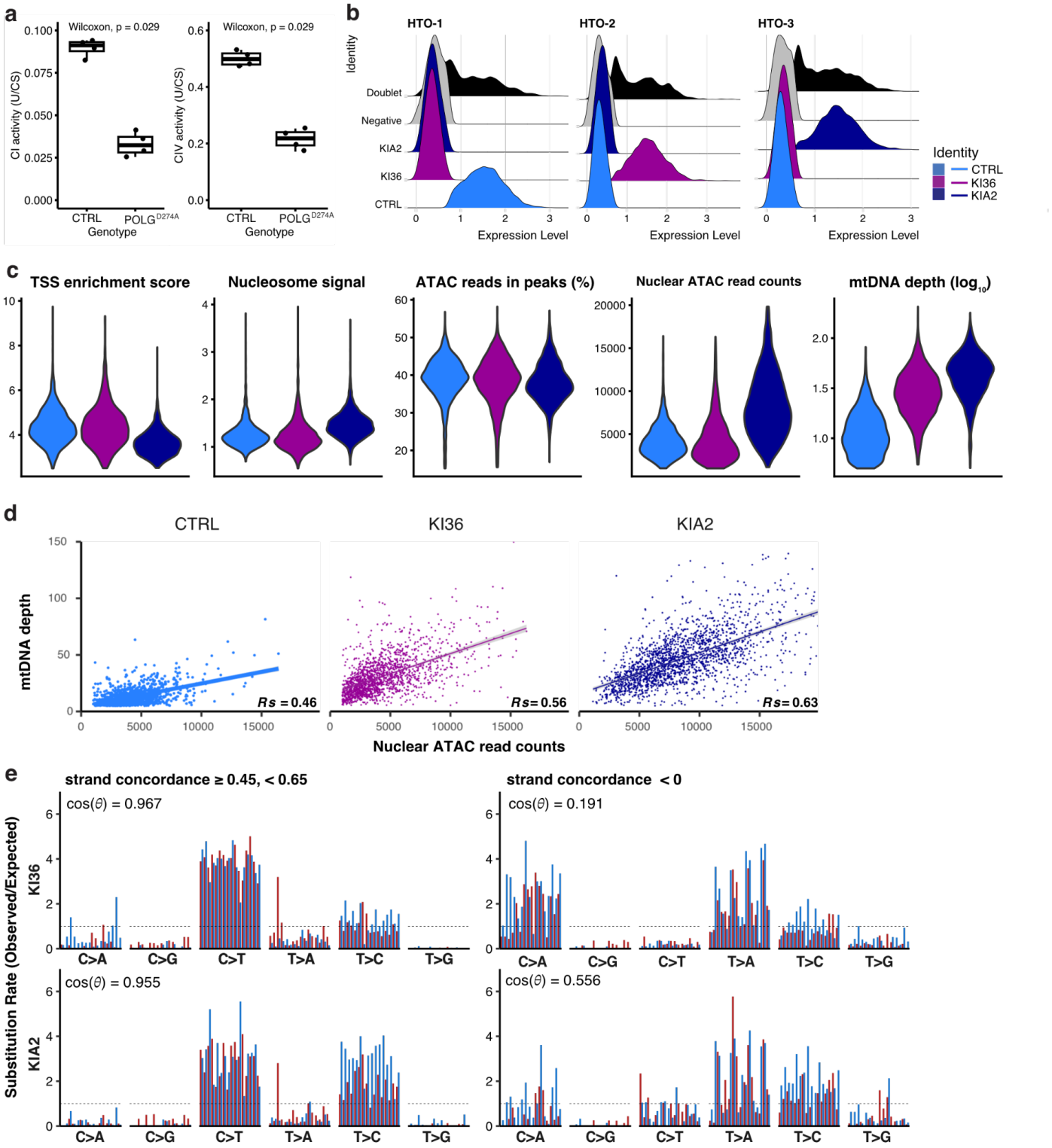
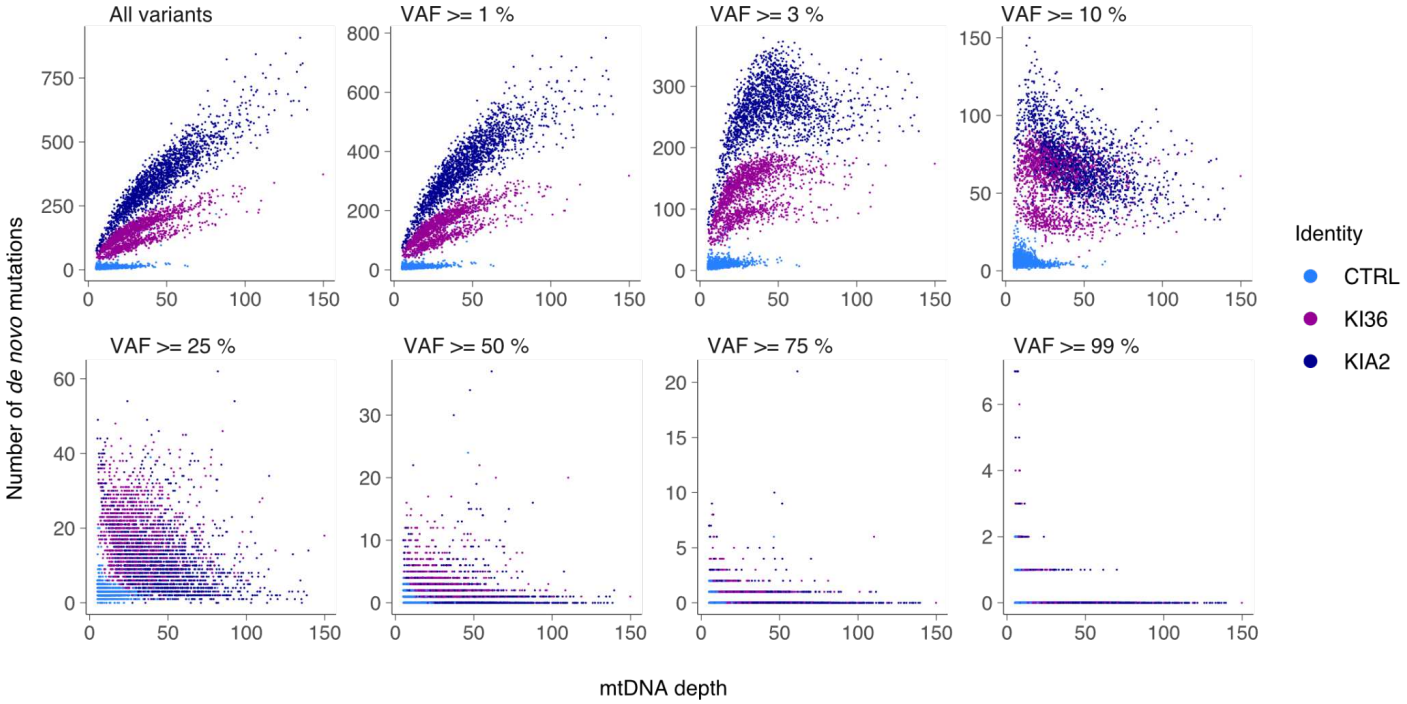

a
Total scmtMPM
150
100
50
0
H05 H47 M29 M35 M60 M80
Variant with the highest VAF + −
b
CI CIII CIV CV rRNA tRNA
log10(mtscMPM + 1)
log2(scwMSS + 0.01)
Donor ID H05 H47 M29 M35 M60 M80
c
MT-ND3 MT-ND4L
log10(scmtMPM + 1)
log2(scwMSS + 0.01)
d
MT–ND4L variants in H05
10599G>A
Pseudobulk VAF
Variants-Positive (%)
B cell
NK/ILC
Monocyte (24.6%)
CD8 Naive
CD8 Eff/Mem
CD4 Naive
CD4 Eff/Mem
MT–ND3 variants in M80
10398A>G
Rank
Monocyte
CD8 Naive (25.8%)
MT–ND3 variants in M60
10270T>C
CD8 Eff/Mem (76%)
e
mtSNV VAF
VAF
Density
Complex I scMMB
scMPM
scwMSS
H05
M60
M80
m.10559 G>A
m.10270 T>C
m.10398 A>G
Heteroplasmy (%)
Heteroplasmy Density
min
max
CI scMPM
CI scwMSS
q2
q98

**Figure 6 – 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 m.3243A>G. The white dots indicate the pseudobulk value. **(b)** Violin plots of total scmtMPM (upper) and scwMSS (lower) of human PBMCs stratified by indicated mitochondrial gene functional groups. Dotted boxes highlight individuals with high complex I (left) and tRNA (right) scMMB metrics. **(c)** Violin plot of scmtMPM (upper) and scwMSS (lower) for 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 indicated mitochondrial genes and individuals (left). The radar plots depict the percentage (%) of the cells that carry the mtDNA variant in the indicated immune cell subsets (right). **(e)** UMAP projection colored by the heteroplasmy (left) and density (middle left) of the identified mtSNVs, and by the scMMB metrics, scmtMPM (middle right), and scwMSS scores (right), of OXPHOS complex I for indicated individuals H05 (upper), M60 (middle), and M80 (lower). Compare to Fig. 5a for cell type annotation.

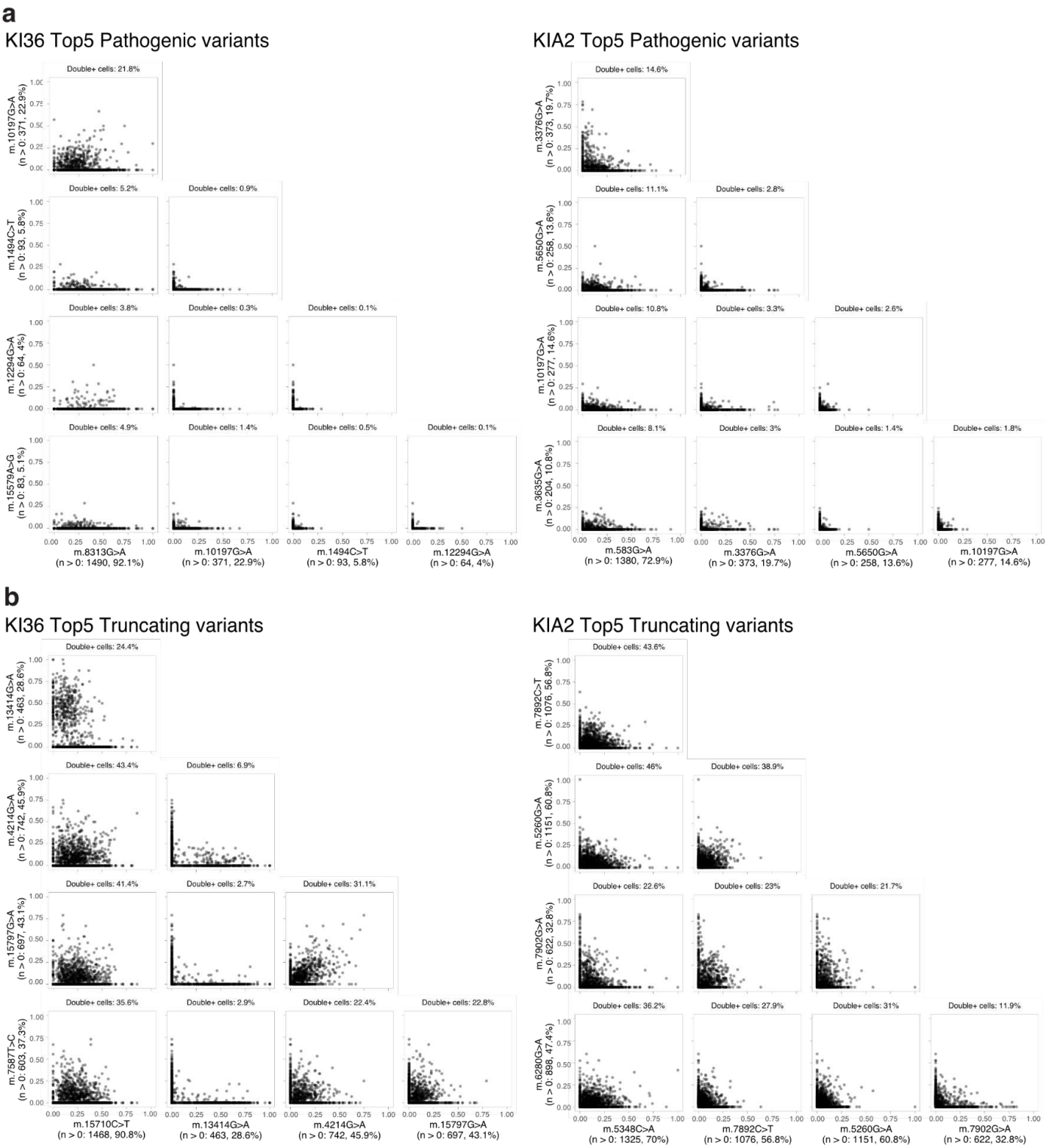



**Extended Data Fig. 1 – mtscATAC-seq quality metrics of *POLG*$^{D274A}$ HEK293 cell lines. (a)** Enzyme activity of mitochondrial complex I (CI, left) and IV (CIV, left). Data were represented as units normalized to citrate synthase (CS) activity (U/CS). **(b)** Ridgeplots of samples after hashtag-antibody demultiplexing. **(c)** Violin plot of TSS enrichment score, nucleosome signal, percent reads in peaks, unique ATAC read counts, and mitochondrial DNA depth. **(d)** Correlations between ATAC read counts and mtDNA depth. Regression lines with the Spearman coefficient ($R_s$) are shown in the figure. **(e)** Mutational signature of mtDNA variants with an intermediate (≥0.45, <0.65, left) or negative (<0, right) strand concordance in both

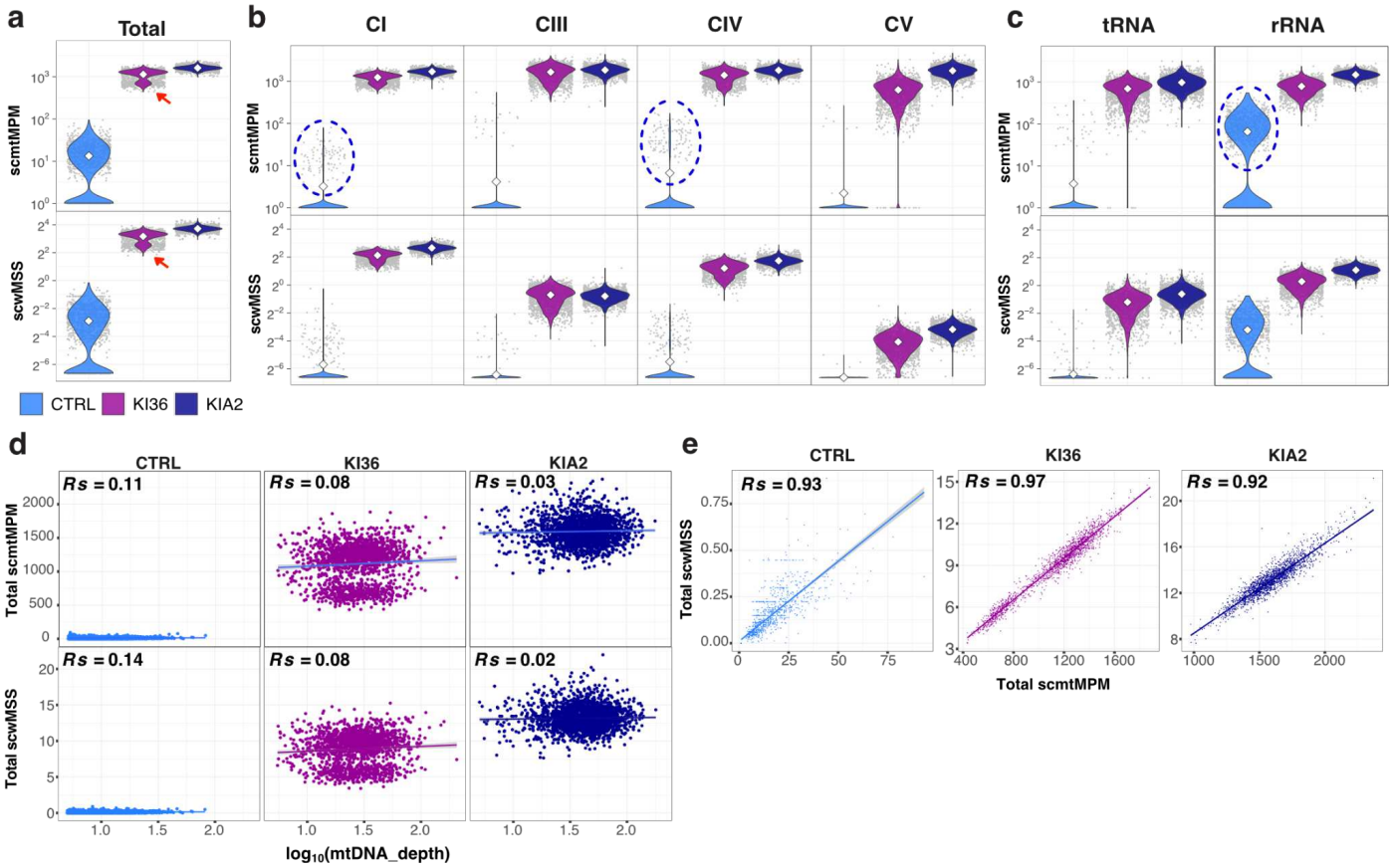
*POLG*$^{D274A}$ (upper, KI36; lower, KIA2) cell lines. The cosine similarity (cos($\theta$)) to the mutational signature of high-confident mtDNA variants (strand concordance > 0.65) is shown.



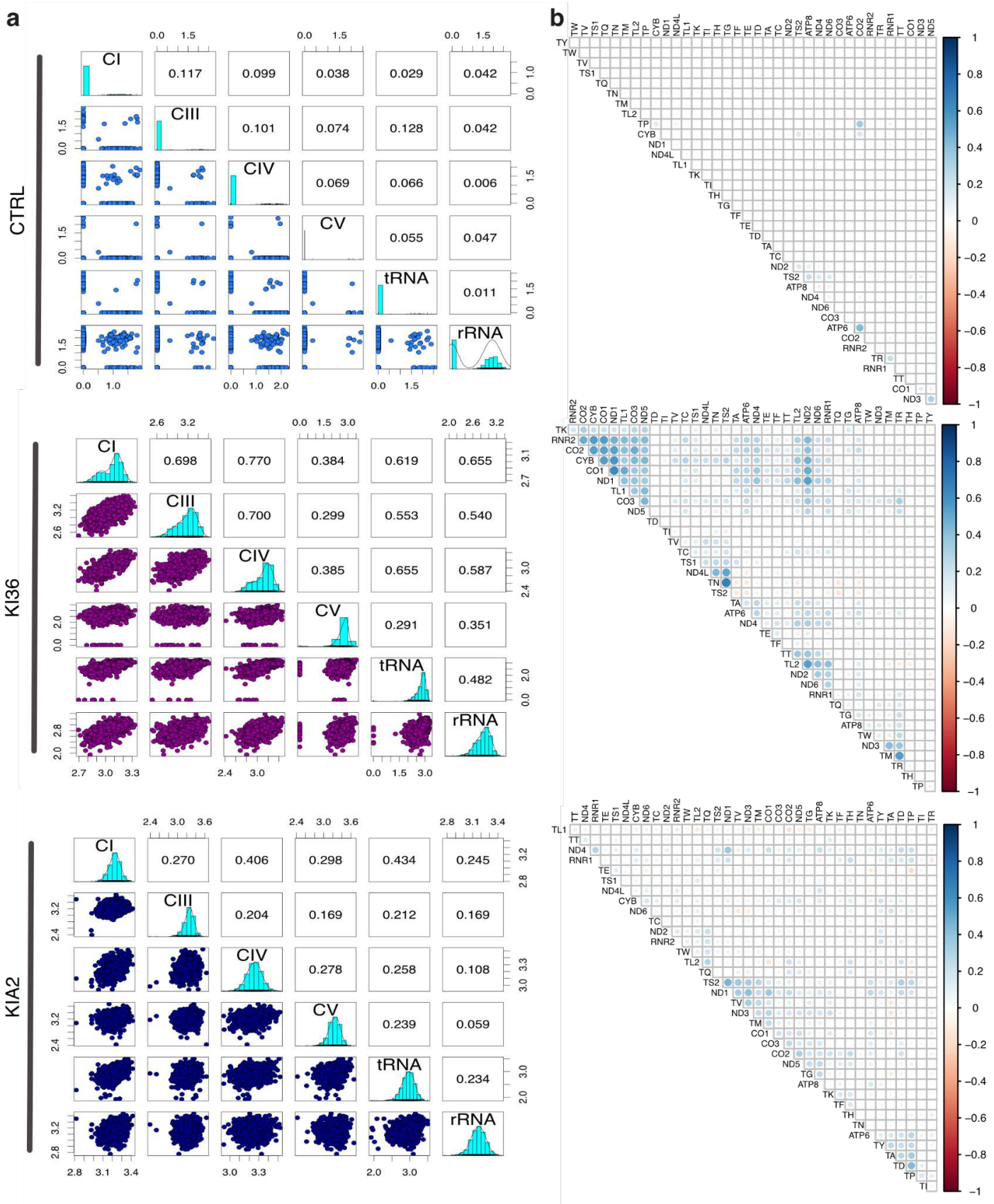

**Extended Data Fig. 2 – Correlations between mitochondrial DNA depth and the number of detected high-confidence *de novo* somatic mtDNA variants in *POLG*$^{D274A}$ cell lines.** The number of high-confidence *de novo* mtDNA variants at different variant allele frequency cut-offs versus mtDNA depth in single cells is plotted and colored by the identity of control or *POLG*$^{D274A}$ cell lines.




**Extended Data Fig. 3 – Pairwise distribution of top-ranked truncating and pathogenic mtDNA variants.** **(a,b)** Scatter plot matrices show pairwise comparisons of top mtDNA variants from each dataset, stratified by mutation type: (a) pathogenic and (b) truncating variants. Each panel displays variants ranked by pseudobulk VAF. 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).

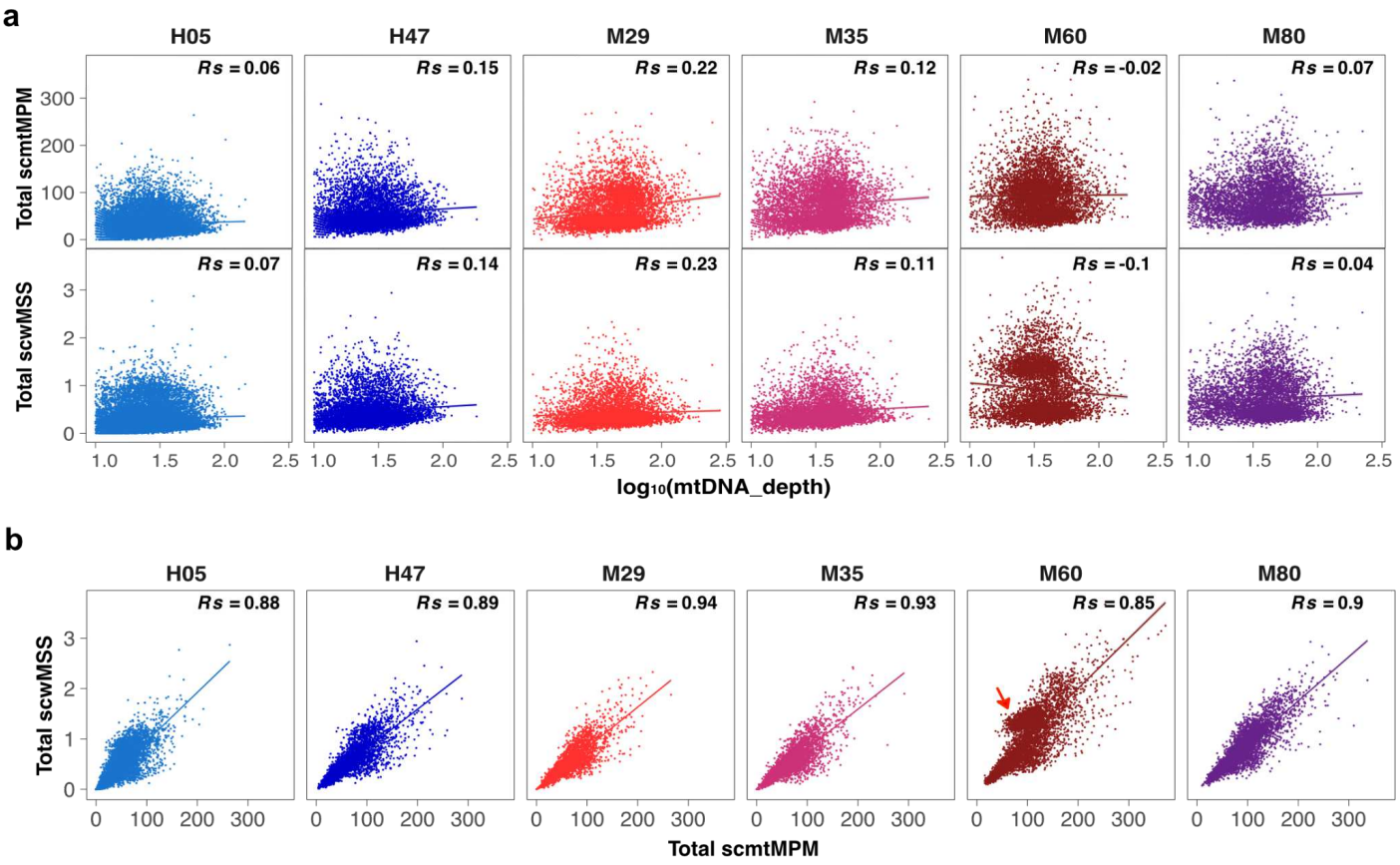



**Extended Data Fig. 4 – Supporting information on mtDNA mutational burden quantifications in *POLG*$^{D274A}$ HEK293 cell lines. (a-c)** Violin plots of Mutations per million bases (scmtMPM) and Heteroplasmy-weighted sum of mitochondrial constraint scores (scwMSS) of mtDNA for indicated respiratory chain complexes (a), tRNA (b), and rRNA (c) genes in *POLG*$^{D274A}$ cell lines. 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 mtDNA mutational burden quantifications. **(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.

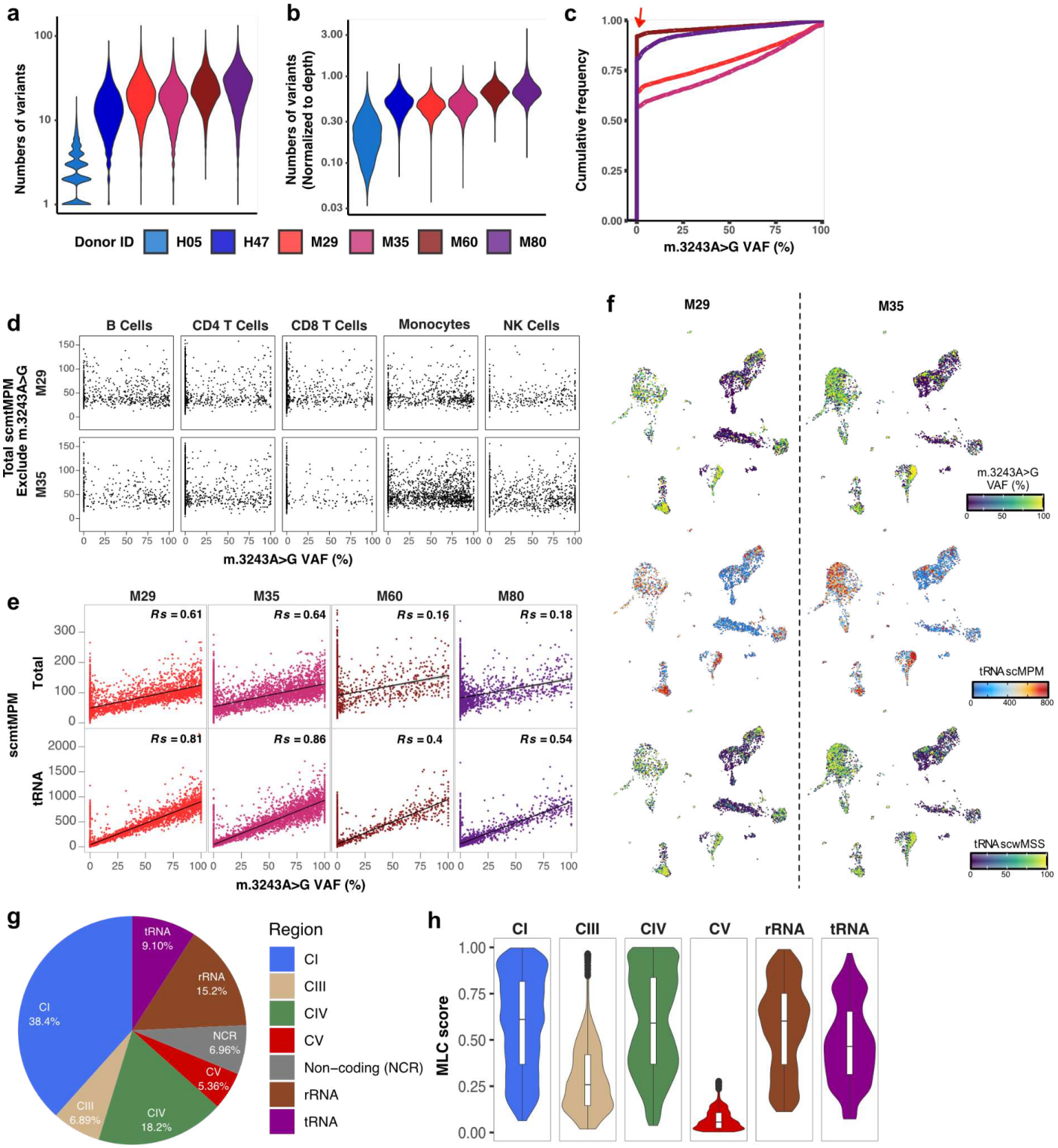



**Extended Data Fig. 5 – Correlative distributions of somatic mtDNA mutational burden across the mitochondrial genome in *POLG*$^{D274A}$ knock-in cell lines. (a)** Pair-plot of scmtMPM for individual cells from the control line (CTRL, top), and the two hyper-mutated lines (K36, middle; and KIA2, bottom), 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.

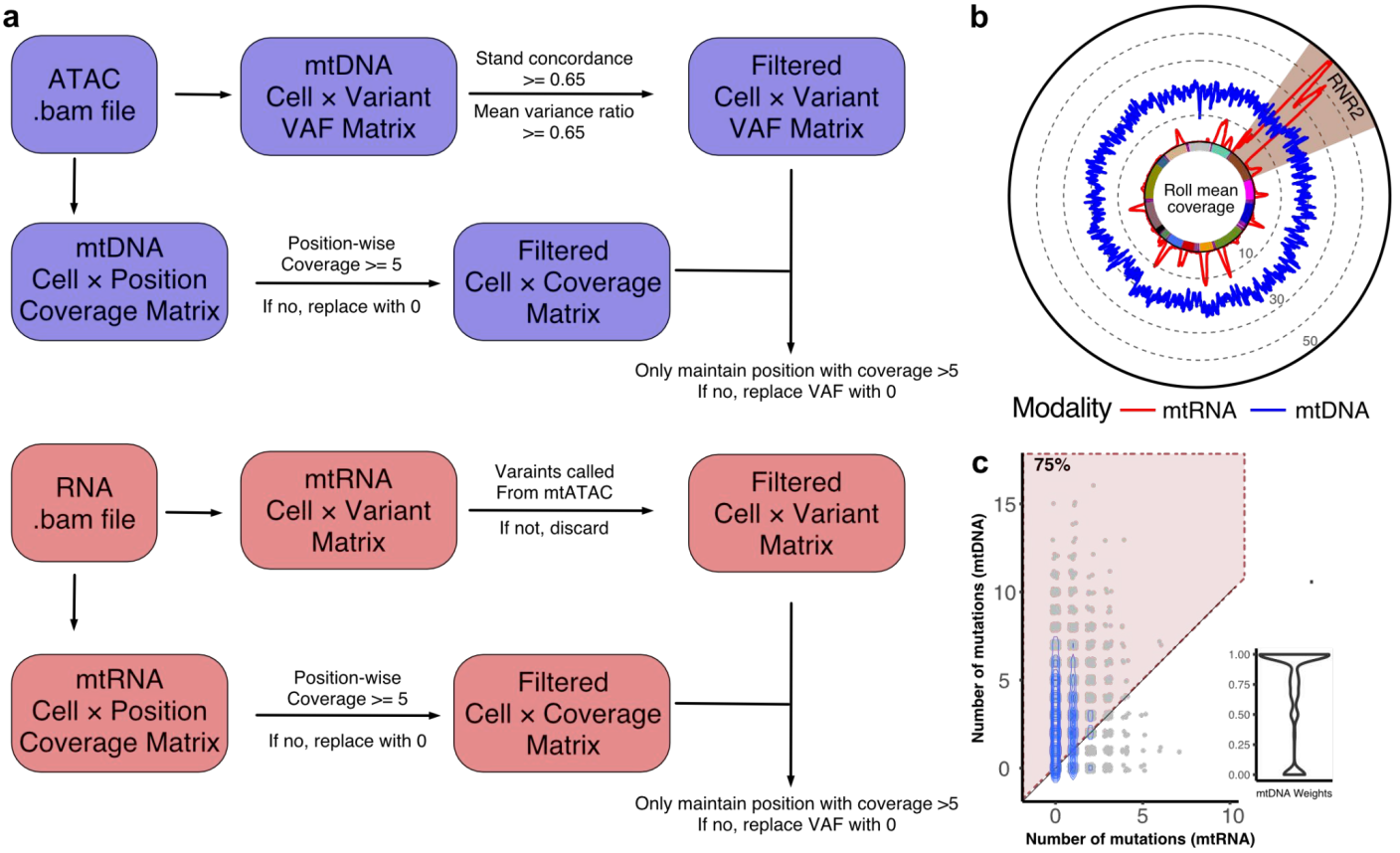



**Extended Data Fig. 6 – Supporting information on mtDNA mutational burden quantification in primary human blood cells. (a)** Correlations between mitochondrial DNA depth and mtDNA mutational burden quantification in individual donors. **(b)** Correlations between scmtMPM and scwMSS. Regression lines with the Spearman coefficient ($R_s$) are shown in the figure.




**Extended Data Fig. 7 – Supporting information on scMMB quantification stratified by functional and gene groups in primary human blood cells. (a,b)** Violin plots of the number of mtSNVs detected with (a) or without (b) normalization to mtDNA depth. **(c)** Cumulative distribution plots of m.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 m.3243A>G heteroplasmy and the total scmtMPM, with the pathogenic m.3243A>G variant being deliberately excluded from the scmtMPM score. **(e)** Correlations between m.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 m.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])




**Extended Data Fig. 8 – 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 weights of mtDNA in detecting variants compared to mtRNA, and 75% of the cells have higher mtDNA weight (>0.5).

**Table S1. Genetic and functional overview of complex I mtDNA variants**

| Position | Variants | Gene | Mutation | Amino acid change | MLC | SIFT[a]/ AlphaMissense[54] | MITOMAP |
|---|---|---|---|---|---|---|---|
| 10599 | G>A | *MT-ND4* | missense | A44T | 0.747 | likely benign | - |
| 10270 | T>C | *MT-ND3* | missense | L71P | 0.969 | likely pathogenic | - |
| 10398 | A>G | *MT-ND3* | missense | T114A | 0.578 | likely benign | pleiotropic [45,55–58] |

[a] Sorting Intolerant from Tolerant (SIFT) [59]

## Methods

*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 described[26]. Briefly, control and KI36 HEK293 cell lines were cultured in basal medium Dulbecco's Modified Eagle Medium (DMEM) 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 a maximum of 2 weeks before single-cell sequencing to limit the accumulation of additional mtDNA mutations.

*Mitochondrial enzyme activity measurements*

The activity of the mitochondrial enzymes NADH:$CoQ_1$ oxidoreductase (complex I, CI), cytochrome *c* oxidase (complex IV, CIV) and citrate synthase (CS) was measured as previously described[60]. All measurements were performed at 30°C with 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. TotalSeq™-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 minutes on ice. Cell lines were then stained with 0,25 µg of individual hashtag antibodies 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.

*Mitochondrial single-cell ATAC-seq (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 TotalSeq™-A hashtag antibodies. Briefly, sorted cells were fixed in 1% formaldehyde (FA; ThermoFisher) in PBS for 10 min at room temperature, quenched with glycine solution to a final concentration of 0.125 M before washing twice with PBS/BSA via centrifugation at 400 g, 5 min, 4°C. Cells were subsequently permeabilized with lysis buffer (10mM Tris-HCL pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% NP40, 1% BSA) for 3 minutes (primary cells) or 5 minutes (cell lines) on ice, followed by adding 1 ml of chilled wash buffer (10mM Tris-HCL pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 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 on 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 96-Deep Well Reaction Module (BioRad). 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 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). Single-cell mtATAC-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 preprocessed 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)[61]. For quality control, cells fulfilling all the following criteria were kept for downstream analysis: > 1.000 but < 25.000 unique nuclear fragments, and TSS enrichment score > 3. Normalization and dimensional reduction were performed using TF-IDF normalization followed by 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 cell matrices were then normalized and scaled using the centered log-ratio (CLR) transformation and 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) method[62].

*Mitochondrial DNA variants calling*

Mitochondrial DNA variants were identified from the ATAC-seq modality using mgatk (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, 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 analysis[49]. 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 **Extended Data Fig. 1d**. However, all other analyses presented throughout the manuscript exclusively utilized variants identified via the more strict filtering criteria. The list of confirmed pathogenic mtDNA variants was obtained from the MITOMAP website [36,63] (http://www.mitomap.org).

*Single-cell mitochondrial DNA mutational burden (scMMB)*

We adapted metrics previously used to assess mtDNA mutational burden [6] and local constraint[25,40,41]. 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 level of individual genes, respiratory chain complexes (I, III, IV and V), and at the whole mtDNA genomic level, as previously conducted in whole-exome sequencing data[7]. Mutation counts were calculated by first multiplying the `allele_frequency` and `coverage` matrices, and then summing up across respective genomic regions. The total genomic length sequenced was calculated by summing up the coverage matrix of respective genomic regions in individual cells. 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)[25] 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 donor[15] using a cross-calling and filtering workflow (**Extended Data Fig. 8a**). 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 the uneven coverage of mtRNA transcripts, we further filtered the cell-by-position mtDNA/mtRNA coverage matrices and only retained positions with at least five reads in individual cells. 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 (**Fig. S7C**). 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).

**Code availability**

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

## Data availability

Raw .fastq files for publicly accessible datasets were downloaded from the Gene Expression Omnibus (GEO) and the Genotypes and Phenotypes database (dbGaP). The healthy adult mtscATAC-seq dataset was downloaded from GEO under the accession number GSE142745. mtscATAC datasets from MELAS patients (M29, M35, M60) were available on dbGAP under the study accession number phs002217.v1.p1. HEK293 cell line data is available at GEO accession GSE291877 (reviewer token: **szghasqoftsrxkd**).


## ACKNOWLEDGEMENTS

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 Programme (LSL), an Emmy Noether fellowship (LU 2336/2-1) and grants by the German Research Foundation (DFG, LU 2336/3-1, LU 2336/6-1, STA 1586/5-1, TRR241, SFB1588, Heinz Maier-Leibnitz Award to LSL).


## AUTHOR CONTRIBUTIONS

YHH and LSL conceived and designed the project. YHH led, designed, and performed experiments with JL. YHH led all analyses with input from PK, AG, LN, MS, CAL, and LSL. GT, GZ, and WSK contributed to HEK293-*Polg*$^{D274A}$ cell lines and biological perspectives. PK and LN contributed to healthy sample data generation. CAL and LSL supervised various aspects of this work. YHH and LSL wrote the paper with input from all authors.

## 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.

## References


1. Nitsch, L., Lareau, C. A. & Ludwig, L. S. Mitochondrial genetics through the lens of single-cell multi-omics. *Nat. Genet.* **56**, 1355–1365 (2024).
2. Campbell, P. *et al.* Mitochondrial mutation, drift and selection during human development and ageing. (2023) doi:10.21203/rs.3.rs-3083262/v1.
3. 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).
4. Sanchez-Contreras, M. *et al.* The multi-tissue landscape of somatic mtDNA mutations indicates tissue-specific accumulation and removal in aging. *Elife* **12**, (2023).
5. Kang, E. *et al.* Age-Related Accumulation of Somatic Mitochondrial DNA Mutations in Adult-Derived Human iPSCs. *Cell Stem Cell* **18**, 625–636 (2016).
6. Greaves, L. C. *et al.* Clonal Expansion of Early to Mid-Life Mitochondrial DNA Point Mutations Drives Mitochondrial Dysfunction during Human Ageing. *PLOS Genetics* **10**, e1004620 (2014).
7. Gorelick, A. N. *et al.* Respiratory complex and tissue lineage drive recurrent mutations in tumour mtDNA. *Nature Metabolism* **3**, 558–570 (2021).
8. Yuan, Y. *et al.* Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* **52**, 342–352 (2020).
9. Smith, A. L. M. *et al.* Age-associated mitochondrial DNA mutations cause metabolic remodeling that contributes to accelerated intestinal tumorigenesis. *Nature Cancer* **1**, 976–989 (2020).
10. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat Rev Dis Primers* **2**, 16080 (2016).
11. Gorman, G. S. *et al.* Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann. Neurol.* **77**, 753–759 (2015).
12. 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).
13. Stewart, J. B. & Chinnery, P. F. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nat. Rev. Genet.* **22**, 106–118 (2020).
14. Walker, M. A. *et al.* Purifying Selection against Pathogenic Mitochondrial DNA in Human T Cells. *N.*

*Engl. J. Med.* **383**, 1556–1563 (2020).

15. 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).
16. Zhang, J. *et al.* Antigen receptor stimulation induces purifying selection against pathogenic mitochondrial tRNA mutations. (2023) doi:10.1172/jci.insight.167656.
17. Franklin, I. G. *et al.* T cell differentiation drives the negative selection of pathogenic mitochondrial DNA variants. *Life Science Alliance* **6**, (2023).
18. Trifunovic, A. *et al.* Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417–423 (2004).
19. Kujoth, G. C. *et al.* Mitochondrial DNA Mutations, Oxidative Stress, and Apoptosis in Mammalian Aging. *Science* (2005) doi:10.1126/science.1112125.
20. Normal Levels of Wild-Type Mitochondrial DNA Maintain Cytochrome c Oxidase Activity for Two Pathogenic Mitochondrial DNA Mutations but Not for m.3243A→G. *The American Journal of Human Genetics* **81**, 189–195 (2007).
21. Rossignol, R. *et al.* Mitochondrial threshold effects. *Biochem. J* **370**, 751–762 (2003).
22. Lareau, C. A. *et al.* Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2020).
23. Kim, M. *et al.* Single-cell mtDNA dynamics in tumors is driven by coregulation of nuclear and mitochondrial genomes. *Nat. Genet.* **56**, 889–899 (2024).
24. 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).
25. Lake, N. J. *et al.* Quantifying constraint in the human mitochondrial genome. *Nature* 1–8 (2024).
26. Peeva, V. *et al.* Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nat. Commun.* **9**, 1727 (2018).
27. 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).
28. Deep sequencing of mitochondrial DNA and characterization of a novel POLG mutation in a patient

785 with arPEO. *Neurology Genetics* https://www.neurology.org/doi/10.1212/NXG.0000000000000391.
786 29. Filograna, R. *et al.* Modulation of mtDNA copy number ameliorates the pathological consequences of a
787 heteroplasmic mtDNA mutation in the mouse. *Science Advances* (2019) doi:10.1126/sciadv.aav9824.
788 30. Mitochondrial genotype modulates mtDNA copy number and organismal phenotype in Drosophila.
789 *Mitochondrion* **34**, 75–83 (2017).
790 31. Maclaine, K. D., Stebbings, K. A., Llano, D. A. & Havird, J. C. The mtDNA mutation spectrum in the
791 PolG mutator mouse reveals germline and somatic selection. *BMC Genom Data* **22**, 52 (2021).
792 32. Ni, T. *et al.* MitoRCA-seq reveals unbalanced cytocine to thymine transition in Polg mutant mice.
793 *Scientific Reports* **5**, 1–13 (2015).
794 33. Origins of human mitochondrial point mutations as DNA polymerase γ-mediated errors. *Mutation*
795 *Research/Fundamental and Molecular Mechanisms of Mutagenesis* **599**, 11–20 (2006).
796 34. Liu, M. H. *et al.* DNA mismatch and damage patterns revealed by single-molecule sequencing. *Nature*
797 **630**, 752–761 (2024).
798 35. Kotrys, A. V. *et al.* Single-cell analysis reveals context-dependent, cell-level selection of mtDNA.
799 *Nature* **629**, 458–466 (2024).
800 36. Lott, M. T. *et al.* mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Current Protocols in*
801 *Bioinformatics* **44**, 1.23.1–1.23.26 (2013).
802 37. Muyas, F. *et al.* De novo detection of somatic mutations in high-throughput single-cell profiling data
803 sets. *Nat. Biotechnol.* **42**, 758–767 (2023).
804 38. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421
805 (2013).
806 39. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**,
807 321–327 (2018).
808 40. Hong, Y. S. *et al.* Deleterious heteroplasmic mitochondrial mutations are associated with an increased
809 risk of overall and cancer-specific mortality. *Nat. Commun.* **14**, 6113 (2023).
810 41. Hong, Y. S. *et al.* Mitochondrial heteroplasmy improves risk prediction for myeloid neoplasms. *Nature*
811 *Communications* **15**, 1–15 (2024).
812 42. Lawless, C., Greaves, L., Reeve, A. K., Turnbull, D. M. & Vincent, A. E. The rise and rise of

mitochondrial DNA mutations. *Open Biology* (2020) doi:10.1098/rsob.200061.

43. Schaefer, P. M. *et al.* Combination of common mtDNA variants results in mitochondrial dysfunction and a connective tissue dysregulation. *Proceedings of the National Academy of Sciences* **119**, e2212417119 (2022).
44. An, J. *et al.* Mitochondrial DNA mosaicism in normal human somatic cells. *Nature Genetics* **56**, 1665–1677 (2024).
45. 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).
46. Bhat, A. *et al.* The possible role of 10398A and 16189C mtDNA variants in providing susceptibility to T2DM in two North Indian populations: a replicative study. *Human genetics* **120**, (2007).
47. A mitochondrial DNA variant 10398G>A in breast cancer among South Indians: An original study with meta-analysis. *Mitochondrion* **13**, 559–565 (2013).
48. Canter, J. A., Kallianpur, A. R., Parl, F. F. & Millikan, R. C. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* **65**, 8028–8033 (2005).
49. Laricchia, K. M. *et al.* Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome Res.* **32**, 569–582 (2022).
50. Bolze, A. *et al.* A catalog of homoplasmic and heteroplasmic mitochondrial DNA variants in humans. *bioRxiv* 798264 (2020) doi:10.1101/798264.
51. Ludwig, L. S. *et al.* Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).
52. Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology* **39**, 1246–1258 (2021).
53. Wei, W. *et al.* Germline selection shapes human mitochondrial DNA diversity. *Science* **364**, eaau6520 (2019).
54. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* (2023) doi:10.1126/science.adg7492.
55. Tzeng, I. S. Role of mitochondria DNA A10398G polymorphism on development of Parkinson's disease: A PRISMA-compliant meta-analysis. *Journal of clinical laboratory analysis* **36**, (2022).

56. Juo, S. H. *et al.* A common mitochondrial polymorphism 10398A>G is associated metabolic syndrome in a Chinese population. *Mitochondrion* **10**, (2010).
57. Pezzotti, A. *et al.* The mitochondrial A10398G polymorphism, interaction with alcohol consumption, and breast cancer risk. *PloS one* **4**, (2009).
58. van der Walt, J. M. *et al.* Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *American journal of human genetics* **72**, (2003).
59. Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452–W457 (2012).
60. 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).
61. Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biology* **22**, 1–19 (2021).
62. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42**, 293–304 (2023).
63. Kogelnik, A. M., Lott, M. T., Brown, Navathe, S. B. & Wallace, D. C. MITOMAP: a human mitochondrial genome database--1998 update. *Nucleic acids research* **26**, (1998).