










# MitoPerturb-Seq identifies gene-specific single-cell responses to mitochondrial DNA depletion and heteroplasmy

Received: 20 June 2025

Accepted: 26 February 2026

Published online: 01 April 2026

 Check for updates

Stephen P. Burr <sup>1,2,5</sup>, Kathryn Auckland <sup>1,2,5</sup>, Angelos Glynos<sup>1,2</sup>, Abhilesh Dhawanjewar <sup>1,2</sup>, Cameron Ryall <sup>1,2</sup>, Wei Wei <sup>1,2</sup>, Antony Hynes-Allen<sup>1,2</sup>, Malwina Prater <sup>1,2,4</sup>, Matylda Sczaniecka-Clift<sup>3</sup>, Julien Prudent <sup>1</sup>, Patrick F. Chinnery <sup>1,2</sup> ✉ & Jelle van den Ameele <sup>1,2</sup> ✉

Mitochondria contain their own genome, mitochondrial DNA (mtDNA), which is under strict control by the cell nucleus. mtDNA occurs in many copies per cell and mutations often only affect a proportion of them, giving rise to heteroplasmy. mtDNA copy number and heteroplasmy level together shape the tissue-specific impact of mtDNA mutations, eventually giving rise to both rare mitochondrial and common neurodegenerative diseases. Here, we use MitoPerturb-Seq for CRISPR–Cas9-based, high-throughput single-cell interrogation of the nuclear genes and pathways that sense and control mtDNA copy number and heteroplasmy. We screened a panel of mtDNA maintenance genes in mouse cells with a heteroplasmic mtDNA *mt-Ta* mutation. This revealed both common and perturbation-specific aspects of the integrated stress response to mtDNA depletion caused by *Tfam*, *Opa1* and *Polg* knockout. These responses are only partially mediated by ATF4 and cause cell-cycle stage-independent slowing of cell proliferation. MitoPerturb-Seq, thus, provides experimental insight into disease-relevant mitochondrial–nuclear interactions and may inform development of therapies targeting cell-type- and tissue-specific vulnerabilities to mitochondrial dysfunction.

Mitochondria store their own genetic information in a ~16.5-kb small circular genome, the mtDNA, which encodes 37 genes required for mitochondrial oxidative phosphorylation (OXPHOS). Each cell may contain hundreds to thousands of copies of mtDNA and their replication and transcription are strictly controlled by nuclear-encoded mitochondrial localized proteins. mtDNA copy number (CN), the absolute number of mtDNA molecules per cell, shows considerable tissue-specific and cell-type-specific variability<sup>1,2</sup>. Furthermore, an age-dependent decrease in mtDNA CN may contribute to aging and neurodegeneration<sup>2,3</sup>. Mutations in the mtDNA often only

affect a proportion of mtDNA molecules within a cell, a state called heteroplasmy<sup>4,5</sup>. During development and aging, the number and proportion of mutated mtDNA molecules can increase to high levels in single cells. Symptoms arise when deleterious heteroplasmy levels reach a cell-type-specific threshold, leading to biochemical OXPHOS defects, cell dysfunction and cell death. Age-related clonal expansion of mtDNA mutations has been implicated in the pathogenesis of rare mitochondrial diseases<sup>6</sup> and common neurodegenerative disorders, including Alzheimer's and Parkinson's disease<sup>7,8</sup>, as well as in specific types of cancer<sup>9</sup>. Identification of factors and pathways that regulate

<sup>1</sup>MRC Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK. <sup>2</sup>Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. <sup>3</sup>MRC Laboratory of Molecular Biology, University of Cambridge, Cambridge, UK. <sup>4</sup>Present address: Altos Labs Cambridge Institute, Cambridge, UK.

<sup>5</sup>These authors contributed equally: Stephen P. Burr, Kathryn Auckland. ✉e-mail: [pfc25@cam.ac.uk](mailto:pfc25@cam.ac.uk); [jv361@cam.ac.uk](mailto:jv361@cam.ac.uk)

mtDNA CN and reduce heteroplasmy, even by a few percent below the threshold, offers the possibility to reverse biochemical defects and confer protection against a range of rare and common diseases.

Genome-wide association studies have shown the importance of nuclear loci modulating mtDNA CN and heteroplasmy but little is known about the specific genes involved. Most studies to date have been correlative and were based on bulk tissue analysis containing diverse cell types in varying proportions, with inconsistent and contradictory findings<sup>4,10–16</sup>. As a result, for many nuclear-encoded genes and genomic loci, a direct causal link between gene activity and mtDNA dynamics remains to be established. Moreover, given the extraordinary degree of mtDNA mosaicism across tissues, with virtually all cells containing different mixtures and levels of wild-type (WT) and mutant mtDNA<sup>17–19</sup>, it remains unclear how mtDNA CN and heteroplasmy levels interact within individual cells to cause a downstream tissue-specific phenotype.

Here, we describe and deploy MitoPerturb-Seq, allowing us to determine the impact of specific nuclear genetic perturbations on mtDNA CN, heteroplasmy and the resulting transcriptomic response at single-cell resolution. Targeting a library of candidate nuclear genetic modifiers of mtDNA dynamics in heteroplasmic mouse embryonic fibroblasts (MEFs), we show gene-specific modulation of mtDNA CN and heteroplasmy variance, affecting cell-cycle progression and the associated nuclear transcriptional response. DamID-seq-based chromatin profiling confirmed ATF4 as a key mediating transcription factor for some but not all nuclear genes responding to decreased mtDNA levels. MitoPerturb-Seq, thus, offers a unique opportunity for unbiased high-throughput interrogation of disease-relevant interactions between nuclear and mitochondrial genomes within single, isogenic cells.

## Results

### Single-cell CRISPR screening with whole-cell multiome in heteroplasmic cells

We established MitoPerturb-Seq to understand whether disruption of candidate nuclear genes modulates mtDNA at the single-cell level. This would allow us to identify genes involved in regulating mutation burden and/or mtDNA CN in cells with heteroplasmic mtDNA mutations, where changes in these parameters are known to affect disease phenotypes<sup>20,21</sup>. MitoPerturb-Seq combines CROPseq for pooled single-cell CRISPR screening<sup>22</sup> with a 10X Genomics multiome-based approach<sup>23,24</sup> for combined single-cell (sc)ATAC-seq and scRNA-seq in whole cells. This enables simultaneous profiling of mtDNA sequence, CN and heteroplasmy from scATAC-seq and the detection of guide RNA (gRNA) sequences from the single-cell transcriptome (Fig. 1a and Extended Data Fig. 1a). We generated Cas9-expressing MEFs (Extended Data Fig. 1b) from heteroplasmic mice carrying an m.5024C>T point mutation in the mitochondrial tRNA<sup>Ala</sup> gene (*mt-Ta*)<sup>11,25</sup>, which corresponds to the human disease-causing heteroplasmic m.5650G>A tRNA<sup>Ala</sup> variant<sup>25–27</sup> (Fig. 1a). Selected Cas9-transgenic MEFs had a mean heteroplasmy of 60.8% ± 5.8% (mean ± s.d.), just below the threshold for biochemical defects in most tissues<sup>25</sup>, with low intercellular variability (Extended Data Fig. 1c). MEFs were transduced with a pooled gRNA library (60 gRNAs, three gRNAs per gene, six control genes, three nontargeting (NT) gRNAs) to perturb 13 nuclear genes encoding proteins previously suggested to affect mtDNA CN and/or heteroplasmy, either through their role in mtDNA maintenance (*Akap1* (ref. 28), *Nnt13*, *Polg*<sup>12,29</sup> and *Tfam*<sup>10,30,31</sup>), mitochondrial membrane remodeling (*Dnm11* (also known as *Drp1*)<sup>32</sup>, *Mtfp1* (ref. 33), *Opa1* (ref. 34,35) and *Snx9* (ref. 36,37) or mitochondrial biogenesis and mitophagy (*Atg5* (ref. 38), *Oma1* (ref. 13), *Pink1* (ref. 28,39), *Ppargc1a*<sup>40</sup> and *Prkn*<sup>39,41</sup>) (Fig. 1b and Supplementary Table 1). Then, 10 days after transduction, gRNA-expressing cells were processed for whole-cell multiome (Fig. 1a and Extended Data Fig. 1d–f). Following quality control (QC) of raw sequencing reads, barcode processing and cell filtering, 5,718

high-confidence single cells were identified, with an average of 5,967 genes and 11,783 unique ATAC-seq peaks per cell; data were visualized by weighted nearest neighbor (WNN) uniform manifold approximation and projection (UMAP) after cell-cycle correction (Fig. 1c–e and Extended Data Fig. 1g–m). In total, 0.0022% of RNA-seq reads aligned to gRNAs (Extended Data Fig. 2a) and 15.8% of ATAC-seq reads aligned to mtDNA, corresponding to 31.7× mean mtDNA sequencing depth per cell.

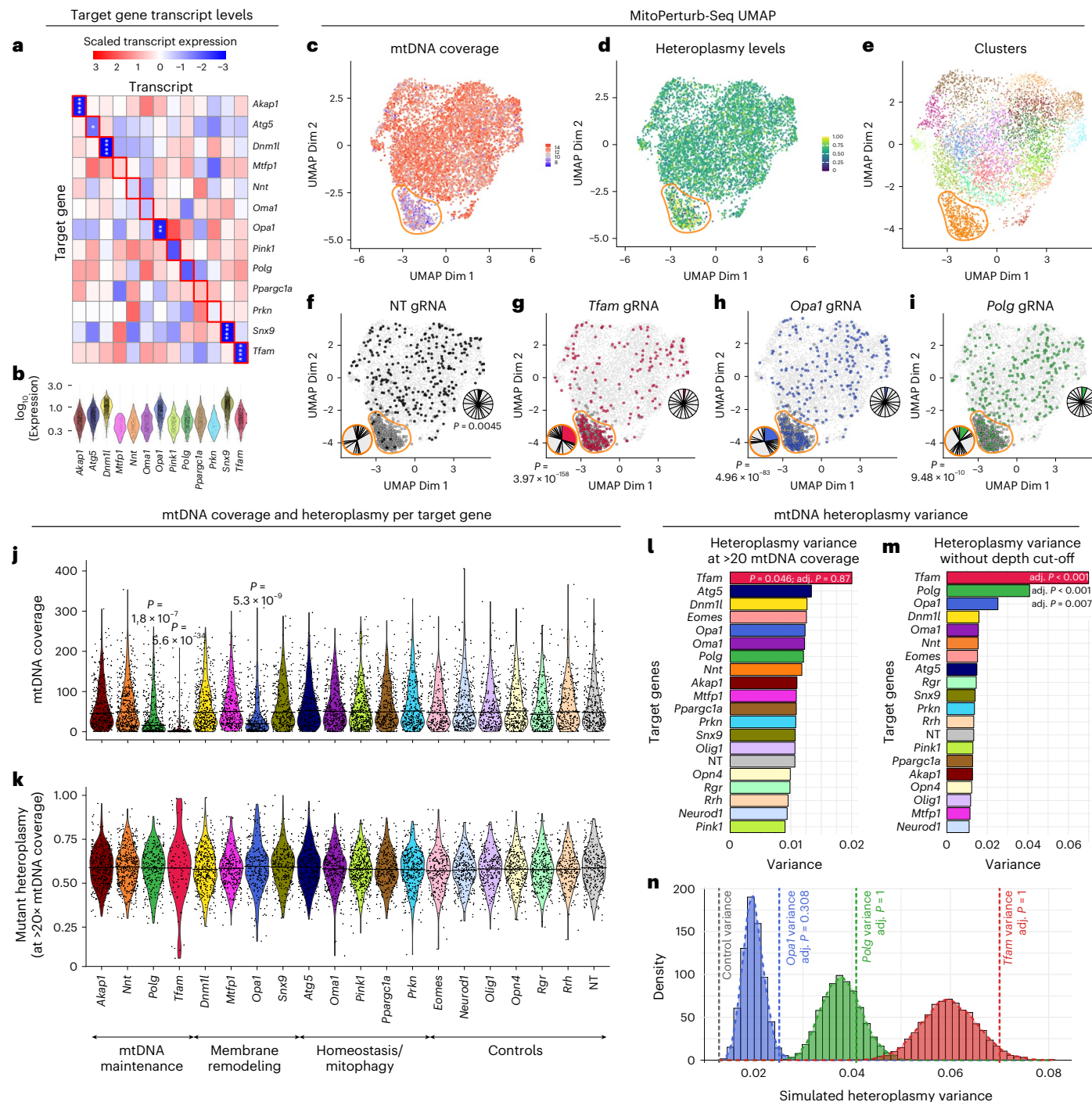
To improve gRNA assignment, we used TAP-seq<sup>42</sup> for PCR-based enrichment of gRNA transcripts. This increased the number of cells assigned to a specific gRNA from 609 (10.7%) to 3,586 (62.7%), expanding the number of cells in each target gene group to 179 ± 41 (mean ± s.d.) (Fig. 1f). As the confidence of estimating true heteroplasmy levels increases with higher read depth<sup>43</sup> (Extended Data Fig. 2b,c), we also sought to further increase single-cell mtDNA coverage. Masking nuclear-localized mitochondrial sequences (NUMTs) in the reference genome to prevent dual alignment to mtDNA and NUMT sequences<sup>17</sup> increased average sequencing depth per cell to 47 × ± 37 × (21.0% of ATAC-seq reads) (Fig. 1g,h and Extended Data Fig. 2d,e). In addition, we used a hybridization-capture-based enrichment of mtDNA from scATAC-seq libraries (Fig. 1a and Supplementary Table 2), similar to the recently reported ReDeeM approach<sup>44</sup>, to generate a separate sequencing library consisting of 99.9% mtDNA reads (Fig. 1h), with higher sequencing saturation (mtDNA read duplication rate increased from 38% to 77%) (Extended Data Fig. 2f), evenly spread across the mitochondrial genome (Fig. 1g and Extended Data Fig. 2g). This increased the average mtDNA sequencing depth per cell 1.7-fold, regardless of initial mtDNA coverage, except in those cells with very low initial coverage (<2×), where depth rose 2–3-fold (Fig. 1i,j). Together, this resulted in an 80 × ± 61 × average mtDNA sequencing depth per cell (Extended Data Fig. 2d). A replicate experiment under identical conditions resulted in 6,272 cells (Extended Data Fig. 3a–g), of which 2,965 (43%) could be assigned a unique gRNA identity, with 31.7 × ± 30 × mtDNA coverage after enrichment (Extended Data Fig. 3h,i). A range of QC metrics showed consistent results across both replicates (Extended Data Figs. 1g–j and 3a–d) and UMAP clustering of both datasets displayed an even distribution of cells across all clusters (Extended Data Fig. 4a), confirming the reproducibility of our MitoPerturb-Seq approach in heteroplasmic MEFs.

### MitoPerturb-Seq identifies mtDNA depletion following targeted gene perturbation

We combined both experiments into a single integrated dataset (Extended Data Fig. 4a,b) to maximize the potential for discovery of perturbation-related mitochondrial phenotypes. This yielded 11,990 cells with 60.6 × average mtDNA sequencing depth. The data were filtered to only retain cells with a unique gRNA assignment, resulting in a total of 6,551 cells (Extended Data Fig. 4c and Supplementary Table 3). As expected<sup>45</sup>, target gene knockdown (KD) efficiency for each gRNA was proportional to baseline expression level (Fig. 2a,b and Extended Data Fig. 4d). For example, *Mtfp1* expression could only be detected at low levels (≤2 unique counts) in 4.6% of cells, likely explaining why average mRNA expression did not decrease with any of the three gRNAs targeting this gene.

We next assessed how mtDNA CN and heteroplasmy level were affected by the genetic perturbation in each cell. Relative mtDNA CN was measured as the average single-cell mtDNA coverage from ATAC-seq reads<sup>2,46</sup>. Per-cell heteroplasmy level was quantified as the proportion of all reads matching the m.5024C>T mutant mtDNA molecule. The precision of per-cell heteroplasmy measurements was enhanced by combining reads across seven heteroplasmic single-nucleotide variants (SNVs) (Extended Data Fig. 4e–g), previously shown to be in *cis* or *trans* with the pathogenic m.5024C>T allele<sup>47</sup>, which we confirmed by calculating per-cell correlation of SNV heteroplasmy levels (Extended Data Fig. 4h,i) and long-read





**Fig. 2 | MitoPerturb-Seq identifies mtDNA depletion following targeted gene perturbation.** **a, b**, Scaled transcript levels (**a**) and per-cell log-transformed transcript counts (**b**) of candidate target genes across all perturbation groups. Expression values normalized on a per-column basis in **a**; cells with zero counts are not plotted in **b**. Two-tailed  $t$ -tests were used to evaluate statistical differences in target transcript expression in the relevant perturbation group versus NT gRNA group. \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\*\* $P < 0.0001$ . **c–e**, UMAP shaded according to mtDNA coverage (**c**) or heteroplasmy level (**d**) and following guided clustering (**e**). Cluster 4 is highlighted by the yellow outline. **f–i**, UMAP with cells assigned to NT (**f**; black), *Tfam* (**g**; red), *Opa1* (**h**; blue) and *Polg* (**i**; green) gRNAs highlighted. Pie charts indicate the distribution of perturbation groups in cluster 4 (indicated by yellow outline and dark gray points) or all other clusters. Chi-squared tests were used to evaluate statistical differences. Adjusted  $P$  values for all perturbation groups are shown in Extended Data Fig. 5b. **j, k**, Per-cell mtDNA coverage (**j**) and heteroplasmy calls after read-depth filtering (**k**), for each perturbation group

including controls. Each violin represents the combined data from all three gRNAs targeting the corresponding gene. Only cells with a combined read depth of  $\geq 20$  at the seven heteroplasmic SNVs are shown in **k**. Pairwise  $t$ -tests with multiple testing correction (Bonferroni) were used to evaluate statistical differences.  $P$  values are specified in the figure, with all  $P$  values  $< 0.05$  shown. Horizontal lines indicate the median (**j**) or mean (**k**). **l, m**, Variance in heteroplasmy levels for each perturbation group, using only cells with >20 mtDNA read depth (**l**) or without read-depth filtering (**m**). Statistical significance was assessed using Brown-Forsythe tests (one-way analysis of variance (ANOVA)) against NT control group, with Bonferroni–Holm multiple testing correction. **n**, Heteroplasmy variance distributions from simulations modeling the stochastic sampling bottleneck. Control group heteroplasmy were downsampled to the read depths of *Opa1*-KD, *Polg*-KD and *Tfam*-KD groups ( $n = 5,000$  simulations). Dashed lines indicate observed variances (colored) and the control group baseline (gray). Empirical two-tailed  $P$  values (Bonferroni–Holm-corrected) were determined.

sequencing (Extended Data Fig. 4j,k). Per-cell heteroplasmy calls before and after mtDNA enrichment were highly correlated for all alleles (Extended Data Fig. 4f). When we visualized mtDNA coverage (Fig. 2c) and heteroplasmy levels (Fig. 2d) on a clustered UMAP plot (Fig. 2e), we noted that one cluster had markedly lower per-cell mtDNA content than the rest. This cluster had a distinct gene expression and accessibility profile compared to the other clusters and was enriched for gRNAs targeting *Tfam*, *Opa1* and *Polg*, with all other gRNA assignments under-represented (Fig. 2f–i and Extended Data Fig. 5a,b). As expected<sup>29,30,34</sup>, measuring mtDNA CN in each perturbation group confirmed that KD of *Tfam*, *Opa1* and *Polg* caused a reduction in mean per-cell mtDNA read depth, with no concurrent change in nuclear ATAC-seq fragment counts (Fig. 2j and Extended Data Fig. 5c,d). This was accompanied by a decrease in mtDNA transcripts (Extended Data Fig. 5e). Each of the three gRNAs targeting these genes caused comparable mtDNA CN reduction, apart from one of the *Polg* gRNAs (*Polg-6*) that only slightly decreased *Polg* expression levels (Extended Data Fig. 4d; fold change = 0.97,  $P = 0.86$ ) and, thus, had no impact on mtDNA CN (Extended Data Fig. 5d). We, therefore, removed this gRNA group from downstream analyses. None of the perturbations affected mean single-cell heteroplasmy levels, (Fig. 2k and Extended Data Fig. 5f–h), possibly reflecting limited power to detect changes of <5% (Extended Data Fig. 5i).

Although there was no difference in mean heteroplasmy, we observed greater heteroplasmy variance in *Tfam*-KD cells compared to the NT gRNA group ( $P = 0.046$ , adjusted  $P = 0.87$ ; Fig. 2l). Extending this analysis to also include cells with lower read depths (<20 across all heteroplasmic SNVs), resulted in a significantly higher heteroplasmy variance in *Tfam*-KD (adjusted  $P = 3.31 \times 10^{-22}$ ), *Opa1*-KD (adjusted  $P = 6.76 \times 10^{-3}$ ) and *Polg*-KD (adjusted  $P = 1.78 \times 10^{-8}$ ) groups than in the NT group (Fig. 2m). To determine whether this increased variance was driven by stochastic sampling during mtDNA CN depletion, we modeled the expected variance using binomial sampling of control cells at the depths observed in each perturbation group, collectively spanning a wide range of mtDNA CN. Observed variances for each perturbation group were consistent with their corresponding expected variances, including for those exhibiting significant mtDNA CN reduction (adjusted  $P = 1$  for *Tfam* and *Polg* KD, adjusted  $P = 0.308$  for *Opa1* KD; Fig. 2n and Extended Data Fig. 5j). We also modeled the relationship between read depth and heteroplasmy variance on the basis of the population of cells from groups without mtDNA depth reduction (that is, all cells excluding *Tfam*, *Polg* and *Opa1* KD) (Extended Data Fig. 5k). Comparing this to the heteroplasmy variance in *Tfam*, *Polg* or *Opa1* KD showed similar depth-dependent trajectories in the nondepleted and the mtDNA-depleted populations (Extended Data Fig. 5k). Together, these results indicate that increased heteroplasmy variance may be explained by a perturbation-mediated genetic bottleneck through a reduction in mtDNA CN and read depth. Thus, in addition to technically

validating MitoPerturb-Seq, our data indicate that mtDNA CN can be manipulated to modulate the range of heteroplasmy levels within single cells, potentially influencing the number of cells above or below the threshold required to cause an OXPHOS biochemical defect.

### mtDNA depletion affects nuclear gene expression

A key feature of MitoPerturb-Seq is the ability to study transcriptional responses to mtDNA depletion at the single-cell level in an isogenic nuclear background. A priori, the mechanisms linking *Polg*, *Tfam* and *Opa1* KD with mtDNA depletion are likely to differ. *Polg* encodes the DNA polymerase responsible for mtDNA replication<sup>29</sup>, *Tfam* encodes a protein essential for mtDNA transcription and compaction<sup>48</sup> and *Opa1* encodes a GTPase thought to have an indirect effect on mtDNA maintenance through regulation of mitochondrial membrane dynamics<sup>34</sup>. We, therefore, asked whether these pleiotropic effects had an impact on the downstream transcriptional response to mtDNA CN depletion. We identified 203 *Tfam*, 118 *Opa1* and 13 *Polg* differentially expressed genes (DEGs; log<sub>2</sub> fold change > 0.25, adjusted  $P < 0.05$ ), of which 107/215 (50%) were shared between at least two of the three groups and 12/215 (5.5%, all mtDNA-encoded) were shared among all three groups (Extended Data Fig. 6a and Supplementary Table 4). Interestingly, while mtDNA CN depletion (Fig. 2j and Extended Data Fig. 6b) and the reduction in mtDNA gene expression (Fig. 3a) caused by *Tfam* KD were more severe than upon *Opa1* or *Polg* KD, the nuclear gene expression profile of *Opa1*-KD cells was more similar to *Tfam*-KD than to *Polg*-KD cells (Fig. 3a,b and Extended Data Fig. 6b). The differences in nuclear transcriptomic response (nuclear first principal component (PC1)) were more prominent when we controlled for absolute mtDNA CN at the single-cell level (Fig. 3c), indicating that, at least for *Opa1*-KD cells, mtDNA depletion itself was only partially responsible for the nuclear transcriptomic response.

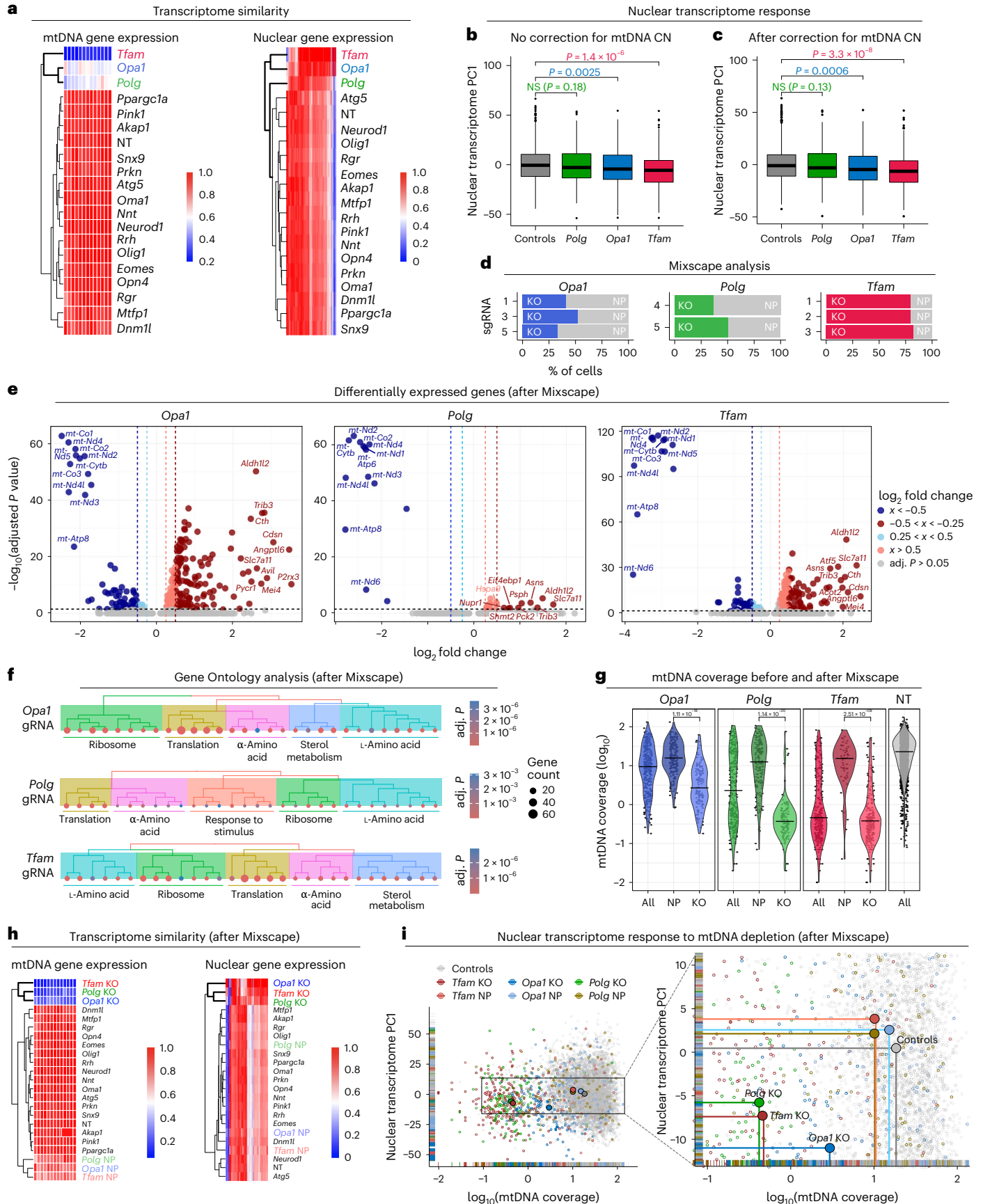
To focus our analysis on the most severely perturbed cells, we used Mixscape<sup>49</sup> to define the difference between severely perturbed (knock-out, KO) and less perturbed or nonperturbed (NP) cells within each gRNA group in an unbiased way (Fig. 3d and Extended Data Fig. 6c,d). Mixscape retained more *Tfam*-KO cells (219 out of 270, 81%) than *Opa1*-KO (124 out of 308, 40%) or *Polg*-KO (118 out of 286, 41%) cells, allowing the identification of 261 DEGs in *Opa1*-KO, 250 DEGs in *Tfam*-KO and 46 DEGs in *Polg*-KO cells (Fig. 3e and Supplementary Table 5). In total, 42 out of 325 (13%) DEGs were shared among all three KO groups and 190 out of 325 (58%) DEGs were shared between at least two of the three KO groups (Extended Data Fig. 6e). Gene Ontology (GO) analysis of DEGs showed a similar response in all three groups, with an upregulation of mitochondrial integrated stress response (mtISR) pathways, including genes involved in cytoplasmic protein translation, ribosome biogenesis and amino acid metabolism (Fig. 3f and Supplementary Table 6). Sterol metabolism genes were selectively

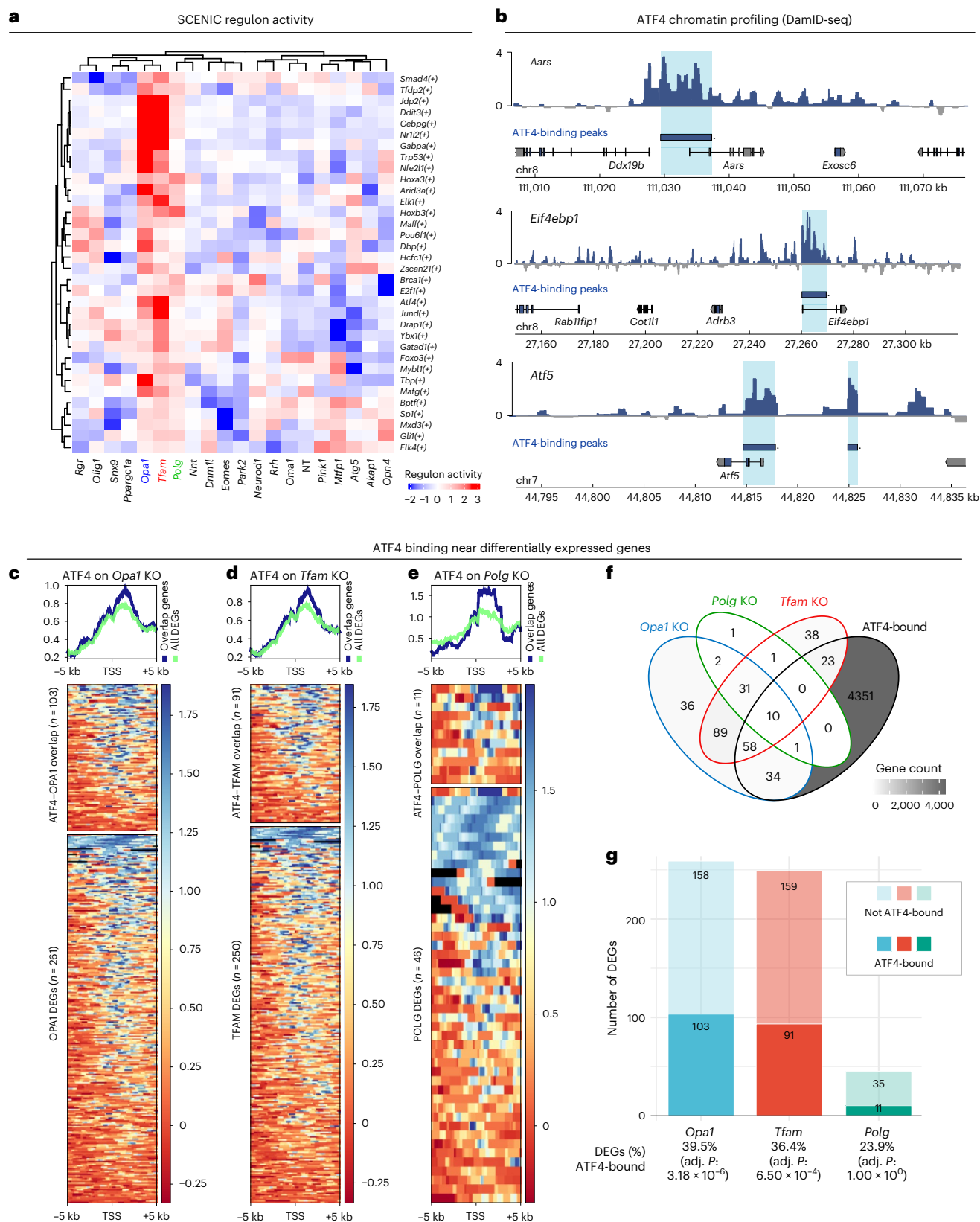
**Fig. 3 | mtDNA depletion affects nuclear gene expression.** **a**, Average gene expression of mtDNA (left) and nuclear (right) genes. Genes with zero expression were removed, filtered for adjusted  $P < 0.05$  and scaled to the gene with highest expression. **b,c**, Distribution of PC1 of nuclear transcriptome across cells in each indicated perturbation group before (**b**) and after (**c**) correction for per-cell mtDNA coverage. A two-sided Wilcoxon signed-rank test with Benjamini–Hochberg correction for multiple comparisons was used to test for statistical differences. NS, not significant. Cell numbers per perturbation group are provided in Supplementary Table 3. Boxes encompass the 25th–75th percentiles, the center line indicates median and whiskers extend to the most extreme data points within 1.5× the interquartile range from the hinges; data points beyond this range are plotted as individual outliers. **d**, Percentage of cells assigned one of the three gRNAs targeting *Tfam* (red), *Opa1* (blue) and *Polg* (green) designated as KO following Mixscape analysis. The remaining cells were assigned as NP (gray). **e**, DEGs (adjusted  $P < 0.05$ ) with a positive or negative log<sub>2</sub> fold change of >0.25 (light colors) and >0.5 (dark colors) in *Opa1*-KO, *Polg*-KO and *Tfam*-KO cells compared to controls. The top ten

upregulated and downregulated genes for each comparison are labeled. A two-sided Wilcoxon rank-sum test with Bonferroni correction for multiple comparisons was used to test for statistical differences. The full list of DEGs is provided in Supplementary Table 5. **f**, Significantly enriched GO terms, grouped by biological process, identified in the *Opa1*-KO, *Polg*-KO and *Tfam*-KO groups. A one-sided Fisher’s exact test with Benjamini–Hochberg correction for multiple comparisons was used to test for statistical differences. The full GO term analysis is provided in Supplementary Table 6. **g**, Distribution of mtDNA coverage (log<sub>10</sub> transformed) in control cells and the *Opa1*, *Polg* and *Tfam* perturbed cells before and after assignment to KO and NP groups by Mixscape analysis. KO versus NP comparisons were conducted using pairwise two-sided Welch’s *t*-tests with multiple testing correction (Bonferroni). **h**, Average gene expression of mtDNA (left) and nuclear (right) genes, following Mixscape classification. As in **a**, genes were filtered for adjusted  $P < 0.05$  and normalized to the most highly expressed gene. **i**, Strength of nuclear PC1 in all cells plotted against mtDNA coverage (log<sub>10</sub> transformed), with *Opa1*, *Polg* and *Tfam* perturbation groups highlighted, separated by KO and NP cells after Mixscape.

depleted only in *Tfam*-KO and *Opa1*-KO but not *Polg*-KO cells (Fig. 3f). Enrichment of genes involved in cholesterol homeostasis has been reported in *Tfam*-KO mouse alveolar macrophages<sup>50</sup>, further confirming the physiological relevance of our MEF-based screening approach.

However, interferon response genes, previously found to be activated upon *Polg* or *Tfam* deficiency<sup>51–53</sup> or in heteroplasmic mice<sup>34</sup> were not differentially expressed (Extended Data Fig. 6f). In the remaining perturbation groups, Mixscape analysis only detected KO cells in the *Atg5*





**Fig. 4 | ATF4 only partially contributes to the response to mtDNA depletion.**

**a**, Regulon activity (AUC > 0) for three selected perturbation groups (*Opa1*, *Polg* and *Tfam*) following SCENIC analysis. **b**, ATF4 DamID-seq profiles in heteroplasmic MEFs across three ATF4 target genes (*Aars*, *Eif4ebp1* and *Atf5*). ATF4-binding peaks are shaded light blue. **c–e**, Metagene plots (top) and heat maps (bottom) showing ATF4 DamID-seq signal across TSSs of all genes differentially expressed in *Opa1* (**c**), *Tfam* (**d**) or *Polg* (**e**) perturbation

groups (bottom, green), or only those DEGs associated (top, blue) with nearby ATF4-binding sites. **f**, Overlap between ATF4 DamID-seq target genes and DEGs in each of the indicated perturbation groups. **g**, Number of DEGs (total bar height) and the number and percentage bound by ATF4 DamID-seq (dark shading). A one-sided Fisher's Exact test with Bonferroni correction was used to test significance differences across all unique genes detected in gRNA-containing cells.

gRNA group, indicating that perturbation of most target genes could not elicit a severe nuclear transcriptomic response in our conditions (Extended Data Fig. 6d). *Atg5* KO mainly affected pathways related to lysosome activity and cytoplasmic protein translation and stability (Extended Data Fig. 6g,h). All DEGs and pathways are provided in Supplementary Tables 5 and 6.

In the *Opa1*, *Polg* and *Tfam* gRNA groups, the KO cells retained by Mixscape had significantly stronger mtDNA depletion than NP cells, where mtDNA levels were similar to controls (Fig. 3g). There were no significant differences in mean heteroplasmy between KO and NP or control groups but we did observe significantly higher heteroplasmy variance in all three KO groups compared to NP cells (Extended Data Fig. 6i). mtDNA depletion was less marked in the *Opa1*-KO group (Fig. 3g), rendering the strongly perturbed *Polg*-KO cells more similar to *Tfam*-KO than to *Opa1*-KO cells (Fig. 3h,i). This is in keeping with the nuclear response to *Polg* or *Tfam* perturbation being primarily caused by mtDNA depletion, while other factors likely contribute to the response in *Opa1*-KD/KO cells. Focusing our analysis on mtDNA-encoded gene expression (mtRNA), *Opa1* KD caused a more severe decrease in mtRNA for the same level of mtDNA depletion than *Polg* KD (Extended Data Fig. 6j–m), supporting a role for *Opa1* in regulation of mtRNA transcription or stability, independent of its effect on mtDNA content, as previously suggested<sup>55</sup>.

We independently validated these findings by performing bulk RNA-seq analysis and droplet digital PCR (ddPCR)-based mtDNA CN measurements after single-gRNA transduction using a second m.5024C>T clone (clone 8, 45% heteroplasmy; Extended Data Fig. 1c). *Tfam* KD resulted in a rapid and severe mtDNA CN reduction 3 days after transduction (Extended Data Fig. 7a), implicating active mtDNA degradation, which is in keeping with TFAM having a crucial role in nucleoid compaction and protecting mtDNA from lysosomal or mitochondrial nuclease-mediated degradation<sup>56–58</sup>. Nevertheless, the transcriptomic response to *Opa1* KD 6 days after gRNA transduction was more severe than to *Tfam* KD (82 and 33 DEGs, respectively, at false discovery rate (FDR) < 0.05; overlap of 27 genes) (Extended Data Fig. 7b,c). Together, these data indicate that, although *Opa1*, *Polg* and *Tfam* perturbation all activate a similar set of mtSR genes, retrograde mitochondrial–nuclear signaling may occur through different perturbation-specific pathways, with cells responding to *Opa1* KD independent of mtDNA depletion, in line with OPA1's upstream role in mitochondrial homeostasis and membrane dynamics<sup>59</sup>. The level of mtDNA depletion (more severe in *Tfam* KO than *Opa1* KO) and strength of the nuclear transcriptional response (stronger in *Opa1* KO than *Tfam* KO) was consistent between both MEF clones, each carrying different heteroplasmy levels, suggesting that, at least in this context, the cellular response to these perturbations was independent of the heteroplasmy level and occurred well below the biochemical threshold of the m.5024C>T mutation<sup>25</sup>. Critically, simultaneously probing these mechanisms in nuclear isogenic cells, grown together and exposed to the same environment, allowed us to avoid additional and potentially unknown confounders impacting cellular stress responses.

### ATF4 only partially contributes to the response to mtDNA depletion

To define the factors driving the transcriptional response to mtDNA depletion in *Tfam*-KD, *Opa1*-KD and *Polg*-KD cells, we performed single-cell regulatory network inference and clustering (SCENIC) analysis<sup>60</sup>. SCENIC allows inference of gene regulatory networks (regulons) that show differential activity in each of our perturbation groups (Extended Data Fig. 8a and Supplementary Table 7). We identified 34 regulons that were differentially active in all three KD groups (Fig. 4a), including regulons driven by the transcription factors ATF4, DDIT3, CEBPG or JUND, known to be involved in the cellular response to stress in cell culture<sup>61</sup>, and in *Tfam*-KO or *Opa1*-KO mice<sup>62,63</sup>. Comparing our dataset to a recent systematic review of putative ATF4

target genes<sup>61</sup>, 24 out of 40 (60%) high-confidence ATF4-responsive protein-coding genes were also differentially expressed in at least one of our *Tfam*-KO, *Opa1*-KO and *Polg*-KO groups (Extended Data Fig. 8b). To independently validate these findings, we used DamID-seq<sup>64,65</sup> for genome-wide profiling of ATF4 chromatin-binding sites in our heteroplasmic MEFs (Fig. 4b and Extended Data Fig. 8c). This identified 5,789 ATF4-binding peaks near 4,477 genes (Supplementary Tables 8 and 9 and Extended Data Fig. 8d). The top de novo motif identified by motif enrichment analysis had a high similarity to the canonical AP1 consensus motif (FOS::JUN, JASPAR MA0099.1; weighted Pearson's correlation coefficient = 0.986,  $P = 4.98 \times 10^{-5}$ ). This was confirmed by an analysis of known motifs, where the most highly enriched motifs were AP1 family members that bind the same 5'-TGA[GC]TCA-3' consensus, suggesting ATF, JUN and FOS co-occupancy of these ATF4 DamID-seq binding sites<sup>66</sup>. The canonical ATF4 (bZIP) motif in the HOMER database, 5'-MTGATGCAAT-3', characteristic of ATF4–CEBPG heterodimer binding<sup>67</sup>, was also highly enriched (rank 11,  $P = 1 \times 10^{-17}$ ) (Extended Data Fig. 8e), validating DamID-seq peaks as genuine ATF4-binding sites. Of the 325 DEGs in *Tfam*-KO, *Polg*-KO and *Opa1*-KO cells, 126 (38.8%) were near ATF4-binding sites in heteroplasmic MEFs, with binding mostly near transcription start sites (TSSs) (Fig. 4c–e). However, at our current thresholds (FDR < 0.05;  $\log_2$ -transformed fold change > 0.25), the majority of DEGs upon *Opa1* KO (158 out of 261, 60.5%), *Tfam* KO (159 out of 250, 63.6%) or *Polg* KO (35 out of 46, 76.1%) were not ATF4 bound (irreproducible discovery rate (IDR) < 0.01; Fig. 4f,g), indicating an important but not exclusive role for ATF4 in responding to mtDNA-related stress. Critically, a range of other non-ATF4 transcription factor regulons were found upon SCENIC analysis (Fig. 4a). These are provided for further exploration in Supplementary Table 7 and are likely to coregulate the response to perturbation of these mtDNA maintenance genes or to mtDNA depletion in heteroplasmic cells.

### mtDNA depletion delays cell-cycle progression across all stages

Having identified the transcriptomic changes caused by mtDNA depletion, we next asked how this could impact cellular physiology and behavior. Cell proliferation is a sensitive indicator of mitochondrial activity, with OXPHOS dysfunction previously shown to cause cell-cycle slowing primarily at the G1/S transition<sup>68–70</sup>. We performed cell-cycle stage annotation on our MitoPerturb-Seq dataset using Seurat<sup>71</sup> (Fig. 5a) and continuous cell-cycle pseudotime analysis<sup>72</sup> (Extended Data Fig. 9a). UMAP analysis of heteroplasmic MEFs before cell-cycle correction showed cells mainly segregating on the basis of cell-cycle stage (Fig. 5a). To our surprise, we saw no significant differences in the proportion of cells in each cell-cycle phase between the perturbation groups (Fig. 5b). After Mixscape analysis, only *Opa1*-KO but not *Tfam*-KO or *Polg*-KO cells showed a slight ( $P = 0.034$ , adjusted  $P = 0.10$ ) increase in the proportion of cells in G1 phase compared to control cells (Extended Data Fig. 9b). This indicates that, in contrast to previous reports describing cell-cycle slowing specifically at the G1/S transition upon OXPHOS dysfunction<sup>68–70</sup>, there was no selective delay in G1/S progression in *Tfam*-KO and *Polg*-KO cells, despite severe mtDNA depletion.

To validate these findings, we generated transgenic cell lines expressing a fluorescent ubiquitination-based cell-cycle indicator (FUCCI)<sup>73</sup> (Extended Data Fig. 9c–e and Supplementary Video 1) and analyzed cell-cycle duration in WT and *TFAM*-KO FUCCI-HeLa cells (Extended Data Fig. 9f,g), and in FUCCI transgenic cybrid cells carrying a heteroplasmic large-scale mtDNA deletion encompassing the major arc<sup>74</sup> (FUCCI-DeltaH2.1; Extended Data Fig. 9h). *TFAM*-KO FUCCI-HeLa cells had severe mtDNA depletion (Extended Data Fig. 9g) and displayed an increase in overall cell-cycle length compared to WT FUCCI-HeLa cells, caused by increased duration of both G1 and S/G2 phases (Fig. 5c). These findings in *TFAM*-KO FUCCI-HeLa cells are consistent with our MitoPerturb-Seq data from heteroplasmic *Tfam*-KO (and *Polg*-KO)



**Fig. 5 | mtDNA depletion delays cell-cycle progression and relaxed replication.** **a**, UMAP clustering based on RNA-seq before regression of cell-cycle heterogeneity. Cells are colored by cell-cycle phase. **b**, Proportion of cells in each perturbation group assigned to each cell-cycle stage. Two-sided chi-squared tests with multiple testing correction (Benjamini–Hochberg) were used to test for significant differences. **c**, Cell-cycle phase duration times in FUCCI-expressing WT and *TFAM*-KO HeLa cells. Phase duration was calculated across one full cycle (G1 and S/G2) for 20 cells per condition. Comparisons were conducted using two-tailed unpaired *t*-tests.  $^{**}P < 0.01$  and  $^{****}P < 0.0001$ . Data are presented as the mean  $\pm$  s.d. **d**, Cell-cycle phase duration in  $\Delta$ H2.1 mtDNA deletion cybrid clones stably expressing PIP-FUCCI, cultured in high-glucose or low-glucose medium and imaged every 20 min over a period of 72 h. Phase duration was calculated across one full cycle (G1, S and G2/M) for 20 cells (early,

first 24 h; late, after 24 h). Data are presented as the mean  $\pm$  s.d. One-way ANOVAs were performed per cell-cycle phase for each clone with Tukey’s post hoc test.  $^{****}P < 0.0001$ . All significant pairwise comparisons are shown. N/A, no observed cells progressed to this phase during the experiment. **e**, mtDNA coverage across Seurat cell-cycle phases. A one-way ANOVA with Tukey’s post hoc test was used to test for significant differences.  $^{****}P < 0.0001$ . **f**, mtDNA coverage across cell-cycle pseudotime with cells ranked on the basis of cell cycle pseudotime (tricycle  $\pi$ ) score. The fit line is a smoothed locally estimated scatterplot smoothing regression (span = 0.2). Shaded regions around the fit line indicate the 95% CI. **g**, mtDNA CN in WT PIP-FUCCI-HeLa cells flow-sorted by cell-cycle phase (sorting strategy in Extended Data Fig. 9e). Cell numbers per group are indicated on the graph. Data are presented as the mean  $\pm$  s.d. A one-way ANOVA with Tukey’s post hoc test was used to test for significant differences.  $^{***}P < 0.001$ .

MEFs, where a consistent delay across all stages of the cell cycle would not change the proportion of cells in each individual phase. Clones of FUCCI-DeltaH2.1 cells had no mtDNA depletion (Extended Data Fig. 9i) but carried a range of heteroplasmies (Extended Data Fig. 9j), with higher heteroplasmy severely impacting OXPHOS activity (baseline respiration, ATP-linked respiration and maximum respiratory capacity) (Extended Data Fig. 9k). In contrast to *Tfam*-KO cells, FUCCI-DeltaH2.1 clones with low (20%), mid (35%) and high (90%) heteroplasmy all proliferated normally (Extended Data Fig. 9l). However, the proliferation rates of cells with mid and high (but not low) heteroplasmy decreased when cultured in low-glucose medium (Extended Data Fig. 9m). This was mainly caused by an increase in G1 phase duration, with many cells with mid and high heteroplasmy not progressing from G1 to S phase when cultured in low-glucose medium for > 48 h (Fig. 5d). Together, these data indicate differential requirements and sensing of mtDNA abundance versus OXPHOS activity to sustain cell-cycle progression, with the G1/S transition being most sensitive to heteroplasmy and OXPHOS deficiency. However, when confronted with severe mtDNA depletion, as in *Tfam*-KO cells, cells slow cell-cycle progression across all phases.

### Relaxed replication of mtDNA across the cell cycle

During each cell cycle, the nuclear DNA is replicated only once, in a strictly regulated process during the S phase. In contrast, the mtDNA is thought to undergo relaxed replication, independent of the cell cycle<sup>75,76</sup>. Previous studies found conflicting evidence of whether mtDNA replication rates differ across cell-cycle stages but these lacked temporal resolution, relied on indirect mtDNA CN measurements or were based on cell-cycle synchronization, which directly impacts mitochondrial and cellular metabolism<sup>77–80</sup>. We reasoned that combined RNA and mtDNA profiling from a homogeneous population of proliferating cells, as in our MitoPerturb-Seq dataset, could provide an alternative, unbiased approach to characterize cell-cycle-related mtDNA dynamics at the single-cell level. Analysis of mtDNA scATAC-seq read counts across the cell cycle in MEFs showed a progressive increase in mtDNA coverage across G1/S and S/G2/M stages (Fig. 5e), as well as across continuous cell-cycle pseudotime<sup>72</sup> (Fig. 5f), in keeping with mtDNA replication being fully ‘relaxed’ (ref. 75). Crucially, this also further validates our approach to use ATAC-seq read coverage as a proxy for mtDNA CN, allowing detection of subtle changes in mtDNA content within single cells. Nuclear ATAC-seq read count also increased across cell-cycle phases but most strongly during the S/G2 transition (Extended Data Fig. 10a,b). Heteroplasmy levels remained constant between cell-cycle phases (Extended Data Fig. 10c) and over cell-cycle pseudotime (Extended Data Fig. 10d). The mtDNA depletion caused by *Tfam*, *Polg* or *Opa1* KO was also present in all cell-cycle stages and across cell-cycle pseudotime (Extended Data Fig. 10e–g). To validate these findings, we measured absolute mtDNA CN by single-cell ddPCR<sup>81</sup> in WT FUCCI-HeLa and FUCCI-HEK293T cell lines (Extended Data Fig. 10h,i). This indeed revealed a linear increase in mtDNA CN, not only across cell-cycle phases (G1, S and G2/M) (Extended Data Fig. 10h) but also across the early, mid and late S phase and between the early and late G2 phase (Fig. 5g and Extended Data Fig. 9e). Interestingly, culturing WT FUCCI-HeLa cells in galactose medium to force respiration through OXPHOS induced an increase in mtDNA CN, whilst maintaining a linear increase across cell-cycle stages (Extended Data Fig. 10i), indicating that cells actively adjust mtDNA CN in response to higher dependence on mitochondrial ATP production. Our findings, thus, validate previous work at high temporal resolution, demonstrating fully relaxed replication of WT and mutant mtDNA molecules across the cell cycle but modulated in response to bioenergetic requirements.

### Discussion

Single-cell mtDNA analysis in healthy humans<sup>17</sup> or in mice and persons with inherited or age-related mtDNA mutations<sup>18,19</sup> has found an extraordinary degree of mosaicism, with virtually all cells across

tissues and organisms containing different abundance and mixtures of WT and mutant mtDNA. Although the phenotypes caused by these mtDNA mutations are inherently mosaic at the cellular level, most studies to date have relied on bulk, tissue-wide analysis, averaging out the single-cell consequences of mtDNA-related heterogeneity and, thus, limiting the sensitivity to detect disease-relevant phenotypes and mechanisms. Technical challenges when perturbing and measuring mtDNA at the single-cell level mean that most studies to date have been correlative and little is known about which nuclear-encoded proteins respond to or regulate changes in mtDNA CN and mutation load in individual cells or cell types.

Building on single-cell CRISPR screening methods<sup>22</sup> and simultaneous scRNA-seq and mtDNA-seq<sup>24,82</sup>, we developed MitoPerturb-Seq for high-throughput unbiased forward genetic screening of factors with a direct causal effect on mtDNA CN and heteroplasmy within individual cells. As proof of concept, we screened a panel of 60 gRNAs, targeting 13 nuclear-encoded genes known to be involved in mtDNA dynamics, in Cas9-transgenic MEFs from mice carrying a heteroplasmic mtDNA tRNA<sup>Ala</sup> mutation corresponding to a human pathogenic variant<sup>25,27</sup>. Perturbation of three of these genes, *Tfam*, *Opa1* and *Polg*, caused strong single-cell mtDNA depletion, with large overlap in the single-cell transcriptomic response to each perturbation, driven by comparable transcription factor regulons. One of these transcription factors, ATF4, has been studied extensively as a central factor mediating the response to mitochondrial dysfunction and cellular stress<sup>83–85</sup>. However, in our data and conditions, only approximately one third of DEGs were directly bound by ATF4, which, together with our regulon analysis, indicates that many other factors are involved. We provide a ranked resource of transcription factors involved in modulating mtDNA CN, including established mtISR genes, such as two small MAF transcription factors, MAFF and MAFG, known to be involved in tissue-specific responses to oxidative stress<sup>86</sup>. Cell-type-specific expression of stress-responsive transcription factors, together with differential chromatin accessibility<sup>87</sup>, is likely to underlie tissue-specific and context-dependent resilience and vulnerability to mitochondrial dysfunction<sup>88</sup>; hence, it will be important to extend our approach to other cell types.

We conducted our experiments in heteroplasmic MEFs and targeted some of the few nuclear-encoded genes that were previously found to regulate heteroplasmy levels in other contexts<sup>10,12,13,28–30,32,33,94</sup>. Nevertheless, we could not detect significant changes in heteroplasmy levels in any of the perturbation groups, with several groups also showing no significant transcriptional response to target gene KO. Although technical considerations, including those inherent to pooled single-cell CRISPR screens<sup>89</sup> (low target gene expression in MEFs, cell-to-cell heteroplasmy variability and short duration of perturbation), may underlie this lack of effect to some extent, many of these genes will have context-dependent and cell-type-specific or organism-specific functions. Future experiments including more cells, different cell types, higher mtDNA read depth and longer timeframes will be required to confidently reveal or exclude genes and processes that modulate heteroplasmy levels, although our findings indicate that even severe gene KD will have very subtle effects at best.

How nuclear and mtDNA replication rates are coordinated during cell growth and proliferation remains poorly understood<sup>76–78</sup>. We provide further evidence for completely relaxed mtDNA replication at the single-cell level, independent of cell-cycle stages, in a range of WT, mtDNA-depleted, homoplasmic and heteroplasmic human and mouse cell lines. Interestingly, cell-cycle slowing in response to mtDNA depletion affects the entire cell cycle, equally delaying all phases. This differs from the G1/S-specific delay in response to high heteroplasmy levels or targeted OXPHOS dysfunction that we and others have observed<sup>68–70</sup> and might indicate the presence of a specific, yet unknown nuclear sensing mechanism to coordinate cell-cycle progression across all cell-cycle stages with mtDNA abundance or replication.

Our targeted library provides proof of principle for completely unbiased discovery through genome-wide CRISPR screening. In addition, PCR-based and hybridization-capture-based targeted enrichment approaches allow selective sequencing of gRNAs and mtDNA, potentially reducing cost and increasing throughput >5-fold. Moreover, MitoPerturb-Seq is uniquely placed to conduct screening in disease-relevant postmitotic cell types, including *in vivo*<sup>90</sup> or in organoids derived from human induced pluripotent stem cells. This is particularly relevant, as mtDNA mutations mostly accumulate in and affect postmitotic cells<sup>4,88</sup>. The recent advent of mtDNA-modifying technologies to engineer mtDNA point mutations<sup>91,92</sup> and deletions<sup>32,46</sup> in any cell or model system means that MitoPerturb-Seq can now be used to compare the response to perturbation in isogenic cells with a range of heteroplasmic and homoplasmic mtDNA defects. We envisage this to be combined with other functional measurements, such as flow-cytometry-based mitochondrial membrane potential measurement<sup>93</sup>, and with more affordable combinatorial indexing approaches, such as SHARE-seq<sup>94</sup>, for high-throughput simultaneous detection of gRNAs, mtDNA CN and heteroplasmy in millions of heterogeneous cells from mosaic cultures, tissues and organisms. Increased throughput will also allow combinatorial screening using dual-gRNA vectors to investigate genetic interactions by probing the effects of perturbing more than one gene simultaneously<sup>95</sup>. In conclusion, MitoPerturb-Seq provides a powerful forward genetic screening approach to discover the biological mechanisms driving age-dependent mtDNA CN reduction or clonal expansion of damaging mtDNA mutations, thereby uncovering novel druggable targets to treat both rare and common mtDNA-related and neurodegenerative disorders.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-026-01779-7>.

### References

- Filigrana, R., Mennuni, M., Alsinà, D. & Larsson, N. G. Mitochondrial DNA copy number in human disease: the more the better? *FEBS Lett.* **595**, 976–1002 (2021).
- Wei, W. et al. Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains. *Acta Neuropathol. Commun.* **5**, 13 (2017).
- Mengel-From, J. et al. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet.* **133**, 1149–1159 (2014).
- van den Ameële, J., Li, A. Y. Z., Ma, H. & Chinnery, P. F. Mitochondrial heteroplasmy beyond the oocyte bottleneck. *Semin. Cell Dev. Biol.* **97**, 156–166 (2020).
- 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).
- Trifunov, S. et al. Clonal expansion of mtDNA deletions: different disease models assessed by digital droplet PCR in single muscle cells. *Sci. Rep.* **8**, 11682 (2018).
- Corral-Debrinski, M. et al. Marked changes in mitochondrial DNA deletion levels in Alzheimer brains. *Genomics* **23**, 471–476 (1994).
- Bender, A. et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat. Genet.* **38**, 515–517 (2006).
- Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* **52**, 342–352 (2020).
- Gupta, R. Nuclear genetic control of mtDNA copy number and heteroplasmy in humans. *Nature* **620**, 839–848 (2023).
- Burr, S. P. Cell lineage-specific mitochondrial resilience during mammalian organogenesis. *Cell* **186**, 1212–1229 (2023).
- Chiang, A. C., McCartney, E., O’Farrell, P. H. & Ma, H. A genome-wide screen reveals that reducing mitochondrial DNA polymerase can promote elimination of deleterious mitochondrial mutations. *Curr. Biol.* **29**, 4330–4336 (2019).
- Lechuga-Vieco, A. V. et al. Cell identity and nucleo-mitochondrial genetic context modulate OXPHOS performance and determine somatic heteroplasmy dynamics. *Sci. Adv.* **6**, eaba5345 (2020).
- Gitschlag, B. L. et al. Homeostatic responses regulate selfish mitochondrial genome dynamics in *C. elegans*. *Cell Metab.* **24**, 91–103 (2016).
- Durham, S. E., Brown, D. T., Turnbull, D. M. & Chinnery, P. F. Progressive depletion of mtDNA in mitochondrial myopathy. *Neurology* **67**, 502–504 (2006).
- Battersby, B. J., Loredó-Ostí, J. C. & Shoubridge, E. A. Nuclear genetic control of mitochondrial DNA segregation. *Nat. Genet.* **33**, 183–186 (2003).
- Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).
- Glynos, A. et al. High-throughput single-cell analysis reveals progressive mitochondrial DNA mosaicism throughout life. *Sci. Adv.* **9**, eadi4038 (2023).
- Korotkevich, E., Conrad, D. N., Gartner, Z. J. & O’Farrell, P. H. Selfish mutations promote age-associated erosion of mtDNA integrity in mammals. *Nat. Commun.* **16**, 5435 (2025).
- 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).
- Walker, M. A. et al. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N. Engl. J. Med.* **383**, 1556–1563 (2020).
- Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
- 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).
- 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).
- Kaupilla, J. H. K. et al. A phenotype-driven approach to generate mouse models with pathogenic mtDNA mutations causing mitochondrial disease. *Cell Rep* **16**, 2980–2990 (2016).
- McFarland, R. et al. The m.5650G>A mitochondrial tRNAAla mutation is pathogenic and causes a phenotype of pure myopathy. *Neuromuscul. Disord.* **18**, 63–67 (2008).
- Ratnaike, T. E. et al. MitoPhen database: a human phenotype ontology-based approach to identify mitochondrial DNA diseases. *Nucleic Acids Res.* **49**, 9686–9695 (2021).
- Zhang, Y. et al. PINK1 inhibits local protein synthesis to limit transmission of deleterious mitochondrial DNA mutations. *Mol. Cell* **73**, 1127–1137 (2019).
- Naviaux, R. K. & Nguyen, K. V. POLG mutations associated with Alpers’ syndrome and mitochondrial DNA depletion. *Ann. Neurol.* **55**, 706–712 (2004).
- Ekstrand, M. I. et al. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum. Mol. Genet.* **13**, 935–944 (2004).
- Kremer, L. S. et al. Tissue-specific responses to TFAM and mtDNA copy number manipulation in prematurely ageing mice. *eLife* **14**, RP104461 (2025).
- Kandul, N. P., Zhang, T., Hay, B. A. & Guo, M. Selective removal of deletion-bearing mitochondrial DNA in heteroplasmic *Drosophila*. *Nat. Commun.* **7**, 13100 (2016).

33. Tábara, L. C. et al. MTFP1 controls mitochondrial fusion to regulate inner membrane quality control and maintain mtDNA levels. *Cell* **187**, 3619–3637 (2024).
34. Elachouri, G. et al. OPA1 links human mitochondrial genome maintenance to mtDNA replication and distribution. *Genome Res.* **21**, 12–20 (2011).
35. Kim, J. Y. et al. Mitochondrial DNA content is decreased in autosomal dominant optic atrophy. *Neurology* **64**, 966–972 (2005).
36. Matheoud, D. et al. Parkinson's disease-related proteins PINK1 and Parkin repress mitochondrial antigen presentation. *Cell* **166**, 314–327 (2016).
37. Zecchini, V. et al. Fumarate induces vesicular release of mtDNA to drive innate immunity. *Nature* **615**, 499–506 (2023).
38. Sen, A. et al. Mitochondrial membrane proteins and VPS35 orchestrate selective removal of mtDNA. *Nat. Commun.* **13**, 6704 (2022).
39. Ahier, A. et al. PINK1 and Parkin shape the organism-wide distribution of a deleterious mitochondrial genome. *Cell Rep.* **35**, 109203 (2021).
40. Dillon, L. M. et al. Increased mitochondrial biogenesis in muscle improves aging phenotypes in the mtDNA mutator mouse. *Hum. Mol. Genet.* **21**, 2288–2297 (2012).
41. Suen, D.-F., Narendra, D. P., Tanaka, A., Manfredi, G. & Youle, R. J. Parkin overexpression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. *Proc. Natl Acad. Sci. USA* **107**, 11835–11840 (2010).
42. Schraivogel, D. et al. Targeted Perturb-Seq enables genome-scale genetic screens in single cells. *Nat. Methods* **17**, 629–635 (2020).
43. Griffin, H. R. et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet. Med.* **16**, 962–971 (2014).
44. Weng, C. et al. Deciphering cell states and genealogies of human haematopoiesis. *Nature* **627**, 389–398 (2024).
45. Yang, L. et al. scMAGECK links genotypes with multiple phenotypes in single-cell CRISPR screens. *Genome Biol.* **21**, 19 (2020).
46. Fu, Y. et al. Engineering mtDNA deletions by reconstituting end joining in human mitochondria. *Cell* **188**, 2778–2793 (2025).
47. Zhang, H. et al. Mitochondrial DNA heteroplasmy is modulated during oocyte development propagating mutation transmission. *Sci. Adv.* **7**, eabi5657 (2021).
48. Kaufman, B. A. et al. The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol. Biol. Cell* **18**, 3225–3236 (2007).
49. Papalexis, E. et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021).
50. Gao, X. et al. TFAM-dependent mitochondrial metabolism is required for alveolar macrophage maintenance and homeostasis. *J. Immunol.* **208**, 1456–1466 (2022).
51. Kang, Y. et al. Ancestral allele of DNA polymerase gamma modifies antiviral tolerance. *Nature* **628**, 844–853 (2024).
52. Lepelley, A., Wai, T. & Crow, Y. J. Mitochondrial nucleic acid as a driver of pathogenic type I interferon induction in Mendelian disease. *Front. Immunol.* **12**, 729763 (2021).
53. West, A. P. et al. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–557 (2015).
54. Marques, E. et al. An inherited mitochondrial DNA mutation remodels inflammatory cytokine responses in macrophages and in vivo in mice. *Nat. Commun.* **16**, 10222 (2025).
55. Yang, L. et al. OPA1-Exon4b binds to mtDNA D-loop for transcriptional and metabolic modulation, independent of mitochondrial fusion. *Front. Cell Dev. Biol.* **8**, 180 (2020).
56. Alam, T. I. et al. Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Res.* **31**, 1640–1645 (2003).
57. Kukat, C. et al. Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc. Natl Acad. Sci. USA* **112**, 11288–11293 (2015).
58. Newman, L. E. et al. Mitochondrial DNA replication stress triggers a pro-inflammatory endosomal pathway of nucleoid disposal. *Nat. Cell Biol.* **26**, 194–206 (2024).
59. Tilokani, L., Nagashima, S., Paupe, V. & Prudent, J. Mitochondrial dynamics: overview of molecular mechanisms. *Essays Biochem.* **62**, 341–360 (2018).
60. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
61. Neill, G. & Masson, G. R. A stay of execution: ATF4 regulation and potential outcomes for the integrated stress response. *Front. Mol. Neurosci.* **16**, 1112253 (2023).
62. Kuhl, I. et al. Transcriptomic and proteomic landscape of mitochondrial dysfunction reveals secondary coenzyme Q deficiency in mammals. *eLife* **6**, e30952 (2017).
63. Hinton, A. Jr. et al. ATF4-dependent increase in mitochondrial–endoplasmic reticulum tethering following OPA1 deletion in skeletal muscle. *J. Cell Physiol.* **239**, e31204 (2024).
64. van den Aamele, J., Krautz, R. & Brand, A. H. TaDa! Analysing cell type-specific chromatin in vivo with targeted DamID. *Curr. Opin. Neurobiol.* **56**, 160–166 (2019).
65. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
66. van Dam, H. & Castellazzi, M. Distinct roles of Jun:Fos and Jun:ATF dimers in oncogenesis. *Oncogene* **20**, 2453–2464 (2001).
67. Rodríguez-Martínez, J. A., Reinke, A. W., Bhimsaria, D., Keating, A. E. & Ansari, A. Z. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *eLife* **6**, e19272 (2017).
68. Mitra, K., Wunder, C., Roysam, B., Lin, G. & Lippincott-Schwartz, J. A hyperfused mitochondrial state achieved at G1–S regulates cyclin E buildup and entry into S phase. *Proc. Natl Acad. Sci. USA* **106**, 11960–11965 (2009).
69. Owusu-Ansah, E., Yavari, A., Mandal, S. & Banerjee, U. Distinct mitochondrial retrograde signals control the G1–S cell cycle checkpoint. *Nat. Genet.* **40**, 356–361 (2008).
70. van den Aamele, J. & Brand, A. H. Neural stem cell temporal patterning and brain tumour growth rely on oxidative phosphorylation. *eLife* **8**, e47887 (2019).
71. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
72. Zheng, S. C. et al. Universal prediction of cell-cycle position using transfer learning. *Genome Biol.* **23**, 41 (2022).
73. Grant, G. D., Kedziora, K. M., Limas, J. C., Cook, J. G. & Purvis, J. E. Accurate delineation of cell cycle phase transitions in living cells with PIP–FUCCI. *Cell Cycle* **17**, 2496–2516 (2018).
74. Diaz, F. et al. Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. *Nucleic Acids Res.* **30**, 4626–4633 (2002).
75. Birky, C. W. Jr. Relaxed and stringent genomes: why cytoplasmic genes don't obey Mendel's laws. *J. Hered.* **85**, 355–365 (1994).
76. Chinnery, P. F. & Samuels, D. C. Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.* **64**, 1158–1165 (1999).
77. Sasaki, T., Sato, Y., Higashiyama, T. & Sasaki, N. Live imaging reveals the dynamics and regulation of mitochondrial nucleoids during the cell cycle in Fucci2–HeLa cells. *Sci. Rep.* **7**, 11257 (2017).
78. Seel, A. et al. Regulation with cell size ensures mitochondrial DNA homeostasis during cell growth. *Nat. Struct. Mol. Biol.* **30**, 1549–1560 (2023).
79. Chatre, L. & Ricchetti, M. Prevalent coordination of mitochondrial DNA transcription and initiation of replication with the cell cycle. *Nucleic Acids Res.* **41**, 3068–3078 (2013).

80. Bogenhagen, D. & Clayton, D. A. Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. *Cell* **11**, 719–727 (1977).
81. Burr, S. P. & Chinnery, P. F. Measuring single-cell mitochondrial DNA copy number and heteroplasmy using digital droplet polymerase chain reaction. *J. Vis. Exp.* (185), e63870 (2022).
82. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
83. Bao, X. R. Mitochondrial dysfunction remodels one-carbon metabolism in human cells. *eLife* **5**, e10575 (2016).
84. Forsstrom, S. et al. Fibroblast growth factor 21 drives dynamics of local and systemic stress responses in mitochondrial myopathy with mtDNA deletions. *Cell Metab.* **30**, 1040–1054 (2019).
85. Quirós, P. M. et al. Multi-omics analysis identifies ATF4 as a key regulator of the mitochondrial stress response in mammals. *J. Cell Biol.* **216**, 2027–2045 (2017).
86. Motohashi, H., Katsuoka, F., Shavit, J. A., Engel, J. D. & Yamamoto, M. Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103**, 865–875 (2000).
87. van den Ameele, J. et al. Reduced chromatin accessibility correlates with resistance to Notch activation. *Nat. Commun.* **13**, 2210 (2022).
88. Burr, S. P. & Chinnery, P. F. Origins of tissue and cell-type specificity in mitochondrial DNA (mtDNA) disease. *Hum. Mol. Genet.* **33**, R3–R11 (2024).
89. Schraivogel, D., Steinmetz, L. M. & Parts, L. Pooled genome-scale CRISPR screens in single cells. *Annu. Rev. Genet.* **57**, 223–244 (2023).
90. Jin, X. et al. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, eaaz6063 (2020).
91. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
92. Silva-Pinheiro, P. et al. A library of base editors for the precise ablation of all protein-coding genes in the mouse mitochondrial genome. *Nat. Biomed. Eng.* **7**, 692–703 (2023).
93. To, T.-L. et al. PMF-seq: a highly scalable screening strategy for linking genetics to mitochondrial bioenergetics. *Nat. Metabol.* **6**, 687–696 (2024).
94. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
95. Replogle, J. M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).

**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

## Methods

### Animal models and husbandry

Animals were housed in a facility according to the UK Home Office guidelines upon approval by the University of Cambridge Animal Welfare and Ethical Review Body and the UK Home Office (project license P6C97520A). Mice were kept in individually ventilated cages at 20–24 °C, 45–65% humidity on a 12-h light–dark cycle. The mouse line used in this study was m.5024C>T (allele symbol: *mt-Ta*<sup>mjbst</sup>, MGI 5902095), bred on the C57BL/6 background under protocol 5 of project license P6C97520A (Breeding and Maintenance of Genetically Modified Animals).

### Cell culture and transgenic cell lines

Unless otherwise stated, all cells in this study were maintained in high-glucose DMEM with pyruvate (Gibco, 41966-029) supplemented with 10% FBS (Gibco, 16000-044) and 50 µg ml<sup>-1</sup> uridine (Sigma, U3750) at 37 °C and 5% CO<sub>2</sub> in a humidified incubator. WT HeLa cells were purchased from the European Collection of Authenticated Cell Cultures (93021013). HEK293T cells were purchased from Takara (Lenti-X 293T cells; 632180). None of the cell lines used in this study appear in version 13 of the ICLAC Register of Misidentified Cell Lines.

Primary m.5024C>T MEFs were described previously<sup>11</sup>, isolated from individual E13.5 embryos and immortalized by transfection with the SV40 large T antigen (pBSSVD2005, a gift from D. Ron; Addgene, plasmid 21826). Cas9-expressing m.5024C>T MEFs were generated by transducing with LentiCas9-Blast lentivirus (Addgene, 52962)<sup>96</sup>. Following transduction, cells were selected with 10 µg ml<sup>-1</sup> blasticidin (Sigma, SBR00022). Clonal populations were isolated and heteroplasmy levels assessed by pyrosequencing. To test Cas9 efficiency, m.5024C>T Cas9-blast clones were first transduced with pLJM1-EGFP lentivirus (Addgene, 19319)<sup>97</sup> to stably express eGFP and then transduced with CROPseq–RFP lentivirus expressing eGFP-targeting gRNAs<sup>98</sup>. Editing efficiency was assessed by flow cytometry, with mRFP expression (that is, integration and expression of the gRNA cassette) in absence of eGFP expression (that is, KO of the transgene) indicating successful editing.

*TFAM*-KO HeLa cell lines were generated using gRNAs and lentivirus described below. Following transduction, cells were selected for transgene expression with 1 µg ml<sup>-1</sup> puromycin (Sigma, P4512). To isolate clonal *TFAM*-KO populations, puromycin-selected cells were subcloned, and *TFAM* KO was confirmed by western blot. PIP–FUCCI transgenic cells were generated using plasmids and lentivirus described below. Following transduction, cells were subcloned, selecting cells on the basis of the Cdt1–mVenus G1-phase reporter. Clones were checked after expansion by flow cytometry to identify clones showing expected mVenus and mCherry expression patterns for use in downstream experiments.

ΔH2.1 cybrid cells carrying a heteroplasmic mtDNA deletion were obtained from C. Moraes and were generated by fusing 143B(TK<sup>-</sup>) osteosarcoma cells with enucleated patient-derived fibroblast cells harboring a 7.5-kb partial deletion in the mtDNA<sup>74</sup>.

For OXPHOS experiments, in addition to standard high-glucose DMEM, cells were maintained in low-glucose or galactose medium. For low-glucose medium, DMEM 1 g l<sup>-1</sup> glucose (Gibco, 31885-023) was supplemented with 4.5 g l<sup>-1</sup> glucose, 10% FBS and 50 µg ml<sup>-1</sup> uridine. For galactose medium, no-glucose DMEM (Gibco, 11966-025) was supplemented with 10 mM D-galactose (Sigma, G0750), 110 mg l<sup>-1</sup> sodium pyruvate (Gibco, 11360-070), 10% dialyzed FBS (Gibco, 26400-044) and 50 µg ml<sup>-1</sup> uridine<sup>99</sup>.

Clonal cell populations were obtained by sorting single cells into individual wells of a flat-bottom 96-well culture plate, containing 200 µl of culture medium, using a BD FACS Melody cell sorter as described below. Then, 1 week after sorting, wells were checked for the presence of a single colony of cells to reduce the risk of obtaining nonclonal populations. Clones were expanded to 24-well plates after 2 weeks,

after which further characterization (for example, heteroplasmy measurement) was performed to select suitable experimental clones.

### Plasmids and gRNA library construction

We modified CROPseq-guide-Puro (Addgene, plasmid 86708) lentiviral plasmid<sup>22</sup> to embed the gRNA sequences in the 3' UTR of a polyadenylated RFP transcript, generating the CROPseq–RFP lentiviral vector. This enables simultaneous enrichment of RFP-expressing transduced cells by fluorescence-activated cell sorting and detection of gRNA sequences through 3' gene expression profiling. A modified RFP cDNA was a gift from F. Merkle but carried a stretch of seven adenosines, leading to spurious annealing of oligo-dT primers within the coding sequence. For future experiments, we recommend using the CROPseq–mCherry plasmid that we cloned subsequently and can be requested by contacting the corresponding author.

gRNA sequences for CROPseq (Supplementary Table 1) were designed as described previously<sup>22</sup>, ordered from Twist Biosciences as a ssDNA oligo pool and PCR amplified before cloning. Then, 200 fmol of the amplified oligo pool was cloned using the ClonExpress Ultra one-step cloning kit (Vazyme, C115) into 10 fmol of gel-purified CROPseq–mCherry plasmid digested by BsmBI overnight. The cloning mixture was desalted by dialyzing against nuclease-free water using a 0.05-µm membrane filter (Merck, VMWPO2500) for 30 min. The desalted cloned plasmids were transformed into Endura electrocompetent cells (Lucigen, 60242-1) by electroporation in prechilled 1-mm cuvettes at 25 µF, 200 Ω, 1.5 kV before resuspending in 1 ml of recovery medium to recover and then plating or by heat shock in high-efficiency stable competent *Escherichia coli* (New England Biolabs (NEB), C3040H). Bacterial colonies (~43,000, estimated ~700× per gRNA) were scraped from plates. Plasmid DNA was extracted using a plasmid midi kit (Qiagen, 12143) and run on agarose gel; then, nonrecombined plasmid was extracted using a Monarch DNA gel extraction kit (NEB, T1020).

The single gRNA sequence targeting human *TFAM*<sup>100</sup> was ordered as desalted oligos (Sigma) with overhangs for cloning, annealed by heating at 95 °C for 2 min and cooled at 1 °C min<sup>-1</sup> to room temperature. The annealed oligo was cloned into the LentiCRISPRv2 backbone<sup>96</sup> and transformed in 10-beta competent *E. coli* (NEB, C3019H); then, plasmid DNA was extracted using the QIAprep spin miniprep kit (Qiagen, 27104).

Single gRNA sequences targeting mouse *Tfam* and *Opa1*, as well as NT gRNA, were ordered as desalted oligos (Sigma), annealed by heating at 95 °C for 2 min and cooled at 1 °C min<sup>-1</sup> to room temperature. Annealed gRNAs were cloned into LentiCRISPRv2-RFP670 (Addgene, plasmid 187646) at a 1:5 vector-to-insert ratio by Gibson assembly using the ClonExpress Ultra one-step cloning kit and transformed in 10-beta competent *E. coli*; then, plasmid DNA was extracted using the QIAprep spin miniprep kit (Qiagen, 27104).

The pCAG-mCherry-P2A-i4Dam and pCAG-mCherry-P2A-i4Dam-Atf4 plasmids were constructed by Gibson assembly into pCAG-IRES-GFP, using mCherry as an upstream open reading frame and Dam (Addgene, plasmid 59217)<sup>65</sup> with a C-terminal Myc tag. The *Atf4* gene was amplified from mouse cDNA and cloned in frame into the pCAG-mCherry-P2A-DamID construct 3' of Dam and the MYC tag. A P2A ribosome-skipping sequence was placed between mCherry and Dam, resulting in overexpression of Dam and ATF4 (that is, not targeted<sup>64</sup>), to override post-transcriptional or post-translational regulation of ATF4. To allow transient transfection of DamID<sup>87</sup>, an intron was inserted into Dam by ligating oligos with a modified synthetic intron (IVS)<sup>101</sup> into the BamHI restriction site within Dam, between the third and fourth helices of the DNA-binding domain of the Dam methylase<sup>102</sup>; then, plasmids were transformed in *dam*/*dcm*<sup>-</sup> competent *E. coli* (NEB, C2925SH).

### Flow cytometry

Expression of fluorescent markers was assessed using a BD LSRFortessa cell analyzer with appropriate laser, filter and detector settings.

Postacquisition data analysis was performed in FlowJo version 10. Single-cell sorting was performed using a BD FACS Melody cell sorter according to the manufacturer's instructions. In brief, samples were initially gated using doublet discrimination to select live single cells on the basis of forward scatter and side scatter parameters. If cells were being sorted on the basis of expression of a fluorescent marker, additional gating was performed to identify marker-positive cells. Cells in the desired gate were then sorted into tubes (bulk sorts) or plates (single-cell sorts), all sorts were performed in 'single-cell' sort mode.

### Lentiviral production and transduction

First, 24 h before transfection, HEK293T Lenti-X cells (Takara, 632180) were plated at  $0.75 \times 10^6$  cells per well in six-well plates. Transfections were performed with cells at 80–90% confluency using TransIT-293 transfection reagent (Mirus, MIR2704) according to the manufacturer's instructions. The envelope plasmid (pMD2.G; Addgene, plasmid 12259), vector plasmid and packaging plasmid (psPAX2; Addgene, plasmid 12260) were added to the transfection mix in a 2:3:4 molar ratio. Supernatant containing lentiviral particles was harvested at 48 h after transfection and passed through a 40- $\mu$ m filter to remove cellular debris. Isolated virus was stored at  $-70^\circ\text{C}$  before use in transductions. Cells to be transduced were plated at  $5 \times 10^4$  cells per well in 24-well plates, an appropriate volume of thawed and prewarmed lentiviral supernatant was added and the total volume was brought to 500  $\mu$ l per well with fresh culture medium. Then, 3 h after transduction, an additional 500  $\mu$ l of culture medium was added to bring the total volume to 1 ml. Cells were passaged to six-well plates at 24 h after transduction. To titrate lentiviral stocks, cells were transduced with increasing volumes of viral supernatant as described above. At 7 days, transduced cells were analyzed to assess expression of the transfer plasmid fluorescent marker, which acted as an indication of the transduction efficiency; this was then used to calculate the volume of viral supernatant required to achieve a given proportion of transduced cells.

### Transient transfection and DamID-seq

All transient transfections were performed using Lipofectamine 2000 (Thermo, 11668027). The m.5024C>T MEFs were plated at  $2 \times 10^6$  cells per well in six-well plates in 2 ml of serum-free medium immediately before transfection. Following the manufacturer's protocol, 1  $\mu$ g of each plasmid (pCAG-mCherry-p2A-i4Dam-Atf4 or pCAG-mCherry-p2A-i4Dam) prepared in *dam<sup>-</sup>/dcm<sup>-</sup>* competent *E. coli* (NEB, C2925H) was added to each well of a six-well plate. After 5 h, the medium was replaced with DMEM plus FBS. Then, 48 h after transfection, cells were harvested and the DNA was extracted using a QiaAmp DNA micro kit (Qiagen, 56304) and processed for DamID as described previously<sup>103</sup>. DamID fragments were prepared for Illumina sequencing according to a modified TruSeq protocol. Sequencing was performed as paired-end 50-bp reads by the Cancer Research UK (CRUK) Genomics Core Sequencing facility on a NovaSeq 6000.

### Whole-cell 10X Genomics multiome and sequencing

Cells were prepared for 10X Genomics multiome ATAC and gene expression sequencing following a modified version of the standard 10X Genomics nuclei isolation protocol<sup>23</sup>. In brief,  $5 \times 10^5$  cells were fixed in 0.1% formaldehyde (Thermo) for 5 min at room temperature, followed by permeabilization in 0.1% NP40 (Thermo) for 3 min at  $4^\circ\text{C}$  in the presence of RNase inhibitor (Roche) to prevent mRNA degradation. Fixed and permeabilized cells were resuspended in diluted nucleus buffer (10X Genomics), counted and adjusted to a concentration of 3,000 cells per  $\mu$ l. Following fixing and permeabilization of cells, subsequent steps were performed at the CRUK Cambridge Institute Genomics Core. Transposition, probe hybridization and ATAC and gene expression library preparation were performed using the Chromium Next GEM single-cell multiome ATAC and gene expression reagent kit (10X Genomics, PN-1000283) according to the manufacturer's instructions

(protocol no. CG000810 Rev A). Illumina sequencing was performed by the CRUK Genomics Core Sequencing facility on a NovaSeq 6000, with each run on a single lane of an SP flow cell, returning approximately 650–800 million paired-end reads per lane. Sequencing parameters were set as follows: gene expression, read 1, 28 bp; read 2, 90 bp; i5 index, 10 bp; i7 index, 10 bp; ATAC, read 1, 50 bp; read 2, 50 bp; i5 index, 16 bp; i7 index, 8 bp.

### gRNA enrichment and sequencing

CROPseq gRNA sequences were enriched from 10X Genomics multiome gene expression cDNA libraries using TAP-seq<sup>42</sup>. Briefly, 15  $\mu$ l of cDNA from the 10X GEM reverse transcription reaction was input to a PCR reaction (TAP PCR 1) containing a CROPseq-RFP specific forward primer binding 54 bp upstream of the 3' gRNA sequence (CROPouter) and a reverse primer binding the Illumina Truseq read 1 (partial read 1), producing an amplicon approximately 550–600 bp long covering the gRNA sequence and retaining the 3' 10X cell barcode and unique molecular identifier sequences. To increase specificity, a second nested PCR (TAP PCR 2) was performed on the 10-ng TAP PCR 1 product using a second CROPseq-RFP-specific forward primer carrying the Illumina Truseq read 2 sequence and binding immediately upstream of the CROPseq-RFP gRNA sequence (CROPinner) and the partial read 1 reverse primer. Next, 10 ng of TAP PCR 2 product was input to a third PCR reaction (TAP PCR 3) using a forward primer carrying the Illumina P7 sequencing adaptor and a 10-bp index sequence binding the Truseq read 2 sequence (Illumina P7) and a reverse primer carrying the Illumina P5 sequencing adaptor and binding the Truseq read 1 sequence (targeted 10X). Following TAP PCR 3, enriched libraries were quantified using the KAPA library quantification kit (Roche). Sequencing on the Illumina MiSeq system was performed using the MiSeq reagent nano kit v2 for 300 cycles according to the manufacturer's instructions. Following quantification, sequencing libraries were normalized to 4 nM, denatured, diluted to a final concentration of 10 pM and pooled with 1% 12.5 pM denatured PhiX control before sequencing. Sequencing parameters were identical to the Multiome gene expression libraries described above.

### mtDNA enrichment and sequencing

mtDNA sequences were enriched from multiome ATAC sequencing libraries by hybridization capture, using a custom xGEN hybrid capture panel (Integrated DNA Technologies) containing 270 probes targeting the mouse mtDNA sequence (average of one probe every 60 bp) (Supplementary Table 2). Hybridization capture was performed on 500 ng of ATAC sequencing library using the xGen hybridization and wash kit (Integrated DNA Technologies, 10010351) according to the manufacturer's instructions, including the optional AMPure XP bead (Beckman Coulter, A63880) DNA concentration protocol steps. Post-capture PCR amplification was performed for 12 cycles and the final libraries were quantified using the KAPA library quantification kit and pooled for Illumina sequencing with parameters identical to the Multiome ATAC libraries described above.

### Bulk RNA-seq of single-gRNA CRISPR cells

Following lentiviral transduction with CROPseq-RFP single-gRNA CRISPR vectors, 50,000 RFP-positive cells per sample were sorted at day 6 after transduction. Total RNA was extracted from the cells using the Quick-RNA microprep kit (Zymo Research, R1050) and RNA concentration and RNA integration number equivalent (RIN<sup>e</sup>) values were assessed using the high-sensitivity RNA ScreenTape system (Agilent, 5067-5579/5580). RIN<sup>e</sup> > 8.5 was confirmed for all samples and samples were adjusted to a final concentration of 10 ng  $\mu$ l<sup>-1</sup> before library preparation using the NEBNext single-cell, low-input RNA library prep kit (NEB, E6420S) according to the manufacturer's instructions. Libraries were pooled and sequenced on the Illumina NovaSeq X platform on a single lane of a 1.5B flow cell using 50-bp paired-end reads, yielding approximately 750 million reads.

### mtDNA ddPCR and pyrosequencing

Single-cell mtDNA CN measurements were made using the Bio-Rad QX200 AutoDG ddPCR system<sup>81</sup>. In brief, single-sorted cells in 96-well plates were lysed in lysis buffer containing 1% Tween 20 (Life Technologies, 003005) and 200  $\mu\text{g ml}^{-1}$  proteinase K (Ambion, AM2546) at 37 °C for 30 min followed by 85 °C for 15 min to inactivate proteinase K. Cell lysate was then input to ddPCR reactions containing primer and probe combinations targeting the *mt-Nd1* and *mt-Co3* genes of the mouse mtDNA. Following data acquisition, the CNs obtained from the two independent mtDNA probes were averaged to give a final absolute mtDNA CN measurement for each cell.

Single-cell pyrosequencing was performed<sup>104</sup> using the PyroMark Q48 Autoprep pyrosequencing system (Qiagen) according to the manufacturer's instructions using a primer set specifically targeting the m.5024C>T mutation site.

### Long-range PCR and long-read sequencing

Genomic DNA was extracted from ear clips biopsies, taken from three m.5024C>T mice at 2 weeks of age, using the Monarch genomic DNA purification kit (NEB, T3010) and quantified by a Qubit fluorometer. A segment of mtDNA spanning positions 13200 to 5500 was amplified from 10 ng of genomic DNA by long-range PCR using PrimeSTAR GXL premix (Takara, R051B). The amplicons were isolated by Ampure XP bead cleanup and sequenced by long-read Nanopore sequencing (Plasmidsaurus, Oxford Nanopore Technologies). Reads were aligned to the GRCm39 mouse mtDNA reference genome excluding reads shorter than 8 kb. Variant positions 13614, 13715, 1781, 1866, 3009, 3823 and 5024 were extracted and converted into a binary matrix with the mutant or WT allele. For each pairwise combination of these positions, the percentage of positions matching either the WT or mutant haplotype was calculated to find the co-occurrence of alleles belonging to each haplotype.

### Western blotting and antibodies

Cultured cells were washed, dissociated using trypsin (Gibco, 15400-054) and pelleted before being snap-frozen in liquid nitrogen. Protein extraction was performed by mixing 500  $\mu\text{l}$  of PathScan Sandwich ELISA lysis buffer (1 $\times$ ) (Cell Signaling, 7081) with each cell pellet. Following a 2-min incubation on ice, samples were spun down for 1 min at 14,000g at 4 °C to remove cell debris. Protein concentration was measured using the Pierce BCA protein assay kit (Thermo, 23227). NuPAGE gels were run at 165 V for 45 min. Membrane transfer was performed using an iBlot machine (Life Technologies). A 1-h incubation at room temperature in 5% milk in Tris-buffered saline with 0.1% Tween 20 was used for blocking. The membrane was incubated with the primary antibody overnight at 4 °C, while the secondary incubation was performed at room temperature for 1 h. Primary antibodies were anti-TFAM (Cell Signaling, 8076S) and anti-vinculin (Sigma, V4505). The Clean-Blot immunoprecipitate detection kit (horseradish peroxidase) (Thermo, 21232) was used for detection and imaging was conducted using the Amersham Imager 600 (General Electric).

### High-resolution respirometry

High-resolution respirometry was carried out using the O2k-Respirometer (Oroboros Instruments). Calibration was performed with DMEM plus 10% FBS. A total of 5 million cells suspended in 2 ml of medium were added to each chamber. Once the oxygen consumption rate reached a plateau, three drugs were added sequentially. First, 5  $\mu\text{l}$  of oligomycin A (Merck, O4876). After establishing the new baseline, 2  $\mu\text{l}$  of carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) (Merck, C2759) was added, followed up by successive doses of 1  $\mu\text{l}$  until a plateau of the maximum respiratory capacity was reached. Lastly, 1  $\mu\text{l}$  of rotenone (Merck, R8875) and 2  $\mu\text{l}$  of antimycin A (Merck, A8674) were added in quick succession. Baseline respiration was calculated by subtracting the background (respiration still present after the addition of rotenone and antimycin A from the basal respiration of the sample, before the

addition of any drugs). Proton leak was measured by subtracting the background from the respiration still present when oligomycin A was added. ATP-linked respiration was estimated by subtracting the proton leak from the basal respiration. The maximal respiratory capacity of the sample was calculated when the background was subtracted from respiration in the presence of CCCP.

### Bioinformatic analysis

**Initial analysis and QC.** Raw FASTQ files from Multiome sequencing were combined with the 'lost reads' FASTQ files to recover missing or low-quality reads that were initially discarded during sequencing. The FASTQ files were first run through FastQC (version 0.11.9; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to perform QC checks on raw sequencing data and through FastQ Screen (version 0.14.1; [https://www.bioinformatics.babraham.ac.uk/projects/fastq\\_screen/](https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/)) to check for contamination. The combined FASTQ files were then processed using Cell Ranger ARC (version 2.0.1) from 10X Genomics, which performs alignment, filtering, barcode counting and peak calling, to generate feature–barcode matrices for downstream analyses. Cell Ranger ARC was first run with the standard *Mus musculus* reference genome (GRCm38/mm10, GENCODE vM23/Ensembl 98)<sup>105</sup>, before running with a custom-built *M. musculus* reference genome created with `cellranger mkref`. Some mtDNA regions exhibit low coverage because of homology with nuclear DNA, and hard-masking these NUMTs is recommended<sup>17</sup>. A custom blacklist (<https://github.com/caleblareau/mitoblacklist/>) was incorporated into the original reference, generating a hard-masked, modified *M. musculus* genome. Using the `gex_possorted_bam.bam` file, generated from the Cell Ranger ARC pipeline, the data were run through Qualimap (version 2.2.1)<sup>106</sup> to evaluate the overall mapping quality of our sequencing data.

The filtered feature–barcode matrices, stored as sparse matrices in hdf5 format, were loaded into R (version 4.3.3) to perform data QC and standard preprocessing steps with the Seurat R package (version 5.1.0)<sup>71</sup> for RNA data processing and the Signac R package (version 1.13.0)<sup>107</sup> for ATAC data processing. Data QC was performed for both scRNA-seq and scATAC-seq data. For RNA, this involved removing cells with total RNA reads less than 1,000 and greater than 60,000, number of detected genes less than 1,000 and greater than 10,000, mitochondrial RNA percentage greater than 10 and ribosomal percentage greater than 30. For ATAC, this involved removing total ATAC reads less than 1,000 and greater than 150,000, number of accessible regions less than 500 and greater than 55,000, nucleosome banding pattern signal greater than 2 and TSS enrichment score less than 1.

Following QC, RNA-seq and ATAC-seq data were independently normalized and scaled, after which linear dimensional reduction was performed. For the RNA-seq data, cell-cycle scoring was conducted to minimize the impact of cell-cycle heterogeneity, assigning each cell a score on the basis of the expression of G2/M and S phase markers. This was initially calculated with the Seurat R package (`CellCycleScoring`), followed by pseudotime analysis with the tricycle R package (version 1.12.0)<sup>72</sup>, as described below.

As with QC, filtering and initial data processing, integration was performed separately for RNA-seq and ATAC-seq data before being combined into one multiome object for downstream analyses. RNA-seq data integration was performed by first merging the two Seurat objects, followed by `sctransform` normalization while regressing out cell-cycle effects. Next, linear dimensional reduction was applied and cell clustering was determined by computing *k*-nearest neighbors and constructing a shared nearest neighbor graph. Lastly, UMAP was used for dimensionality reduction and visualization. Data integration was performed using the `IntegrateLayers` function with the canonical correlation analysis (CCA) method. CCA was selected because the datasets shared common cell types while potentially containing technical or batch-related differences. This method is particularly effective for correcting subtle batch effects while preserving strong biological signals. Following integration,

linear dimensional reduction, clustering and UMAP visualization were repeated. ATAC-seq data integration began by merging the two Seurat objects, followed by linear dimensional reduction, clustering and UMAP visualization. Integration was then performed using 'FindIntegratingAnchors' and 'IntegrateEmbeddings', after which linear dimensional reduction, clustering and UMAP visualization were repeated. The two integrated Seurat objects were combined and the 'FindMultiModal-Neighbours' function was used to construct a WNN graph, followed by UMAP visualization.

For mtDNA CN analysis, mtDNA coverage is correlated with total (nuclear + mtDNA) ATAC-seq coverage per cell and others have previously included a normalization step<sup>46</sup>. We chose not to include normalization as it would have prevented the cell-cycle analysis in Fig. 5 and Extended Data Figs. 9 and 10 because total per-cell ATAC-seq read count increases from G1 to S and G2, in line with nuclear DNA replication (Extended Data Fig. 10a,b). Normalization of the mtDNA read depth to total depth would in this case remove crucial biological variation and insight. Normalization could be applied for comparison between perturbation groups (Fig. 2j) but we opted for a homogeneous approach to estimate mtDNA CN across the manuscript.

**gRNA detection.** gRNA sequences were present in the 90-bp read 2 sequences of the gene expression library, these reads were extracted and assigned to individual cells using a bespoke pipeline. First, raw gene expression FASTQ files were processed with UMI-tools (version 1.0.1)<sup>108</sup> extract to embed the IOX cell barcode sequence contained in read 1 into the corresponding read 2 read name string. Barcoded read 2 FASTQs were then aligned using bwa-mem (version 0.7.17)<sup>109</sup> to a custom reference genome containing the sequence for each of the 60 gRNAs in the CROPseq library, flanked on both sides by 85-bp anchor sequences from the surrounding CROPseq-RFP backbone, ensuring sufficient reference sequence to successfully align all 90-bp read 2 sequences that overlapped at least 5 bp of the gRNA sequence. The resulting SAM output was processed in R; aligned reads were filtered to retain those with MAPQ  $\geq 30$  and individual reads were assigned to single cells by cross-referencing the embedded barcode sequence in the QNAME SAM field with the corresponding barcodes allocated by Cell Ranger ARC. Cells were given a single gRNA assignment if they had  $\geq 2$  separate barcoded gRNA reads identified and, in cases where multiple gRNAs were identified, the ratio of reads from the most abundant gRNA to total gRNA reads was  $>2.3$ .

**Heteroplasmy calling.** Single-cell mtDNA variant identification was performed using the mgatk package (version 0.7.0)<sup>17</sup> implemented in `tenx` mode, with the NUMT-masked Cell Ranger ARC `atac_pos-sorted_bam.bam` output file and corresponding known cell barcode list as inputs, which includes removal of PCR duplicate reads performed at the level of individual cells to give single-cell deduplicated per-base mtDNA coverage data. High-confidence heteroplasmic variants were reported in the `mgatk.variant_stats.tsv` output. To call single-cell heteroplasmy values, per-base mtDNA sequencing data contained in the `mgatk.signac.rds` output file was first imported into the Seurat object containing the corresponding Cell Ranger ARC gene expression and ATAC assays using the `ReadMGATK` Seurat command. Next, base calls at the seven high-confidence heteroplasmic variants corresponding to the reference and mutant alleles were extracted from the mgatk assay and the ratio of mutant to WT alleles at each variant position was used to calculate heteroplasmy. Investigation of the correlation between the heteroplasmy calls at individual variant positions, combined with previously published data<sup>10</sup>, confirmed that the two principal variant positions, m.5024C>T and m.13715C>T, were linked on a single mtDNA haplotype and were inversely correlated with the remaining five variants, m.1781C>T, m.1866A>G, m.3009G>T, m.3823T>C and m.13614C>T, all linked on a second mtDNA haplotype. On the basis of this confirmed linkage among all seven variants, we

were able to treat the coverage at each separate variant position as an independent observation of the same underlying heteroplasmy, thus allowing us to combine the WT and mutant base calls at all seven variant sites to increase the effective coverage and maximize the accuracy of the single-cell heteroplasmy calls.

**Heteroplasmy modeling.** In silico modeling of heteroplasmy calling on the basis of subsampling of a simulated heteroplasmic mtDNA population indicated that the sample size, corresponding to the read depth at heteroplasmic sites, has a strong influence on the accuracy of single-cell heteroplasmy estimation, with greater depth resulting in increasingly accurate calls. Previous studies suggested implementing a minimum coverage cutoff of 20 reads to confidently identify a heteroplasmic variant<sup>17,21</sup>. To model the effect of mtDNA coverage on the accuracy of heteroplasmy calls, we generated in silico cell populations with simulated heteroplasmy values according to the distribution of heteroplasmy in our MitoPerturb-Seq dataset. On the basis of our experimental data, we used an mtDNA CN of 1,750 (measured by ddPCR; NT gRNA cells in Extended Data Fig. 7a) and a mean heteroplasmy of 58.2% with s.d. of 11.3% (calculated using all cells in the integrated dataset with combined heteroplasmic SNV coverage  $\geq 20$ ) to generate a set of normally distributed 'true' heteroplasmy values. To investigate the impact of sampling heteroplasmy at different mtDNA coverage, we simulated a population of 2,000 cells and randomly subsampled each cell with sample sizes of 5, 20, 50 and 100. These samples were then used to calculate 'sampled' heteroplasmy values for the cell and, for each sample size, the sampled heteroplasmy was plotted against the simulated true heteroplasmy for the corresponding cell, with regression line and  $R^2$  value calculated using ggplot2 (version 3.5.1)<sup>111</sup> `geom_smooth()`. To model the likely impact of applying a coverage threshold of 20 reads to our integrated MitoPerturb-Seq dataset, we simulated a matching population of 6,510 cells (equal to the number of cells with coverage  $> 0$  at all seven heteroplasmic SNV sites combined) as described above and sampled these cells using the same distribution of post-mtDNA enrichment coverage that we saw in our integrated data to calculate the 'sampled' heteroplasmy value for each cell. Sampled heteroplasmy was then plotted against the simulated true heteroplasmy, with cells sampled at depth  $< 20$  highlighted. Together, this allowed us to conclude that, in our case, a threshold of  $\geq 20$  reads at the combined heteroplasmic variant sites represented a good compromise, effectively eliminating the cells with the most inaccurate heteroplasmy calls whilst retaining the majority of cells for downstream analysis. We note that, in datasets with higher average per-cell mtDNA coverage (for example, from cells with high-mtDNA CN, like hepatocytes<sup>19</sup>), it may well be possible to increase this read-depth threshold to improve overall accuracy without significantly impacting the number of cells retained in the dataset.

**Simulating the effect of reduced sequencing depth on heteroplasmy variance.** To test whether the increased heteroplasmy variance observed in *Opal*-KD, *Polg*-KD and *Tfam*-KD cells could be attributed solely to the reduced mtDNA depth at relevant genomic sites, we performed a computational simulation modeling the expected heteroplasmy variance from sampling heteroplasmy levels in control gRNA cells at mtDNA depths characteristic of the KD cells. For each simulation, cells from the control gRNA group were sampled without replacement to match the number of cells in each KD group. For each sampled control group cell, a mtDNA depth value was sampled with replacement from the KD group to approximate the mtDNA depth distribution. A simulated alternative allele count was estimated for each cell through binomial sampling, with the sampled mtDNA depth as the number of trials and the heteroplasmy level from the sampled control cell as the probability of picking the alternative allele. The heteroplasmy level for each cell was then calculated as the ratio of its simulated alternative allele count to its assigned sampled depth.

The null distribution of simulated heteroplasmy variances was generated from 5,000 simulations performed as described above and the observed *Opa1*-KD, *Polg*-KD and *Tfam*-KD heteroplasmy variance was assessed against this distribution using a two-sided empirical *P* value, adjusted for multiple comparisons using the Bonferroni–Holm method at a significance level of  $\alpha = 0.05$ .

**Modeling depth-dependent heteroplasmy variance.** To characterize the global relationship between read depth and heteroplasmy variance, single cells were aggregated into bins of equal sample size to ensure stable variance estimates. Cells from the population without mtDNA reduction (aggregated across all gRNAs excluding *Tfam*, *Polg* and *Opa1*) were grouped into 100 bins, while the rest were grouped into 15 bins. We modeled the expected baseline variance ( $\text{Variance} \propto \frac{1}{\text{depth}}$ ) using quantile regression (`quantreg`) on the nondepleted bins, fitting the 5th and 95th percentiles. To robustly estimate uncertainty boundaries, particularly at low read depths, we calculated 95% confidence intervals (CIs) derived from 200 bootstrap iterations of resampled NT bins.

**Cell-cycle pseudotime analysis.** In addition to performing cell-cycle annotation with Seurat, we used the tricycle R package (version 1.12.0)<sup>72</sup> to predict, analyze and visualize cell-cycle states in scRNA-seq. Unlike traditional cell-cycle scoring methods that rely on discrete phase markers, tricycle estimates a continuous cell-cycle trajectory using a reference-based projection approach. By leveraging a pretrained internal reference using the `Runtricycle` function, tricycle first projects data onto the cell-cycle embeddings using the reference and then estimates cell-cycle position. The estimated cell-cycle positions range from  $0\pi$  to  $2\pi$ , with  $0.5\pi$  being the start of the S stage,  $\pi$  being the start of the G2/M stage and  $1.5\pi$  being the middle of the M stage. The results of this function are added as metadata to the Seurat object. To evaluate tricycle's performance, we examined the expression of key genes relative to cell-cycle position. Cell-cycle annotations from Seurat and tricycle showed high concordance, with 82.3% of annotations aligning. Of the 17.7% that did not align, 9.7% of these were a mismatch between G1 and S annotation, probably because Seurat assigns S and G2M scores on the basis of the expression of predefined marker genes, with cells exhibiting low scores for both phases inferred to be in G1 phase. Additionally, chi-square tests for independence with Bonferroni multiple testing correction were performed for each gRNA across the different cell-cycle phases to assess association between gRNA presence and cell-cycle phase distribution.

**Mixscape analysis.** For unbiased perturbation assessment from scRNA-seq datasets, we used the Mixscape method<sup>49</sup>, implemented within the Seurat R package. We first calculated perturbation signatures (`CalcPerturbSig`), setting the number of nearest neighbors to 20; when clustering by these signatures, technical variation is removed and a specific perturbation cluster is identified. Using these signatures, the `RunMixscape` function assumes that each target gene is a mixture of two Gaussian distributions (KO and NP) and that NP cells have the same distribution as those expressing NT gRNAs. Each cell is assigned a posterior probability of belonging to the KO group; cells with a probability greater than 0.5 are labeled KO. For the present study, Mixscape analysis was only conducted in cells that were already confidently assigned a gRNA, without cells labeled as either unknown or containing multiple gRNAs during prior gRNA assignment. All cells labeled with a negative control gene (*Eomes*, *Neurod1*, *Olig1*, *Opn4*, *Rgr* and *Rrh*) or NT gRNA were used as the control population. Following Mixscape assignment, the `PlotPerturbScore` function was used to examine posterior probabilities and perturbation score distributions, comparing those assigned as KO or NP to those assigned NT. Differential expression analyses were performed and visualized with the Mixscape `heatmap` function to see whether KO cells exhibited reduced expression. To maximize class separability, dimensionality reduction was performed with linear discriminant analysis and visualized.

**Differential gene expression.** Before differential gene expression, the `PrepSCTFindMarkers` function from the Seurat R package was used to prepare SCTransform-normalized data for differential gene expression analysis when using the `FindMarkers` set of functions. For the present data, the `FindAllMarkers` differential expression analysis function in Seurat was used to identify marker genes for all clusters in the single-cell dataset. This involved comparing each gRNA group to the control group (negative control genes and NT gRNAs) to identify genes that are significantly expressed in each group. Differential expression analyses were also conducted for cells classified by the Mixscape analysis as KO, comparing each gRNA with a KO classification to the control group (negative control genes and NT gRNA). Differential gene expression results were filtered for an adjusted *P* value  $< 0.05$  and a  $\log_2$  fold change of 0.25. We used a  $\log_2$  fold change threshold of 0.25 (equivalent to  $\pm 20\%$  transcript abundance) to identify genes with a significant but relatively small change in expression, which allowed us to detect consistent subtle changes across related families of genes (Supplementary Tables 4 and 5). To highlight DEGs with more significant changes in transcription, we applied an additional  $\log_2$  fold change threshold of 0.5 (equivalent to  $\pm 50\%$  transcript abundance) when displaying the results of the differential gene expression analysis.

Functional enrichment analysis was performed using the clusterProfiler R package (version 4.12.1)<sup>112,113</sup>. GO enrichment analysis was conducted with the `enrichGO` function, compared to a background gene set (`org.Mm.eg.db`, version 3.20.0). After performing `enrichGO`, the `enrichPlot` R package `pairwise_termsim` function with the Wang method<sup>114</sup> was applied to evaluate the semantic similarity between the enriched GO terms on the basis of the topological structure of the GO graph. Relationships between the enriched GO terms were visualized as a hierarchical tree using the `treemap` function from `enrichPlot`.

Nuclear principal component analysis (PCA) was conducted using RNA-seq data after integration, normalization and initial filtering. A linear model was fitted to the data, with corrections applied for mtDNA coverage estimated from the ATAC-seq data. For each nuclear gene *i*, we modeled the relationship between gene expression and mtDNA CN using the following linear regression: gene expression of gene  $i = \beta_0 + \beta_1 \times \text{mtDNA copy number} + \epsilon$ , where  $\beta_0$  represents the intercept,  $\beta_1$  represents the slope coefficient for mtDNA CN and  $\epsilon$  represents the residual error term. The residuals from each linear model, representing gene expression values adjusted for mtDNA CN, were extracted and used for the subsequent analyses. The top 2,000 variable residuals from this model were used for nuclear PCA. Additionally, nuclear PCA was performed before the correction for mtDNA coverage.

**SCENIC.** Raw scRNA-seq counts from the QC filtered and integrated dataset were used as an input to the Python (version 3.12.2) pyScenic pipeline (version 0.12.1)<sup>115</sup>: Gene regulatory network inference was performed using the GRNBoost2 algorithm, followed by regulon prediction with the `pyscenic ctx` command with `--mask_dropouts` set to TRUE. AUCell (version 1.28.0)<sup>60</sup> was used to calculate cellular regulon enrichment scores, which were then used for downstream analysis. The resulting loom file was loaded into R using `SCopeLoomR` (version 0.13.0; <https://github.com/aertslab/SCopeLoomR>). The expression matrices, regulons and AUCell matrix were extracted using `get_dgem`, `get_regulons` and `get_regulons_AUC`, respectively; key column attributes and metadata were also extracted. Cells were split by gRNA and for each group of cells, the AUC matrix was extracted and the mean regulon activity across all of those cells was calculated. The resulting matrix was scaled by performing z-score scaling per regulon and regulons with missing values were removed. To examine the regulon activity scores for select genes, only regulons with a relative activity score  $> 0$  in all three genes of interest (*Opa1*, *Polg* and *Tfam*) were retained for visualization. Heat maps were generated with `ComplexHeatmap` (version 2.18.0)<sup>116</sup>.

**Bulk RNA-seq analysis.** Overall sequencing quality of Raw FASTQ files was first checked using FastQC, confirming no major issues. Next, reads were trimmed to remove Illumina TruSeq adaptor sequences using Trimmomatic (version 0.39)<sup>117</sup> and aligned to the mouse GRCm38/mm10 genome reference using RUM (version 2.0.4)<sup>118</sup>. Aligned BAM files were used as input to HTSeq-count (version 2.0.3)<sup>119</sup>, using union mode, to count reads in features. Final analysis was performed in R using the EdgeR software package (version 4.4.2)<sup>120</sup>. To confirm the heteroplasmy level of the m.5024C>T clone 8 cells used in this experiment, we extracted reads covering the seven heteroplasmic SNV sites from the aligned BAM files of the two NT gRNA technical replicates. Five of seven sites (m.1781C>T, m.1866A>G, m.3009G>T, m.13614C>T and m.13715C>T) had a depth > 500 reads and these were used to call heteroplasmy levels, which were then averaged to give a final heteroplasmy result for each sample. Using this method, we calculated a mean heteroplasmy level of 45.6% for NT Rep 1 and 45.4% for NT Rep 2.

**DamID-seq analysis.** Raw sequencing reads from Dam-only and Dam-ATF4 samples were processed using the damidseq\_pipeline (version 1.5.3)<sup>121</sup>. Paired-end reads were aligned to the GRCm38/mm10 reference genome using Bowtie2 (version 2.3.2)<sup>122</sup> and mapped reads were assigned to GATC fragments. Signal intensities were computed in 300-bp bins using RPM (reads per million) normalization. Each Dam-ATF4 sample was normalized against each Dam-only sample, resulting in six pairwise comparisons. Peak calling on each pairwise comparison was performed using MACS3 (version 3.0.3)<sup>123</sup> with the Dam-ATF4 sample used as the treatment and the corresponding Dam-only sample used as control. Peaks were called in broad mode, with a fixed fragment size of 300 bp and no model estimation (`--nomodel`). The effective genome size was computed from the mm10 reference genome and significance thresholds were set at  $q < 0.05$  with an mfold range of 5–50. Reproducible peaks from the pairwise comparisons were identified across the biological replicates using IDR<sup>124</sup>, broad peaks for each Dam-ATF4 sample were merged across Dam-ATF4 versus Dam-only comparisons and deduplicated on the basis of signal strength. IDR was computed between all pairwise combinations of the three Dam-ATF4 samples using the Python IDR package (version 2.0.3) and peaks with a reproducibility score below an IDR threshold of 0.01 were retained. A multisample intersection was then performed using BEDTools (version 2.31.0)<sup>125</sup> `multiinter`, generating a consensus peak set that captured overlapping peaks across replicates. The resulting peaks were annotated using HOMER (version 5.1)<sup>126</sup> `annotatePeaks.pl` and motif enrichment analysis was performed using HOMER `findMotifsGenome.pl` with a window size of 300 bp around the peaks to identify enriched sequence motifs within the peak regions. Statistical significance of intersections between ATF4-binding targets and DEGs was performed using the R package SuperExactTest (version 1.1.2)<sup>127</sup> and visualization plots were generated using ggplot2 (ref. 111) and ggVennDiagram (version 1.5)<sup>128</sup>. GO enrichment analysis was performed on genes associated with ATF4-binding sites using the clusterProfiler R package. The `enrichGO` function was used to identify significantly enriched biological process terms with a  $q$ -value threshold of 0.05. The background gene universe was set to 17,693, the number of genes detected in the 6,551 cells that were assigned a gRNA, and relationships between enriched GO terms were visualized as a hierarchical tree plot using the `treemap` function in the `enrichPlot` R package (version 1.28.2). Gene set enrichment analysis was conducted using clusterProfiler `gseGO` to assess the ranking of ATF4 target genes within biological pathways. The ATF4-binding signal on the differential expressed genes was visualized as heat maps using the `plotHeatmap` function from deepTools (version 3.5.6)<sup>129</sup>. Genomic tracks for ATF4-binding signals from DamID-seq were plotted using pyGenomeTracks (version 3.9)<sup>130</sup>.

**Quantification and statistical analysis.** Unless described otherwise, data visualizations were generated with inbuilt package functions from the Seurat R package, including `DimPlot`, `FeaturePlot` and `VlnPlot`,

with ggplot2 (ref. 111), alongside ggplot2 extensions `ggalluvial` (version 0.12.5)<sup>131</sup>, `ggpmisc` (version 0.6.1), `ggpubr` (version 0.6.0) and `ggvenn` (version 0.1.10), and with `ComplexHeatmap`<sup>116</sup>, `enrichplot` and `pheatmap` (version 1.0.12). Statistical tests were conducted in R; details of the tests can be found in the accompanying figure legends. Figures in this publication were created in BioRender; van den Ameerle, J. <https://BioRender.com/j3ro7vq> and <https://BioRender.com/jpd1min> (2026) under license.

### Reporting summary

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

### Data availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, J.v.d.A. ([jv361@cam.ac.uk](mailto:jv361@cam.ac.uk)). All sequencing data were deposited to the Gene Expression Omnibus (GEO) under accession numbers [GSE297416](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE297416), [GSE297418](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE297418) and [GSE297491](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE297491). Additional data are supplied in the Supplementary Information. Source data are provided with this paper.

### Code availability

Custom scripts for data analysis are hosted in a dedicated publicly accessible GitHub repository (<https://github.com/JvdAlab/mitoPerturb-Seq>) and from Zenodo (<https://doi.org/10.5281/zenodo.19008664>)<sup>132</sup>.

### References

- Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- Sancak, Y. et al. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **320**, 1496–1501 (2008).
- Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Dott, W., Mistry, P., Wright, J., Cain, K. & Herbert, K. E. Modulation of mitochondrial bioenergetics in a skeletal muscle cell line model of mitochondrial toxicity. *Redox Biol.* **2**, 224–233 (2014).
- Castellani, C. A. et al. Mitochondrial DNA copy number can influence mortality and cardiovascular disease via methylation of nuclear DNA CpGs. *Genome Med.* **12**, 84 (2020).
- Xu, D. H. et al. SV40 intron, a potent strong intron element that effectively increases transgene expression in transfected Chinese hamster ovary cells. *J. Cell Mol. Med.* **22**, 2231–2239 (2018).
- Horton, J. R., Liebert, K., Bekes, M., Jeltsch, A. & Cheng, X. Structure and substrate recognition of the *Escherichia coli* DNA adenine methyltransferase. *J. Mol. Biol.* **358**, 559–570 (2006).
- Marshall, O. J., Southall, T. D., Cheetham, S. W. & Brand, A. H. Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586–1598 (2016).
- Nash, P. A., Silva-Pinheiro, P. & Minczuk, M. A. Genotyping single nucleotide polymorphisms in the mitochondrial genome by pyrosequencing. *J. Vis. Exp.* (192), e64361 (2023).
- Church, D. M. et al. Modernizing reference genome assemblies. *PLoS Biol.* **9**, e1001091 (2011).
- Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
- Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
- Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

109. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
110. Zhang, H., Burr, S. P. & Chinnery, P. F. The mitochondrial DNA genetic bottleneck: inheritance and beyond. *Essays Biochem.* **62**, 225–234 (2018).
111. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 1st edn (Springer, 2009).
112. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
113. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
114. Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. & Chen, C. F. A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23**, 1274–1281 (2007).
115. van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
116. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
117. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
118. Grant, G. R. et al. Comparative analysis of RNA-seq alignment algorithms and the RNA-seq unified mapper (RUM). *Bioinformatics* **27**, 2518–2528 (2011).
119. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
120. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L. & Smyth, G. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, gkaf018 (2025).
121. Marshall, O. J. & Brand, A. H. damidseq\_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371–3373 (2015).
122. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
123. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
124. Li, Q. H., Brown, J. B., Huang, H. Y. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
125. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
126. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
127. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).
128. Gao, C. H. et al. ggVennDiagram: intuitive Venn diagram software extended. *iMeta* **3**, e177 (2024).
129. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
130. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
131. Brunson, J. ggalluvial: layered grammar for alluvial plots. *J. Open Source Softw.* **5**, 2017 (2020).
132. Burr, S. P. et al. MitoPerturb-Seq: Analysis code for 'MitoPerturb-Seq identifies gene-specific single-cell responses to mitochondrial DNA depletion and heteroplasmy'. *Zenodo* <https://doi.org/10.5281/zenodo.19008664> (2026).

## Acknowledgements

We thank all lab members, H. Ma and Y. Crow for helpful discussions, R. Horvath and H. Biggs for continuous support and interest, S. Jackson for advice on CRISPR screening, C. Lyons for help with cell-cycle analysis, R. Schulte and G. Grondys-Kotarba from the Cambridge Institute for Medical Research Flow Cytometry Facility for assistance with cell sorting, the CRUK Cambridge Institute Genomics Core Facility for library preparation and 10X Genomics multiome and sequencing services, J. B. Stewart for the m.5024C>T mouse and MEFs, F. Merkle for RFP cDNA and C. T. Moraes for the  $\Delta$ H2.1 cybrid cell line. J.v.d.A. is supported by a Wellcome Clinical Research Career Development Fellowship (219615/Z/19/Z), an Evelyn Trust Medical Research Grant (21-25) and a UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council's (BBSRC) Responsive Mode Research Grant (BB/X00256X/1). P.F.C. is funded by a Wellcome Collaborative Award (224486/Z/21/Z), the UKRI BBSRC (BB/Y003209/1) and the Rosetrees Trust (PGL23/100048) and his research is supported by the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Center (BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. P.F.C. and J.v.d.A. are further supported by a Wellcome Discovery Award (226653/Z/22/Z), a Medical Research Council (MRC) award (MC\_PC\_21046) to establish a National Mouse Genetics Network Cluster in Mitochondrial Diseases (MitoCluster) and the LifeArc Center to Treat Mitochondrial Diseases under grant no. 10748. LifeArc is a charity registered in England and Wales (no. 1015243) and in Scotland (no. SC037861). M.S.C. was funded by CRUK Program Grant C6/ A18796 and Discovery Award DRCPGM\100005. J.P., P.F.C. and J.v.d.A. acknowledge core funding from the UKRI MRC to the MRC Mitochondrial Biology Unit (MC\_UU\_00028/5 (J.P.), MC\_UU\_00028/7 (P.F.C.) and MC\_UU\_00028/8 (J.v.d.A.)). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version arising from this submission.

## Author contributions

S.P.B., P.F.C. and J.v.d.A. conceptualized the project. S.P.B., A.G., C.R., A.H.A. and J.v.d.A. designed and conducted the experiments. K.A., S.P.B., A.D., W.W. and M.P. performed the bioinformatic analysis. M.S.C. and J.P. provided essential reagents. P.F.C. and J.v.d.A. supervised the project and obtained funding. S.P.B., K.A., P.F.C. and J.v.d.A. wrote the paper and all authors edited and approved the final manuscript. S.P.B. and K.A. contributed equally. A.G. and A.D. contributed equally.

## Competing interests

The authors declare no competing interests.

## Additional information

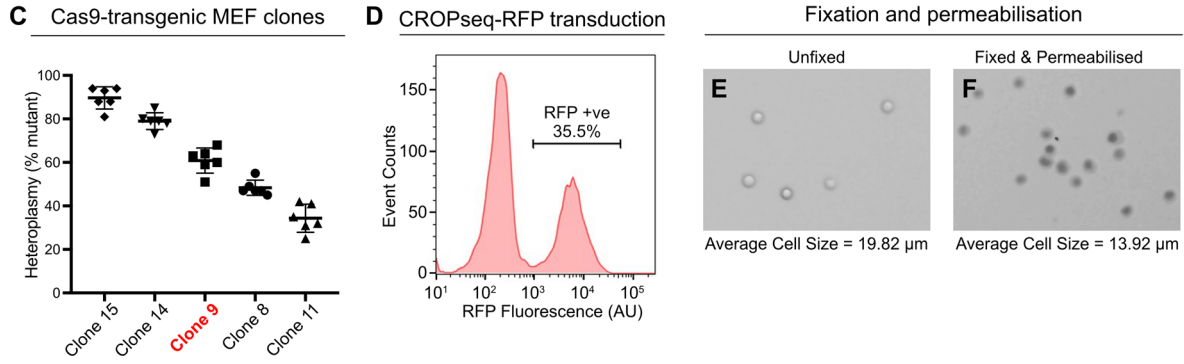
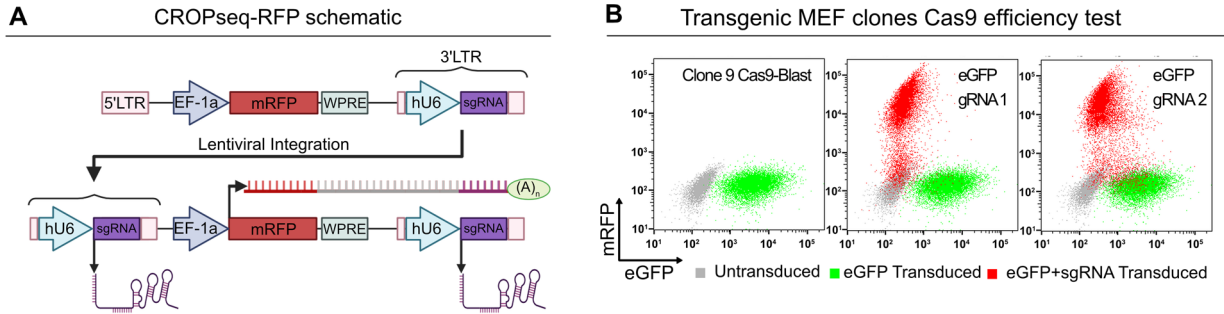
**Extended data** is available for this paper at <https://doi.org/10.1038/s41594-026-01779-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41594-026-01779-7>.

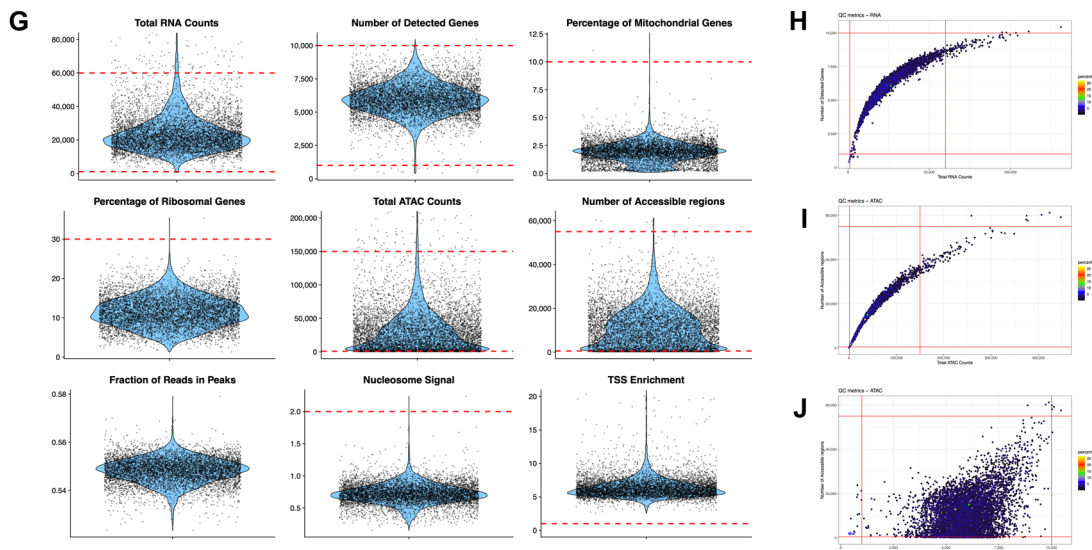
**Correspondence and requests for materials** should be addressed to Patrick F. Chinnery or Jelle van den Ameele.

**Peer review information** *Nature Structural & Molecular Biology* thanks Leif Ludwig and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Melina Casadio and Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

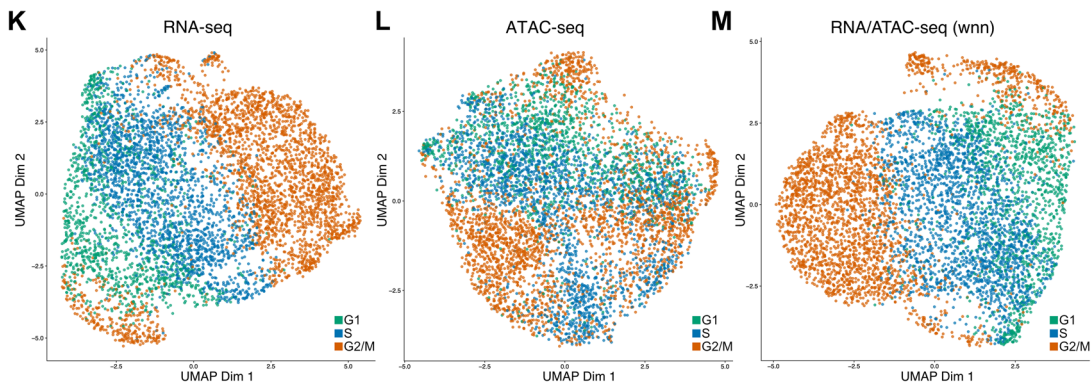
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



MitoPerturb-Seq Experiment 1 QC



MitoPerturb-Seq Experiment 1 UMAP (before cell-cycle correction)

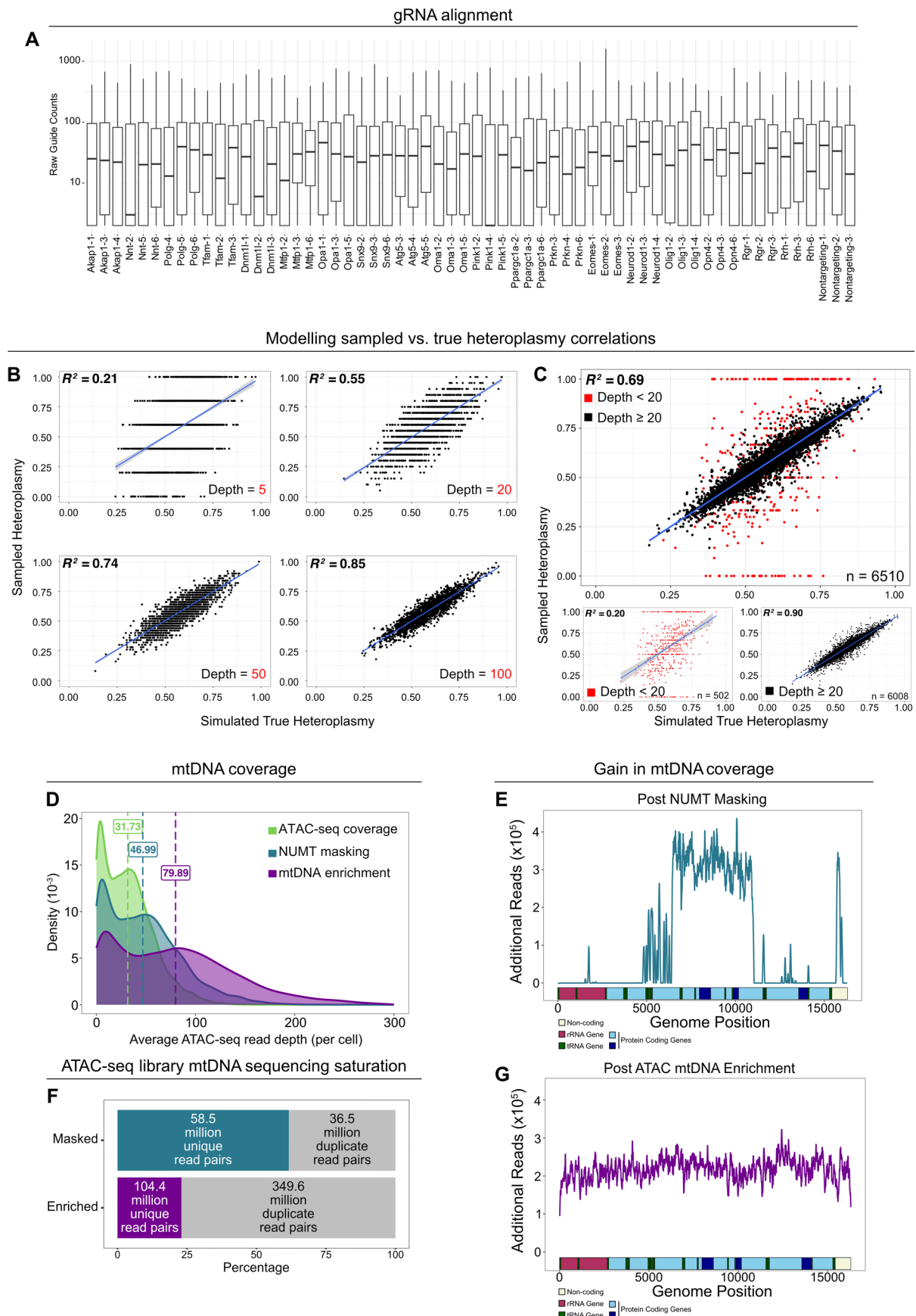


Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | MitoPerturb-Seq and quality-control steps.**

(A) Schematic of the CROPseq-RFP lentiviral cassette showing transcript expression and poly-adenylation of a gRNA-containing transcript following genomic integration. (B) Flow cytometry plots confirming genome editing efficiency in transgenic Cas9-expressing MEF clones. Clonal cells stably expressing eGFP were transduced with CROPseq-RFP lentivirus carrying two separate gRNAs targeting the eGFP sequence. RFP-high cells displayed loss of eGFP fluorescence, confirming effective CRISPR-Cas9 editing of the target gene. (C) Single-cell heteroplasmy measurements from five Cas9 transgenic MEF clones, six cells per clone. Clone 9, used for MitoPerturb-Seq experiments, is highlighted in red. Data presented as mean values  $\pm$  SD. (D) Flow cytometry plot showing RFP expression 10 days post-transduction with the pooled CROPseq-RFP gRNA library. Transduction efficiency of 35.5% (MOI = 0.35) indicates that the majority of cells received a single viral particle. (E,F) Bright field images of unfixed/unpermeabilized (E) or fixed/permeabilized (F) MEFs stained with trypan blue. Viable cells (E) have not taken up the dye and remain bright with a

well-defined membrane, while permeabilized cells (F) have taken up the dye, with darker-staining nuclei. Average cell size, calculated by the Countess cell counter, is indicated. (G) QC violins from Seurat for RNA-seq and ATAC-seq, dashed red lines indicate thresholds used for filtering, as described in the methods. Outliers with high y-axis values have been removed from selected plots: Total RNA Counts, 27 outliers; Percentage of Mitochondrial Genes, 2 outliers; Total ATAC Counts, 47 outliers; TSS Enrichment 6 outliers. (H-J) QC plots of nCountRNA (H), nCountATAC (I) and nFeatureRNA (J) versus nFeatureRNA (H) or nFeatureATAC (I, J), colored by percent mitochondrial reads, with lines indicating cutoff thresholds shown in violin plots in (G). (K-M) UMAPs showing MitoPerturb-Seq cells, clustered based on overall RNA-seq (K), ATAC-seq (L) or by Weighted Nearest Neighbor (WNN) of combined RNA- and ATAC-seq (M) data, prior to regression of cell cycle heterogeneity. Each point represents an individual cell. Cells are colored by cell cycle phase. Figure created in BioRender. Van den Aamee, J. (2026) <https://BioRender.com/j3ro7vq>.



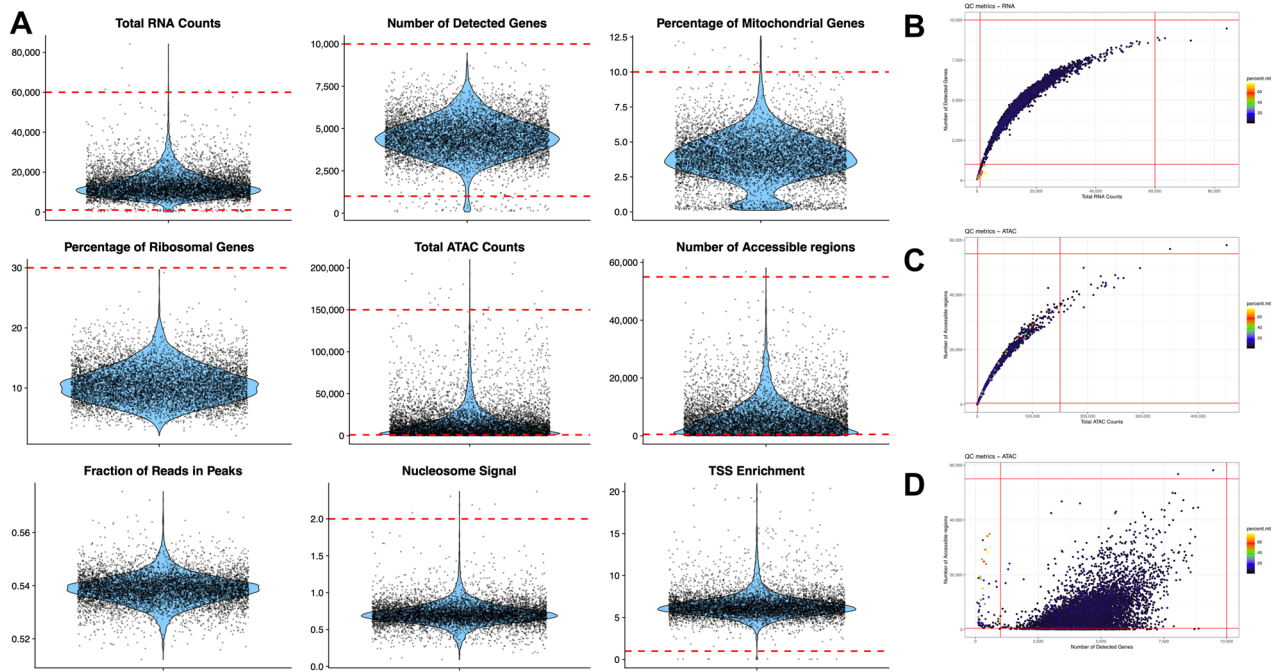
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | gRNA detection and targeted enrichment of mtDNA.**

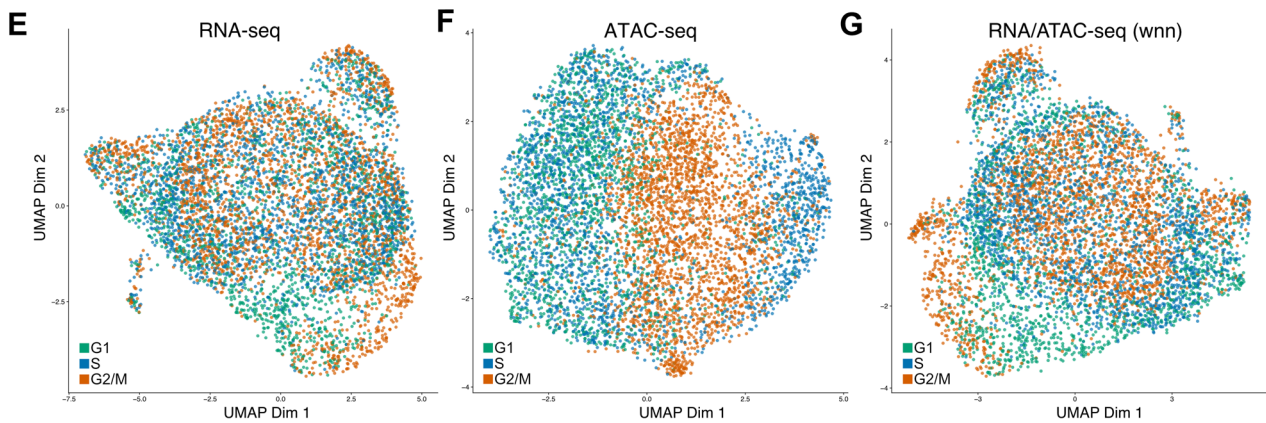
(A) Raw read counts per cell aligned to each of the gRNA sequences included in the gRNA library. For each gRNA, all cells with at least 2 counts are shown, cell numbers per gRNA group are provided in Supplementary Table 3. Boxes encompass 25<sup>th</sup>-75<sup>th</sup> percentiles, mid-line indicates median, whiskers extend to extreme values, no data points fall outside of the whisker range. (B) Sampled heteroplasmy vs. true heteroplasmy correlation at different mtDNA coverage. A population of 2,000 cells was simulated, and randomly sub-sampled at increasing sequencing depth, with sampled heteroplasmy calls showing better correlation with true heteroplasmy as sampling depth increases. Fit line displays smoothed conditional mean, shaded regions around the fit line are 95% CI. (C) Sampled heteroplasmy vs. true heteroplasmy in an *in silico* simulated version of the MitoPerturb-Seq dataset. A population of 6,510 cells was simulated, and sub-sampled using the same mtDNA coverage distribution as obtained from post-mtDNA enrichment mgatk analysis. Cells sampled at depth <20 (n = 502) are highlighted in red, confirming that a depth threshold of  $\geq 20$  successfully

removed cells with the most inaccurate heteroplasmy calls. Smaller panels plot the <20 depth and  $\geq 20$  depth populations from the same simulation separately, with corresponding  $R^2$  values and cell numbers shown. Fit line displays smoothed conditional mean, shaded regions around the fit line are 95% CI. (D) Distribution of mtDNA coverage in the MitoPerturb-Seq dataset following alignment of ATAC-seq reads to the standard mm10 reference (light green) and to a NUMT-masked version of mm10 pre- (dark green) and post- (purple) mtDNA hybridization capture-based enrichment. Mean mtDNA coverage values for each approach are indicated. (E) Line plot showing per-base additional mtDNA coverage gained when aligning ATAC-seq reads to a NUMT-masked version of the mm10 genome compared to the standard mm10 reference. (F) Percentage of unique vs duplicate sequencing reads in pre- and post-mtDNA enrichment ATAC-seq libraries. Raw read-pair numbers for each library are indicated. (G) Line plot showing per-base additional mtDNA coverage gained following hybridization capture enrichment of mtDNA-specific sequences, aligned to the NUMT-masked mm10 genome.

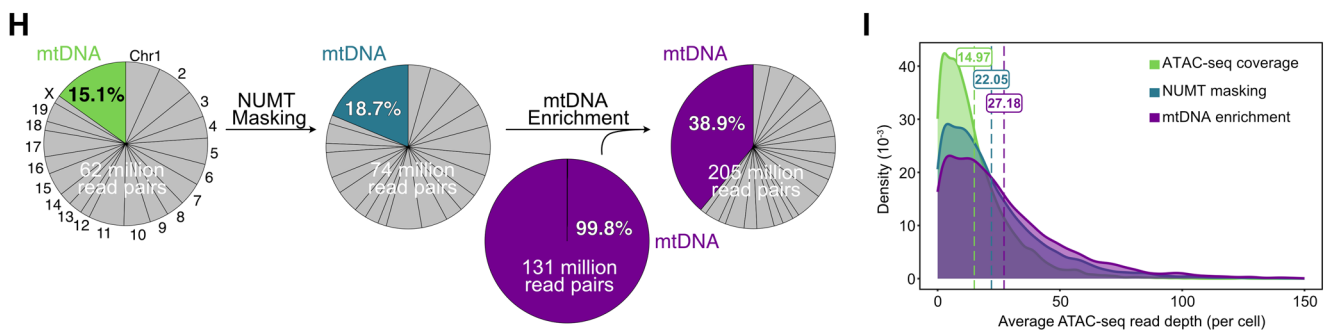
MitoPerturb-Seq Experiment 2 QC



MitoPerturb-Seq Experiment 2 UMAP after cell-cycle correction



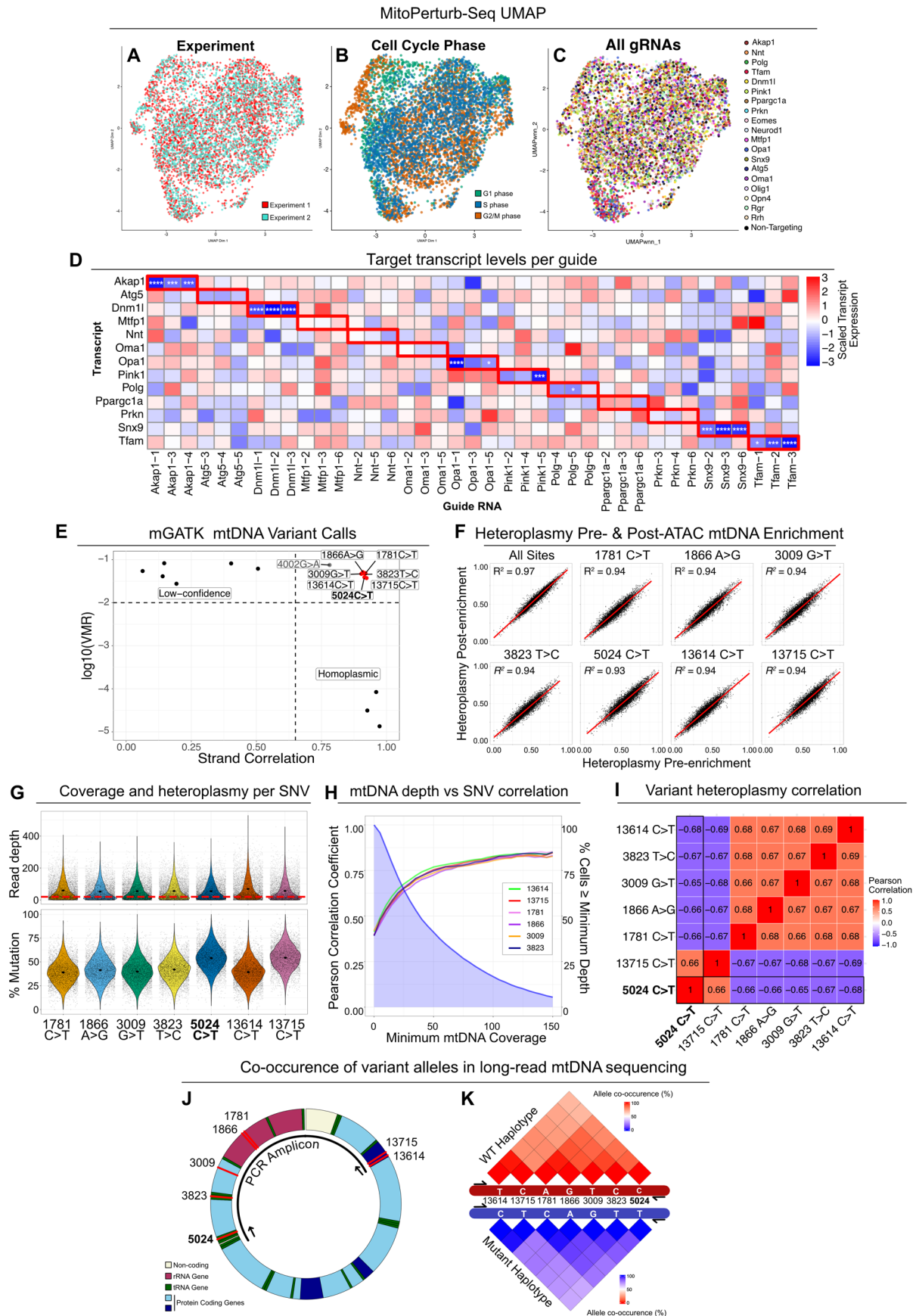
Experiment 2 ATAC-seq mtDNA enrichment



Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Quality control steps of experiment 2.** (A) QC violins from Seurat for RNA-seq and ATAC-seq, dashed red lines indicate thresholds used for filtering, as described in the methods. Outliers with high y-axis values have been removed from selected plots: Percentage of Mitochondrial Genes, 27 outliers; Total ATAC Counts, 16 outliers; TSS Enrichment, 30 outliers. (B-D) QC plots of nCountRNA (B), nCountATAC (C) and nFeatureRNA (D) versus nFeatureRNA (B) or nFeatureATAC (C,D), colored by percent mitochondrial reads, with lines indicating cutoff thresholds shown in violin plots in (A). (E-G) UMAPs showing cells from the MitoPerturb-Seq dataset in Experiment 2, clustered based on overall RNA expression patterns (E), chromatin accessibility

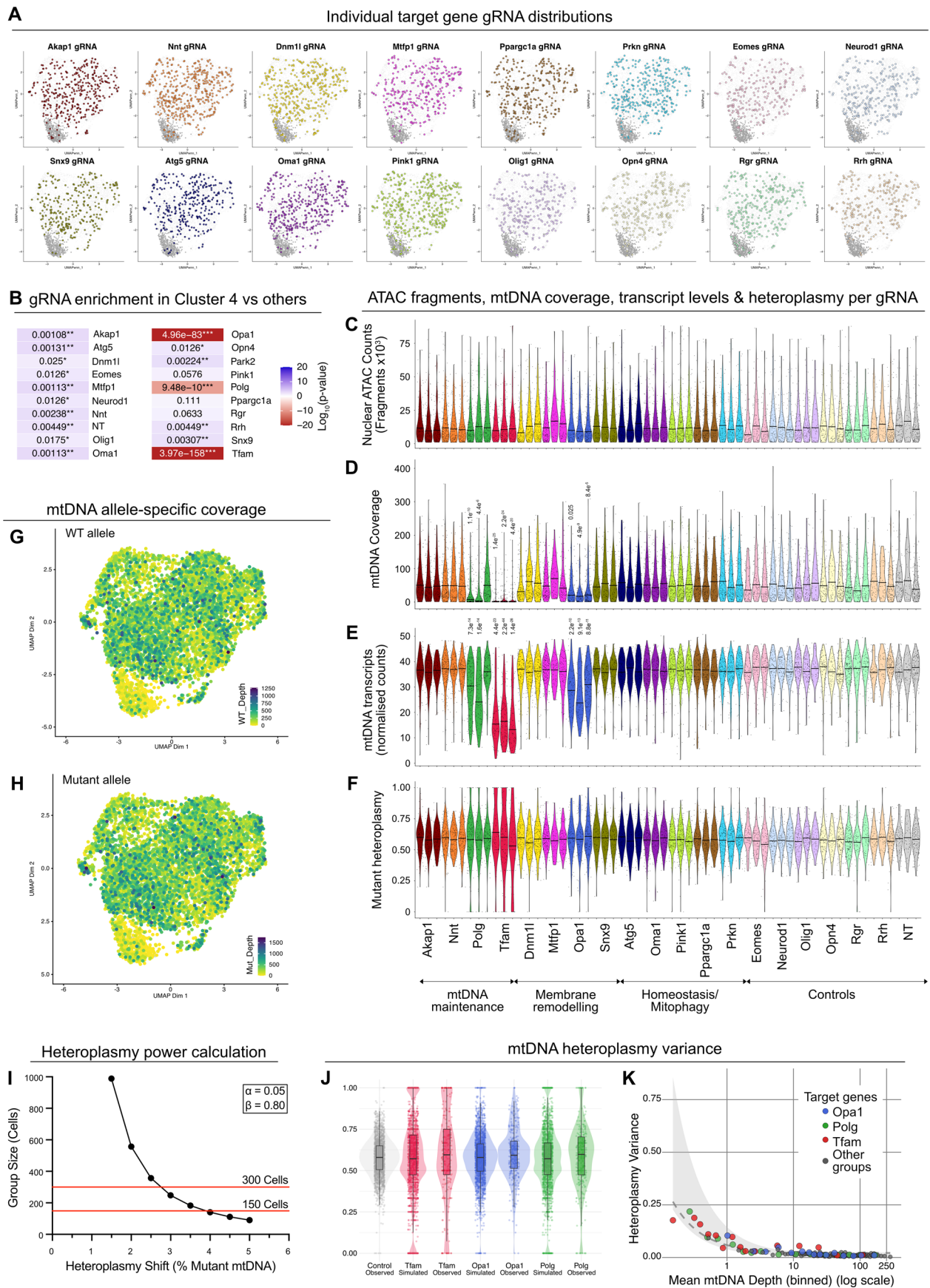
profiles (F) and a combined Weighted Nearest Neighbor (WNN) analysis (G), following regression of cell cycle heterogeneity from the RNA-seq dataset. Each point represents an individual cell. Cells are colored by cell cycle phase. (H,I) Percentage of ATAC-seq reads (H) and distribution of mtDNA coverage (I) in the replicate MitoPerturb-Seq dataset aligning to the mtDNA using the standard mm10 reference (light green), a NUMT-masked version of the mm10 genome (dark green), and following addition of enriched mtDNA reads following hybridization capture (purple). Mean coverage values for each group are indicated in (I).



Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Data-integration, gRNA efficiency, co-segregation of heteroplasmic mtDNA variants.** (A-C) UMAPs of the integrated MitoPerturb-Seq dataset, WNN clustered following cell cycle regression from RNA-seq. Cells are colored by technical replicate (A), cell cycle phase (B) and target gene gRNA assignment (C). (D) Scaled transcript levels for each target gene across all perturbation groups, split by individual gRNA assignment. Expression values normalized per-column. Two-tailed t-tests comparing target transcript expression in the relevant perturbation group vs NT gRNA group, \*  $p < 0.05$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (E) High-confidence heteroplasmic variants present in the m.5024 C > T MEFs, identified by mgatk. Seven variants (red) were previously reported, with an additional variant, m.4002 G > A, likely to be a clone-specific *de novo* mutation not present in the mouse strain. (F) Correlation of per-cell heteroplasmy calls pre- and post-mtDNA enrichment from the scATACseq library at read-depth  $\geq 20$  for each individual SNV and all 7 sites combined. (G) Per-cell read depth (top) and heteroplasmy (bottom) at each of the seven main variant sites (SNVs). Heteroplasmy distributions match the reported

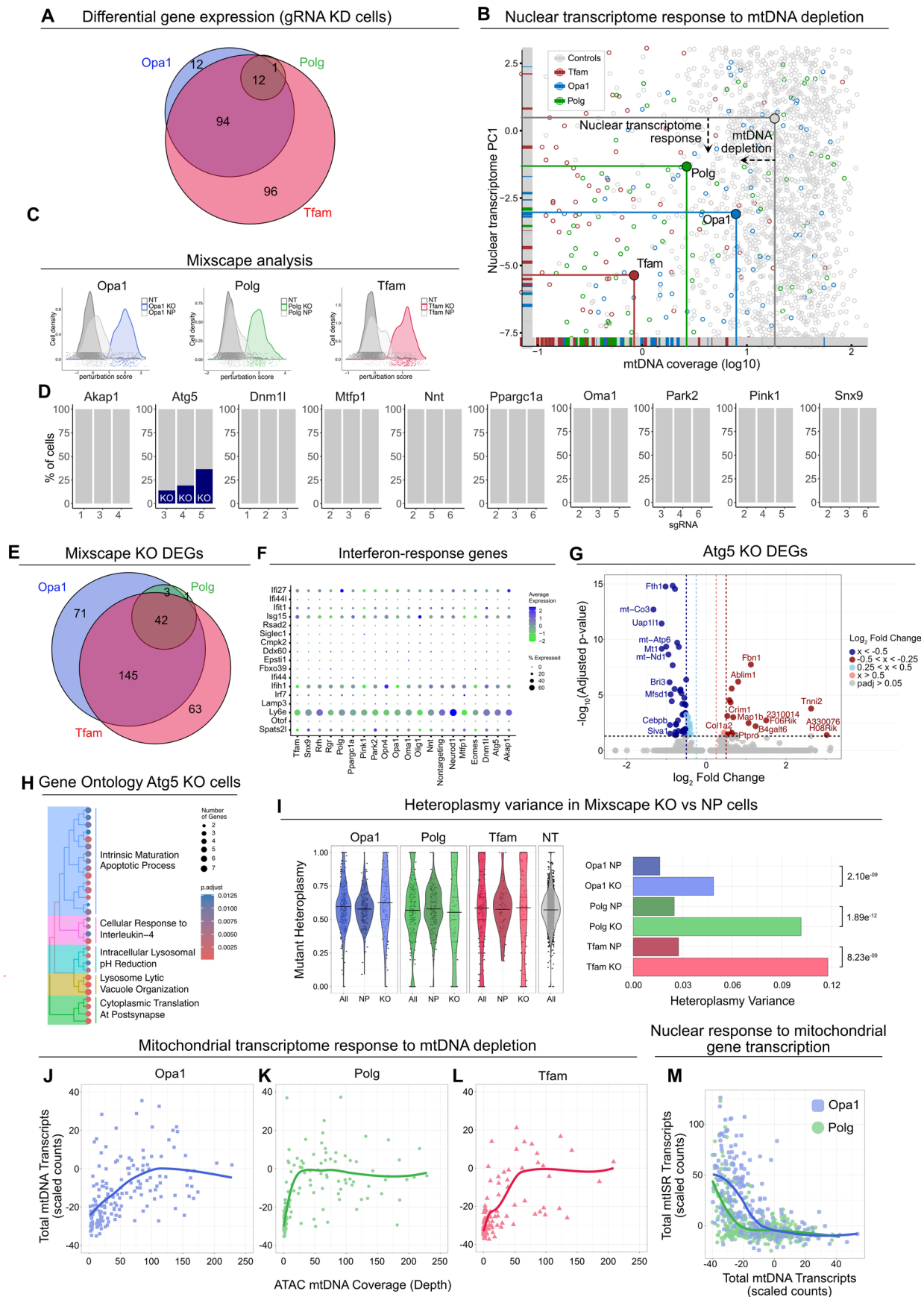
linkage of m.5024 C > T *in cis* with m.13715 C > T and *in trans* with m.1781 C > T, m.1866A > G, m.3009 G > T, m.3823 T > C and m.13614 C > T<sup>58</sup>. On the heteroplasmy plot, only cells with mtDNA coverage  $\geq 20$  (indicated by the red line on the upper panel) at the corresponding SNV position are shown. (H) Pearson correlation coefficients (left axis) between heteroplasmy calls at each SNV and m.5024 C > T with increasing minimum read depth threshold (horizontal axis). Blue shaded area indicates the percentage of cells above the depth threshold (right axis). (I) Pearson correlation between all variant sites with minimum mtDNA depth threshold  $\geq 20$ . (J,K) Schematic of the mtDNA showing the location of the seven main SNVs and the PCR amplicon (J) used for long-read sequencing of mtDNA from skin biopsies of 2-week-old m.5024 C > T mice (1 male & 2 female) (K). Percentage of long-reads with co-occurrence of each allele with any of the other alleles on the same haplotype (WT, red, top; mutant, blue, bottom). Co-occurrence is less strong for distant SNVs, due to PCR-mediated recombination. SNV positions are highlighted in red.



Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | MitoPerturb-seq identifies mtDNA depletion following targeted gene perturbation.** (A) Cells assigned to individual gRNA groups (NT, *Tfam*, *Opa1* and *Polg* UMAPs shown in Fig. 2f-i). (B) Relative enrichment of individual gRNA groups in cluster 4 (low mtDNA CN) compared to all other clusters. Target genes highlighted red are over-represented and blue under-represented in cluster 4. Two-sided chi-squared tests with multiple testing correction (Benjamini-Hochberg) and odds ratio calculations. All adjusted p-values shown; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . NB: Scaled colouring is capped at  $p = 1 \times 10^{-20}$  for clarity, exact p-values are on plots. (C-F) Per-cell nuclear ATAC fragment count (C) mtDNA coverage (D), mtDNA transcript levels (E) and heteroplasmy (F) for each gRNA group, including controls. Each violin represents one of three gRNAs targeting the corresponding gene. Pairwise two-sided Welch's t-tests with multiple testing correction (Bonferroni), p-values in figure, all values  $< 0.05$  shown. Black horizontal lines indicate median (C-E) or mean (F). Plots include all cells in the final integrated dataset, without additional filtering. (G,H) Per-cell mtDNA coverage of the WT (G) and mutant (H) allele (no depth cut-

off applied). (I) Group size required to confidently identify mean heteroplasmy shifts at power  $> 0.8$  and  $p < 0.05$ . A mean group size of 179 would suffice to identify mean heteroplasmy shifts of 3.5-4%, assuming all cells shifted in the same direction. (J) Observed heteroplasmy distributions for *Tfam*, *Opa1* and *Polg* KO compared to simulated distributions generated by binomial downsampling of control cells to matched read depths ( $n = 5,000$ ). Boxes 25<sup>th</sup>-75<sup>th</sup> percentiles, mid-line median, whiskers extend to most extreme datapoints within  $1.5 \times$  IQR from hinges, all datapoints plotted. (K) Relationship between mtDNA depth and heteroplasmy variance. Single cells aggregated into read-depth bins of equal sample size to calculate local variance. Gray points without mtDNA depth reduction (all gRNAs excluding *Tfam*, *Polg*, and *Opa1*); coloured points represent *Tfam*, *Polg*, and *Opa1* perturbation groups. The gray dashed line represents the 95<sup>th</sup> percentile fit from quantile regression on control cells (95% of control cells have technical variance at or below this line at a given depth). The gray ribbon represents bootstrapped 95% confidence intervals for the expected variance range modelled on the non-depleted population.

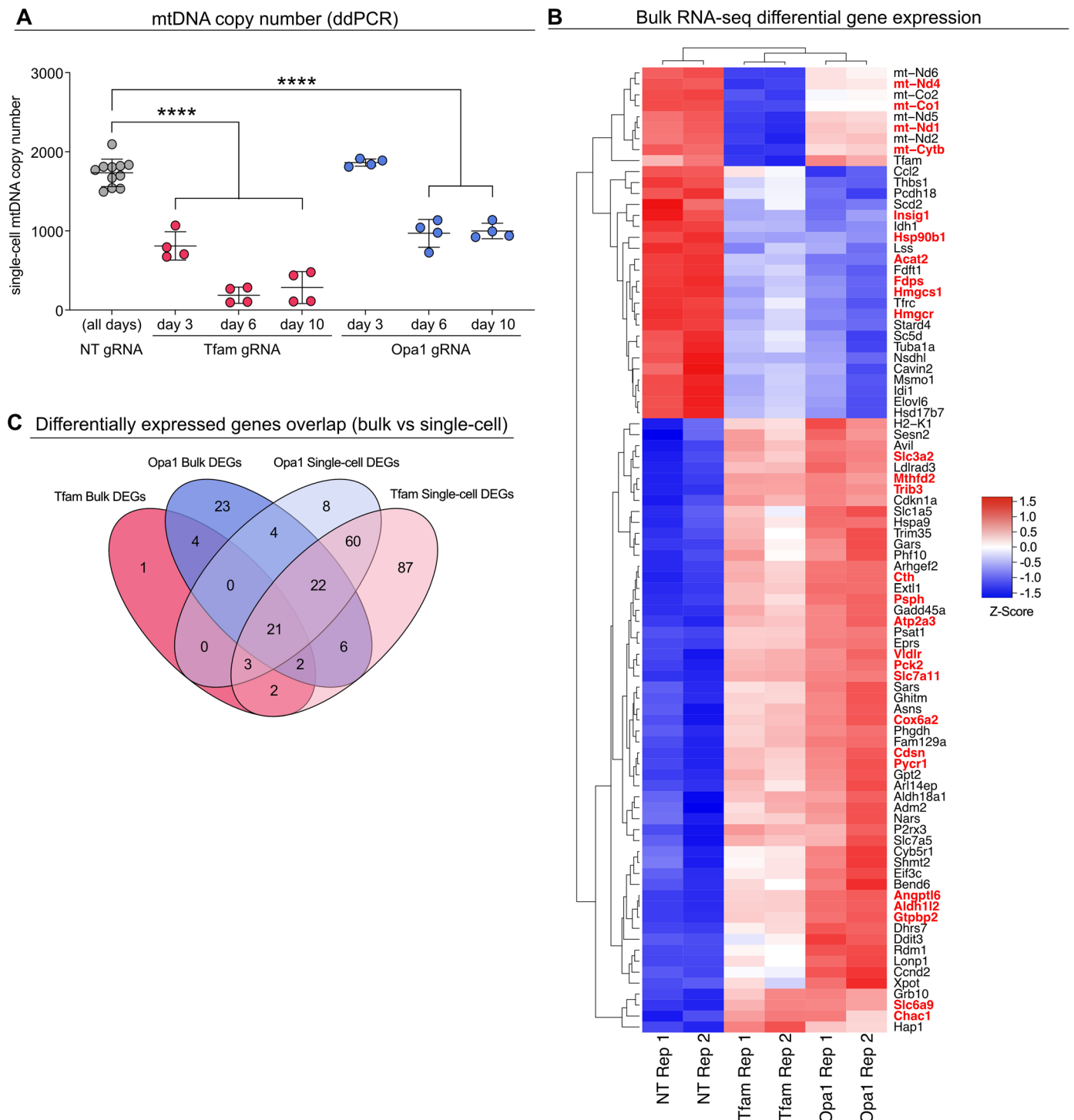


Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Differential gene expression upon gRNA perturbation.**

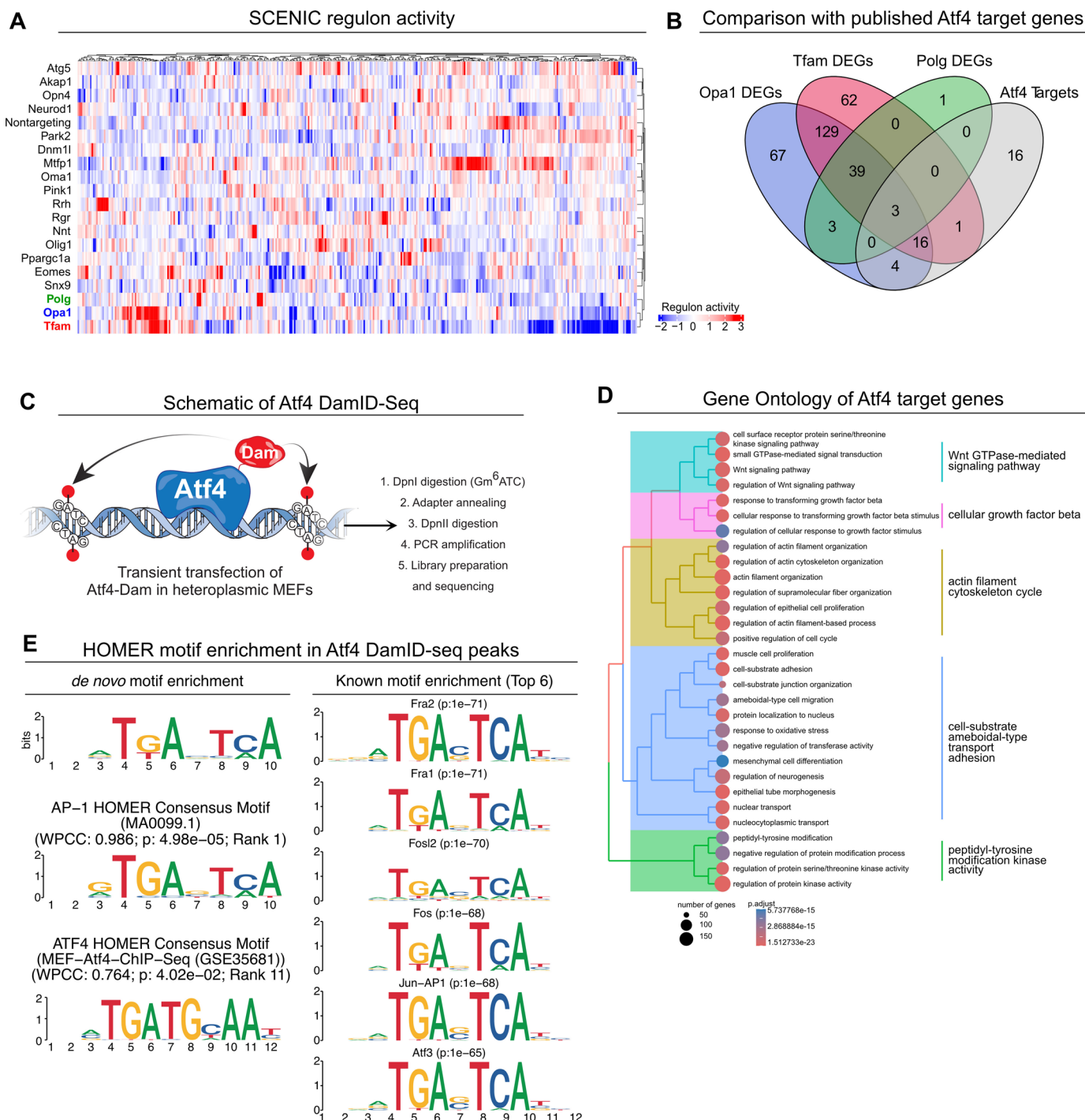
(A) Overlap of DEGs identified in *Tfam*, *Opa1* and *Polg* perturbation groups. (B) Strength of principal component (PC1) of nuclear transcriptome in all cells plotted against mtDNA coverage (log<sub>10</sub> transformed), with *Polg*, *Tfam* and *Opa1* perturbation groups and Controls highlighted, before Mixscape. (C,D) Perturbation scores of cells in the *Tfam*, *Opa1*, *Polg* gRNA groups (C) and proportion of cells assigned one of the three gRNAs targeting the indicated gene (D) following Mixscape. Remaining cells (grey) assigned as non-perturbed (NP). (E) Overlap of DEGs identified in the *Tfam*, *Opa1*, *Polg* perturbation groups after Mixscape. (F) Average expression of interferon-response genes in cells across the various perturbation groups, indicating overall low expression of these genes, and no activation. (G) Significantly differentially expressed genes (adjusted p-value < 0.05) with positive/negative log<sub>2</sub> fold change of >0.25 (light colours) and >0.5 (dark colours) in *Atg5* KO cells compared to controls. Top 10 up- and down-regulated genes labelled. Two-sided Wilcoxon rank-sum test with Bonferroni correction for multiple comparisons. Full list of DEGs in

Supplementary Table 5. (H) Significantly enriched Gene Ontology (GO) terms, grouped by biological process, identified in the *Atg5* KO group following Mixscape. One-sided Fisher's exact test with Benjamini-Hochberg correction for multiple comparisons. Full details of identified GO terms are in Supplementary Table 6. (I) Distribution of mtDNA heteroplasmy (without mtDNA depth cut-off applied) in control cells and in *Opa1*, *Polg* and *Tfam* perturbation groups before and after assignment to KO and NP groups by Mixscape. The corresponding bar plot shows heteroplasmy variance in KO vs NP cells for each perturbation group. KO vs NP comparisons are Brown-Forsythe tests (one-way ANOVA) with multiple testing correction (Holm). (J-L) Per-cell mtDNA transcript levels against mtDNA coverage for *Opa1* (J), *Polg* (K) and *Tfam* (L) perturbation groups. (M) mtDNA transcript levels against combined transcript levels of genes involved in the mitochondrial integrated stress response in *Opa1* and *Polg* perturbation groups, indicating increased activation of the stress response in *Opa1* KO cells compared to *Polg* KO cells at similar mtDNA transcript levels.



**Extended Data Fig. 7 | Bulk RNA-seq upon targeted perturbation of Tfam and Opa1.** (A) Timecourse of mtDNA CN in m.5024 C > T MEFs (Clone 8 in Extended Data Fig. 1C, at 45% mean heteroplasmy) at 3 days, 6 days and 10 days following transduction with non-targeting (NT, grey), *Tfam* (red) or *Opa1* (blue) gRNAs. Each point represents a 20-cell bulk ddPCR measurement normalized to cell number (n = 11 measurements for NT gRNA and 4 measurements for all other groups). Data presented as mean values  $\pm$  SD. One-way ANOVA with Tukey's

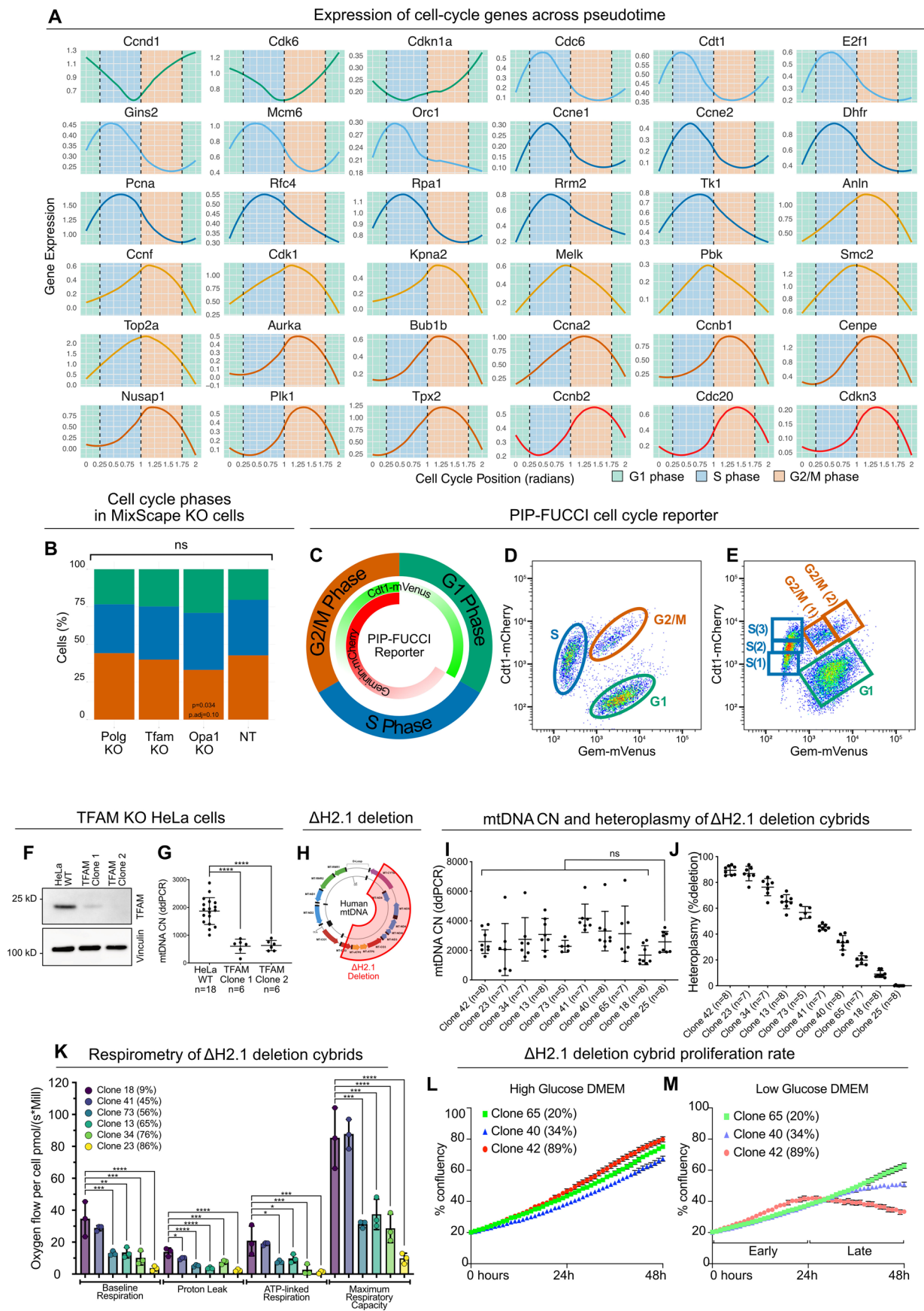
post-hoc test, pairwise comparisons to NT control, \*\*\*\*p < 0.0001. (B) Relative expression of DEGs in bulk RNA-seq data from *Tfam*- or *Opa1*-gRNA cells compared to non-targeting controls. All genes shown were significantly differentially expressed (FDR < 0.05 and log<sub>2</sub>-fold-change > 0.25) in at least one gRNA condition; genes highlighted in red were differentially expressed in both *Tfam*- and *Opa1*-gRNA cells. (C) Overlap of DEGs identified in bulk RNA-seq and MitoPerturb-Seq *Tfam* and *Opa1* KD cells.



**Extended Data Fig. 8 | ATF4 DamID-seq in heteroplasmic MEFs.**

(A) Regulon activity following SCENIC analysis. Average AUC scores were calculated for each gRNA group; all regulons identified are shown and also listed in Supplementary Table 7. (B) Intersection of *Tfam*, *Opa1* and *Polg* KO MitoPerturb-Seq DEGs (post-Mixscape analysis) with a published list of high-confidence ATF4 target genes. (C) Schematic of ATF4-DamID-seq in heteroplasmic MEFs. Adapted, with permission, from ref. 87., under a Creative Commons Attribution 4.0 International License. (D) Significantly enriched Gene Ontology (GO) terms, grouped by biological process, identified in the ATF4

target genes. One-sided Fisher's exact test with Benjamini-Hochberg correction for multiple comparisons. (E) Motif enrichment analysis of ATF4 DamID-seq peaks. *De-novo* motif discovery (left) identified a top-ranked motif matching the AP-1 consensus (Fos::Jun; Rank 1) as well as the canonical ATF4 motif in the HOMER database, representative of ATF4-CEBPG heterodimer binding (Rank 11). Known motif enrichment (right) shows high prevalence of AP-1 family members. P-values denote significance of enrichment; WPCC, Weighted Pearson correlation coefficient.

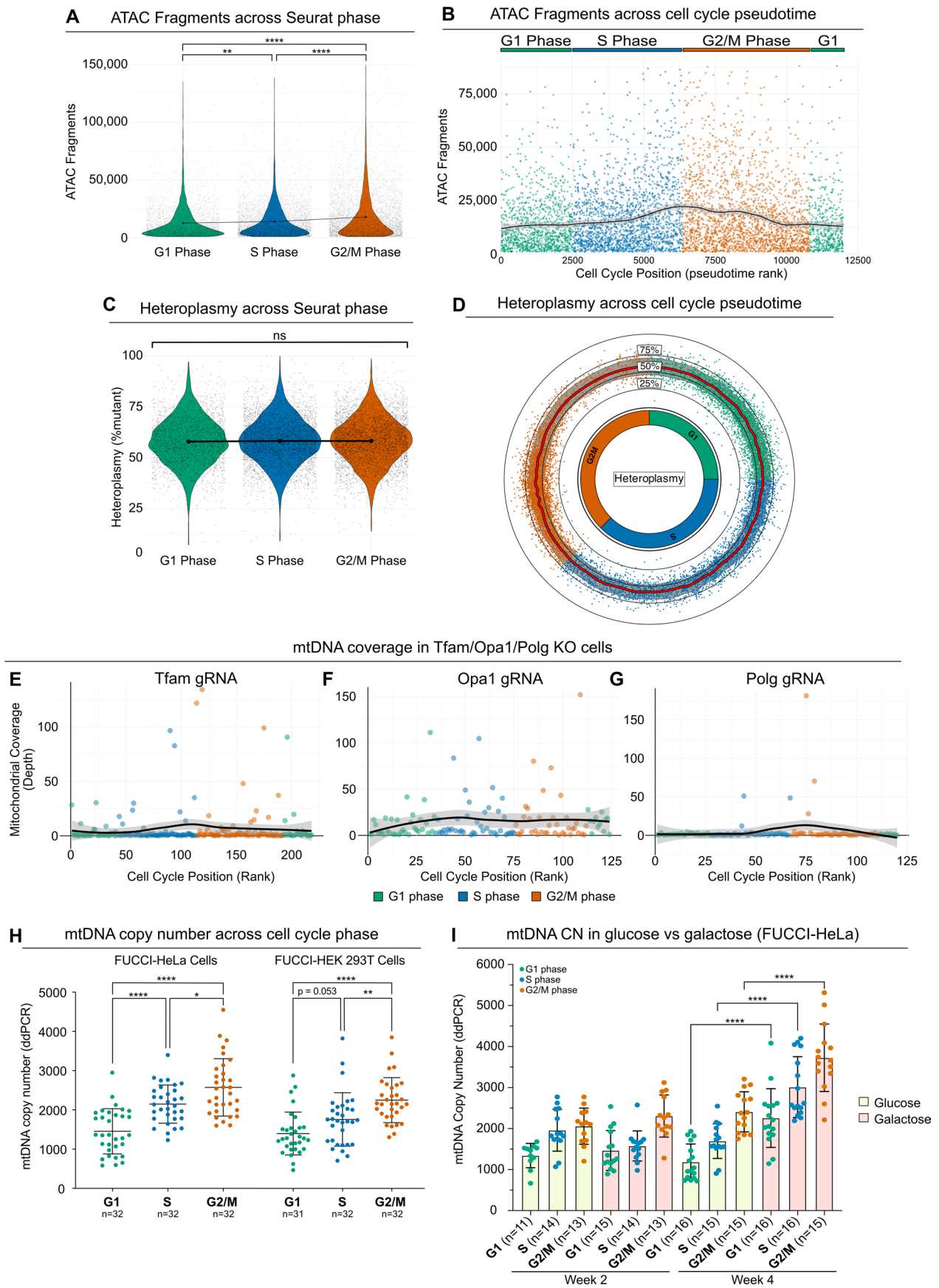


Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Cell cycle slowing in response to mtDNA depletion.**

(A) Expression profiles of indicated cell-cycle genes across cell-cycle pseudotime determined by tricycle. (B) Percentage of cells in each cell-cycle phase in *Tfam*, *Opa1*, *Polg* KO cells, post-Mixscape, compared to non-targeting gRNA cells. Chi-squared tests with multiple testing correction (Bonferroni); ns, not significant. (C) Schematic showing the expression patterns of the PIP-FUCCI fluorescent reporters Cdt1-mVenus and Geminin-mCherry across the cell-cycle. (D,E) Gating strategies used to isolate cells for mtDNA CN measurements based on cell-cycle phase for both inter-phase (D) and intra-phase (E) comparisons. (F,G) Western blot (F) and ddPCR measurements (G) showing KD of TFAM protein (F) and reduced mtDNA CN (G) in HeLa cells transduced with two independent *TFAM* gRNAs. Vinculin is shown as a loading control (F). Number of individual cells per group is indicated in (G), data presented as mean values  $\pm$  SD. One-way ANOVA with Tukey's post-hoc test, \*\*\*\*  $p < 0.0001$ . (H) Map of the human mtDNA, highlighting the 7.5 Kb deletion present in the DeltaH2.1 cybrid cell line. (I,J) mtDNA CN (I) and heteroplasmy (J) of eight clonal populations isolated from

the DeltaH2.1 cybrid cell line. Cell numbers per group indicated on graphs. Data presented as mean values  $\pm$  SD. One-way ANOVA of the mtDNA CN distributions was significant ( $p = 0.015$ ); pairwise comparisons of all heteroplasmic clones to Clone 25 (0% heteroplasmy) using Tukey's post-hoc test; ns, not significant. (K) Respirometry of DeltaH2.1 cybrid cells with increasing heteroplasmy levels, using Oroboros, showing baseline respiration (BR), proton-leak (PL), ATP-linked respiration (AR) and maximal respiratory capacity (MC). Three separate assays per clone. Data presented as mean values  $\pm$  SD. One-way ANOVA with Tukey's post-hoc test performed for each tested parameter, all significant pairwise comparisons are shown. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ . (L,M) Growth curves of DeltaH2.1 cybrid clones carrying different levels of heteroplasmy in high (L) and low (M) glucose culture medium. Confluency values averaged over 8 individual images per timepoint for each clone in each condition, data presented as mean values  $\pm$  SEM. Figure created in BioRender. Van den Aemele, J. (2026) <https://BioRender.com/j3ro7vq>.



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 | Relaxed replication of mtDNA. (A,B)** Nuclear ATAC fragment counts across cell cycle stages using Seurat **(A)** or across cell cycle pseudotime based on tricycle pi score **(B)**. Fit line for **(B)** is a smoothed LOESS regression (span = 0.2), shaded regions around the fit line indicate 95% CI. One-way ANOVA with Tukey's post-hoc test, \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ . **(C,D)** Heteroplasmy levels across cell cycle stages using Seurat **(C)** or across cell cycle pseudotime based on tricycle pi score **(D)**. Red line = mean, grey band = standard deviation. One-way ANOVA with Tukey's post-hoc test; ns, not significant. **(E-G)** mtDNA coverage across cell cycle pseudotime with cells ranked based on tricycle pi score for the *Tfam* **(E)**, *Opa1* **(F)** and *Polg* **(G)** perturbation groups. Fit lines are smoothed LOESS regressions (span = 0.7), shaded regions

around the fit lines indicate 95% CI. **(H)** Single-cell mtDNA CN measurements in PIP-FUCCI-expressing HeLa and HEK293T cells following flow sorting based on cell cycle phase. Cell numbers per group are indicated on graph. Data presented as mean values  $\pm$  SD. One-way ANOVA with Tukey's post-hoc test performed across cell-cycle phases for each cell type, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*\*  $p < 0.0001$ . **(I)** Single-cell mtDNA CN measurements in PIP-FUCCI-expressing HeLa cells following flow sorting based on cell cycle phase. Cells cultured in DMEM supplemented with either high-glucose or galactose for four weeks. Cell numbers per group indicated on graph. Data presented as mean values  $\pm$  SD, one-way ANOVA with Tukey's post-hoc test per cell-cycle phase across time points and culture conditions, \*\*\*\*  $p < 0.0001$ .

## Reporting Summary

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

### Statistics

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

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

- |                 |  |
|-----------------|--|
| Data collection | NovaSeq Control Software v1.7.5  |
| Data analysis   | <p>All software packages used for data analysis are listed and referenced in the methods section, a list with software versions is provided below. Complete code used for bespoke data analysis are available via GitHub (<a href="https://github.com/JvdAlab/mitoPerturb-Seq">https://github.com/JvdAlab/mitoPerturb-Seq</a>).</p> <p>R Statistical Software (v4.3.3), Python (v3.12.2), Seurat (v5.1.0), Signac (v1.13.0), pySCENIC (v0.12.1), AUCCell (v1.28.0), tricycle (v1.12.0), Cell Ranger ARC (v2.0.1), mgatk (v0.7.0), UMI-Tools (v1.0.1), FastQC (v0.11.9), FastQ Screen (v0.14.1), Qualimap (v2.2.1), BWA (v0.7.17), Bowtie2 (v2.4.5), clusterProfiler (v4.12.1), enrichR (v3.2), Trimmomatic (v0.39), RUM (v2.0.4), HTSeq (v2.0.3), edgeR (v4.4.2), damidseq_pipeline (v1.5.3), MACS (v3.0.3), IDR (v2.0.3), bedtools (v2.31.0), HOMER (v5.1), SuperExactTest (v1.1.2), ggplot2 (v3.5.1), ggVennDiagram (v1.5), enrichplot (v1.10.1), deepTools (v3.5.6), pyGenomeTracks (v3.9), ggalluvial (v0.12.5), ggpmisc (v0.6.1), ggpubr (v0.6.0), ggvenn (v0.1.10), ComplexHeatmap (v2.22.0), Pheatmap (v1.0.12), quantreg (v6.1), PyroMark Analysis Software (QIAGEN), FlowJo (v10), QX Manager (Bio-Rad)</p> |

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

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

The 10X Multiome Mitoperturb-Seq, Bulk RNAseq and DamID-seq data produced in this study have been deposited into the Gene Expression Omnibus database and are accessible via the following accession numbers:

10X Multiome ATAC & GEX (MitoPerturb-seq): GSE297418

Tfam/Opa1 CRISPR bulk RNAseq: GSE297416

Atf4 DamID-seq: GSE297491

## Research involving human participants, their data, or biological material

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

Reporting on sex and gender

No reporting on sex and gender is included in the study.

Reporting on race, ethnicity, or other socially relevant groupings

No reporting on race, ethnicity, or other socially relevant groupings is included in the study.

Population characteristics

See above.

Recruitment

No recruitment was involved in this study.

Ethics oversight

No ethical approval was required for the study.

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

## Field-specific reporting

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

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size

Mitoperturb-Seq on MEF cells yielded 11,990 single-cells with an average of 328 cells assigned to each CRISPR gene knockout group, each cell represents an individual biological replicate in this experiment.

Data exclusions

QC of single-cell RNAseq & ATACseq data was performed to remove poor-quality cells from the final dataset, standard thresholds for filtering single-cell data were employed and are described in the methods section "Initial Analysis & Quality Control".

Replication

Mitoperturb-Seq was performed in technical duplicate to yield the 11,990 single-cells described above, no significant batch effects were observed between the two replicates. Bulk RNAseq validation experiments were performed in biological duplicate (two independent transductions per guide RNA) using validated gRNA sequences from the Mitoperturb-Seq pooled library, and were performed in an independent population of MEF cells to those used for the Mitoperturb-Seq experiments, no significant batch effects were seen between the biological replicates upon analysis of the data.

Randomization

Cells in the Mitoperturb-Seq experiments were randomly allocated to gRNA groups by the nature of the CRISPR screening process, therefore we did not need to actively randomise these samples.

Blinding

Blinding was not necessary due to the random nature of the CRISPR gRNA transduction, as described above.

## Reporting for specific materials, systems and methods

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

## Materials & experimental systems

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

## Methods

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

## Antibodies

Antibodies used

Anti-TFAM (D5C8) Rabbit mAb, Cell Signaling #8076S, used at 1:1000 dilution.  
 Anti-Vinculin Mouse mAb, Sigma #V4505, used at 1:1000 dilution.  
 Anti-Mouse Ig/HRP Rabbit pAb, Agilent #P0260, used at 1:1000 dilution.  
 Anti-Rabbit Ig/HRP Swine pAb, Agilent #P0217, used at 1:1000 dilution.

Validation

Anti-TFAM (D5C8) Rabbit mAb has been validated by the manufacturer for use in western blot with confirmed cross-reactivity to human TFAM (<https://www.cellsignal.com/products/primary-antibodies/tfam-d5c8-rabbit-mab/8076>), in the study we included cell lysate from WT HeLa cells as a positive control for TFAM staining when assessing CRISPR KO efficiency.  
 Anti-Vinculin Mouse mAb has been validated by the manufacturer for use in western blot with confirmed cross-reactivity to human VCL (Vinculin, <https://www.sigmaaldrich.com/GB/en/product/sigma/v9131>)

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

m.5024C>T Mouse Embryonic Fibroblasts were isolated from a female embryo and immortalized by Dr JB Stewart, University of Newcastle  
 WT HeLa cells, originally from a female patient, were purchased from the European Collection of Authenticated Cell Cultures, catalogue number 93021013  
 HEK 293T cells, originally from a female fetus, were purchased from Takara (Lenti-X 293T Cells), catalogue number 632180  
 DeltaH2.1 cybrid cells were generated by Dr CT Moraes by fusing 143B(TK-) osteosarcoma cells (female) with enucleated patient-derived fibroblast cells harboring a 7.5 kb partial deletion in the mtDNA, the source of these cells is referenced in the manuscript.

Authentication

The MEF cells were authenticated by the presence of 7 specific SNPs in the mtDNA sequence, which were confirmed as part of the analytical pipeline employed in the study.  
 Human cell lines used in the study were purchased from verified sources:  
 WT HeLa cells - ECACC, #93021013  
 HEK 293T cells - Takara, Lenti-X 293T #632180  
 DeltaH2.1 cybrid cells were authenticated by the presence of the 7.5Kb partial deletion in the mtDNA, which was confirmed by ddPCR assays performed in the study.

Mycoplasma contamination

All cell lines used in this study were tested for mycoplasma and were confirmed to be negative prior to performing experiments.

Commonly misidentified lines  
 (See [ICLAC](#) register)

None of the cell lines used appear in version 13 of the ICLAC Register of Misidentified Cell Lines

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

The mouse line used in this study was m.5024C>T (Allele symbol: mt-Tam1Jbst, MGI ID: 5902095), bred on the C57BL/6 background

Wild animals

The study did not involve wild animals.

Reporting on sex

No reporting on sex is included in the study.

Field-collected samples

The study did not involve samples collected in the field.

Ethics oversight

Mouse husbandry was performed in a Home Office-designated facility, according to the UK Home Office guidelines upon approval by

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

## Plants

Seed stocks

Plants were not used in this study, as highlighted under Materials &amp; experimental systems

Novel plant genotypes

Plants were not used in this study, as highlighted under Materials &amp; experimental systems

Authentication

Plants were not used in this study, as highlighted under Materials &amp; experimental systems

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation

For CROP-RFP expressing MEF cells, and FUCCI expressing HeLa/HEK293T/DeltaH2.1 cells, samples were prepared from in vitro cell cultures by first trypsinising cells, followed by washing 2x in PBS, followed by resuspension in 500ul PBS per sample. Finally, cells were passed through a 50um filter prior to analysis/sorting on the BD FACSMelody Cell Sorter

Instrument

BD FACSMelody Cell Sorter

Software

FlowJo (v10)

Cell population abundance

For the CROP-RFP MEF cell sorts, cells were sorted in 'single-cell' mode and selected for expression of the RFP marker carried on the gRNA lentiviral backbone, since this marker is only expressed following successful integration of the gRNA cassette, this ensured that the post-sort populations collected for MitoPerturb-seq fixation/permeabilisation were highly enriched (close to 100% pure) for single cells expressing a CRISPR gRNA.

Gating strategy

For all flow cytometry experiments, cell populations were initially gated based on 2-dimensional FSC-A/SSC-A signal intensity to select the live-cell population, further gating based on SSC-H/SSC-A was applied to select single cells and exclude doublets. For CROP-RFP MEF sorts, RFP-positive cells were gated by visualising RFP signal on a histogram and applying a sort gate based on the clear division between the RFP -ve and RFP +ve cell populations. For FUCCI HeLa cell sorts, FUCCI signal was visualised on a 2-dimensional mVenus/mCherry plot, with sort gates placed around the G1, S & G2M cell populations as outlined in the manuscript.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.