nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-024-01225-6

Single-nucleoid architecture reveals heterogeneous packaging of mitochondrial DNA

Received: 14 October 2022

Accepted: 9 January 2024

Published online: 12 February 2024

Check for updates

R. Stefan Isaac[1], Thomas W. Tullius[1,4], Katja G. Hansen[1,4], Danilo Dubocanin[2], Mary Couvillion[1], Andrew B. Stergachis[2,3] ✉ & L. Stirling Churchman[1] ✉

Cellular metabolism relies on the regulation and maintenance of mitochondrial DNA (mtDNA). Hundreds to thousands of copies of mtDNA exist in each cell, yet because mitochondria lack histones or other machinery important for nuclear genome compaction, it remains unresolved how mtDNA is packaged into individual nucleoids. In this study, we used long-read single-molecule accessibility mapping to measure the compaction of individual full-length mtDNA molecules at near single-nucleotide resolution. We found that, unlike the nuclear genome, human mtDNA largely undergoes all-or-none global compaction, with most nucleoids existing in an inaccessible, inactive state. Highly accessible mitochondrial nucleoids are co-occupied by transcription and replication components and selectively form a triple-stranded displacement loop structure. In addition, we showed that the primary nucleoid-associated protein TFAM directly modulates the fraction of inaccessible nucleoids both in vivo and in vitro, acting consistently with a nucleation-and-spreading mechanism to coat and compact mitochondrial nucleoids. Together, these findings reveal the primary architecture of mtDNA packaging and regulation in human cells.

Originating from a eubacterial ancestor, mitochondria have retained a small (16.5 kb) circular genome that is present across the mitochondrial network in hundreds to thousands of copies per cell[1]. Human mtDNA encodes 13 core subunits of the oxidative phosphorylation complexes that are assembled with nuclear-encoded subunits at the mitochondrial inner membrane. Most of the mitochondrial genome is coding, with two polycistronic transcripts originating from transcription start sites that direct transcription along either the heavy or light strand; the names of the two strands reflect their distinct sedimentation properties (Fig. 1a). The primary noncoding region (NCR) of mtDNA contains an origin of replication (O_H) and a displacement loop (D-loop) with unknown function. The D-loop is present in a fraction of human mtDNA molecules and contains a 400–650 nucleotide third strand of DNA, the 7S DNA, whose synthesis depends on the transcription and replication machinery[2–9]. Mutations in mtDNA and its associated proteins cause numerous inherited and acquired diseases, and misregulation of mtDNA expression is implicated in neurodegenerative disorders, cancers and aging-related illnesses[10–16].

Although mtDNA has a contour length of over 5 µm, super-resolution imaging studies have shown that it is compacted into 100-nanometer nucleoprotein complexes called nucleoids[17,18]. However, the principles balancing mitochondrial nucleoid compaction, replication and transcription remain poorly understood. Unlike nuclear chromosomes, mtDNA is not packaged with histones and

[1]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [2]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA. [3]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [4]These authors contributed equally: Thomas W. Tullius, Katja G. Hansen. ✉e-mail: absterga@uw.edu; churchman@genetics.med.harvard.edu

https://doi.org/10.1038/s41594-024-01225-6

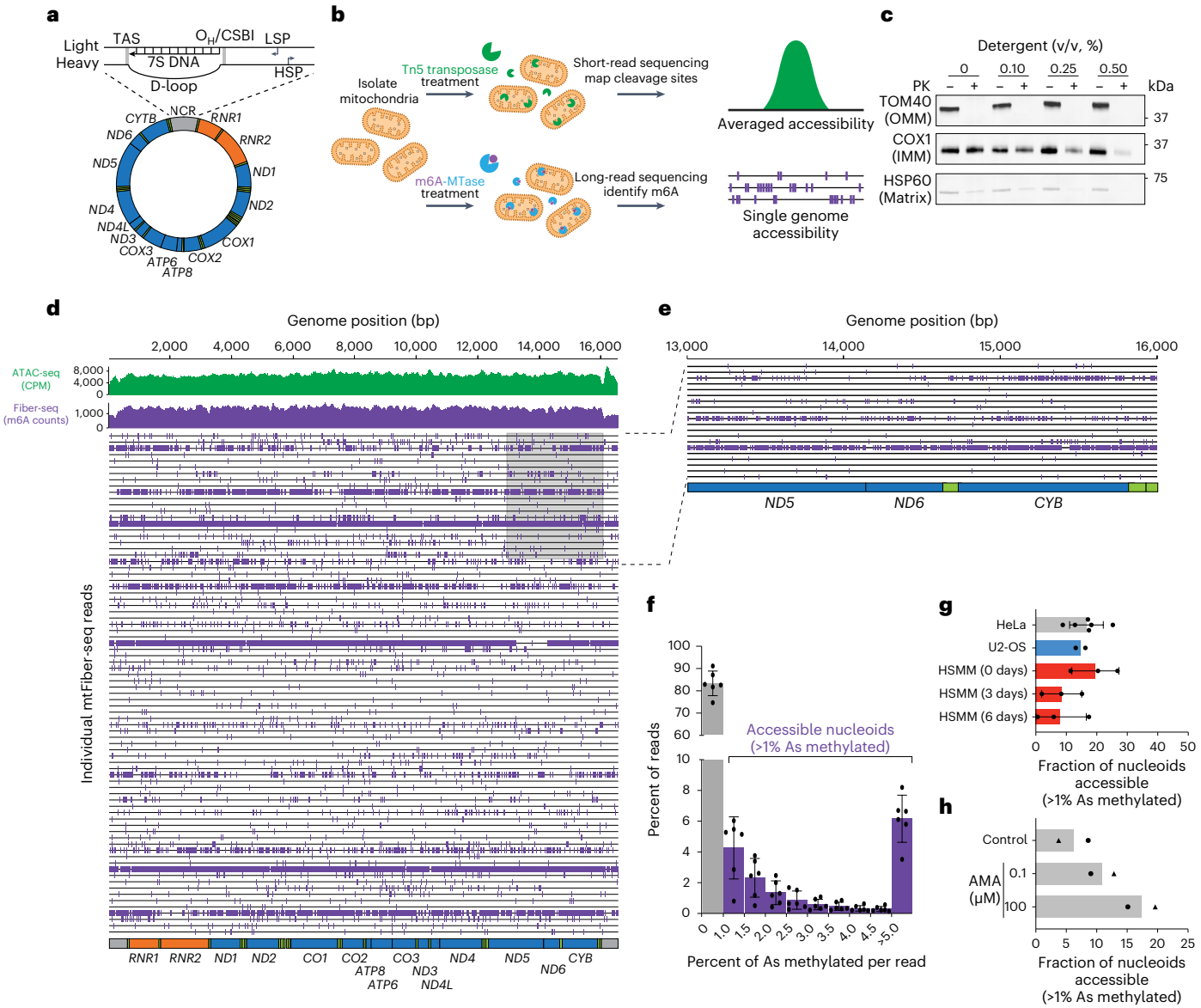


Fig. 1 | Most human mitochondrial nucleoids are inaccessible. a, Schematic of the human mitochondrial genome. Top, NCR of the mitochondrial genome, highlighting the termination associated sequence (TAS), D-loop and 7S DNA, the heavy strand O_H and CSBI, and the light and heavy strand promoters (LSP and HSP, respectively). Bottom, the 16.5-kb circular human mitochondrial genome encodes 13 polypeptides (blue), 22 tRNAs (green) and 2 rRNAs (orange). **b**, Schematic depicting the experimental design. Human cells were subjected to cellular fractionation to isolate mitochondria, which were then permeabilized and treated with Tn5 transposase for ATAC-seq or the Hia5 m6A-MTase for mtFiber-seq. **c**, Permeabilization of mitochondria with increasing concentrations of detergent was assessed by proteinase K (PK) digestion of the outer mitochondrial membrane (OMM) protein TOM40, the inner mitochondrial membrane (IMM) protein COX1 and the matrix-soluble protein HSP60, followed by western blot analysis. Representative blot of two independently repeated experiments. **d**, Mitochondrial genome comparing the signal between ATAC-seq, aggregated methylation from mtFiber-seq and 80 randomly sampled mtFiber-seq reads. Individual reads are indicated by horizontal black lines, and m6A-modified bases are marked by purple vertical dashes. **e**, Zoom-in of positions 13,000–16,000 in the mitochondrial genome showing 20 sampled reads. **f**, Barplot showing the percent of HeLa mtFiber-seq reads per bin based on the number of m6A modifications per read. Most reads (80%) have fewer than 1% adenine bases (As) methylated per molecule. Reads with greater than 5% As methylated were aggregated into a single bin; $n = 6$ biological replicates. **g**, Barplot showing the fraction of nucleoids with >1% As methylated in five cell types: HeLa, U2-OS, undifferentiated skeletal muscle myoblasts, and 3- and 6-day differentiated skeletal muscle myocytes; $n = 6$ biological replicates for HeLa, 2 biological replicates for U2-OS and 3 biological replicates for each timepoint across HSMM differentiation. **h**, Barplot showing the fraction of nucleoids with >1% As methylated in HeLa cells after treatment with antimycin A (AMA) or vehicle control for 24 h; $n = 2$ biological replicates. For all plots, error bars represent s.d. of the mean. CPM, counts per million.

their associated machinery. The mitochondrial transcription factor TFAM is the primary constituent of mitochondrial nucleoids[19–27]. High levels of TFAM compact DNA in vitro[19–21], yet the architecture formed by TFAM packaging and the extent to which other proteins contribute in cells remain unknown.

To address these questions, we measured DNA accessibility of individual nucleoids at near single-nucleotide and single-molecule resolution, generating full-length accessibility profiles of individual mtDNA molecules. The results showed that most nucleoids are in an inaccessible state, with the remainder exhibiting heterogeneity in

packaging. Thus, to first order, mtDNA accessibility is all-or-none, in contrast to nuclear genome organization, which is characterized by local differences in chromatin accessibility. We found that TFAM levels directly modulate the fraction of accessible nucleoids in cells. We identified footprints from transcription and replication factors across the accessible nucleoids, resolving their co-occupancy across single molecules. Comparing footprinting patterns between nucleoids reconstituted in vitro and nucleoids in cells demonstrated that TFAM alone explains most mtDNA packaging. These patterns were consistent with a nucleation-and-spreading model whereby TFAM binds first to high-affinity sites and subsequently binds cooperatively to coat and compact the genome.

Results

Mitochondrial Fiber-seq probes individual nucleoid accessibility

To investigate mtDNA accessibility patterns, we first subjected isolated mitochondria to assay for transposase-accessible chromatin (ATAC-seq) (Fig. 1b), which relies on the selective cleavage of accessible DNA by the hyperactive Tn5 transposase[28]. Similar to other short-read sequencing-based analyses of mtDNA accessibility[29–31], ATAC-seq revealed near-uniform accessibility across the mitochondrial genome (Fig. 1d), irrespective of the NCR and coding regions. Such homogeneity could arise from bulk averaging of accessibility across individual mitochondrial nucleoids. Since hundreds of copies of mtDNA are present in each cell, even single-cell based accessibility methods would result in bulk averaging, and the dynamic, interconnected mitochondrial network prevents a 'single-mitochondrion' approach. To overcome these barriers, we sought to leverage recent single-molecule chromatin fiber sequencing approaches[32–36]. Specifically, we adapted the Fiber-seq method[32] for mtDNA, because it was capable of evaluating the accessibility along both the heavy and light strands from the same 16.5 kb mtDNA molecule. Fiber-seq uses a nonspecific adenine methyltransferase (MTase) to mark accessible DNA, and application of Fiber-seq to the nuclear genome mirrors DNase I cleavage in bulk[32]. The use of an adenine MTase offers superior resolution of the mitochondrial genome relative to an approach based on GC dinucleotide methylation[37,38] due to the frequency of adenines and distances between them (Extended Data Fig. 1a,b). Fiber-seq 'stencils' the architecture of chromatin onto underlying DNA fragments via the selective modification of accessible A/T base pairs with N^6-methyl-deoxyadenosine (m6A). The m6A residues are then identified using highly accurate long-read sequencing of individual double-stranded 15–20 kb DNA fragments. We adapted this protocol to investigate the packaging of individual mitochondrial nucleoids (Fig. 1b). For mitochondrial Fiber-seq (mtFiber-seq), isolated mitochondria were permeabilized (Fig. 1c) and then treated with the m6A-MTase Hia5. After DNA isolation, mtDNA was linearized and subjected to PacBio HiFi single-molecule sequencing, enabling identification of modified adenines along each sequenced molecule of DNA.

Application of mtFiber-seq to human HeLa cells yielded reads spanning the full 16.5-kb mitochondrial genome (Fig. 1d,e and Extended Data Fig. 2a), with aggregated methylation profiles mirroring our observations by ATAC-seq (Fig. 1d). However, methylation patterns along individual reads revealed substantial intra- and inter-read heterogeneity in the accessibility of mitochondrial nucleoids that had previously been obscured by bulk-averaged accessibility measurements. In the vast majority of nucleoids (~80%), less than 1% of their total adenines were methylated, a methylation level approaching that of mtDNA untreated with MTase (Extended Data Fig. 1c), indicating that these nucleoids are compacted in a predominately inaccessible state (Fig. 1f). In the remaining population of nucleoids, the accessibility varied, with a maximum methylation level of ~80% of adenines.

To eliminate the possibility that the observed phenomenon is due to subsaturating reaction conditions, we varied the concentration of MTase and treatment times used in the mtFiber-seq reactions. These conditions did not substantially alter the proportion of nucleoids within the inaccessible state, demonstrating that this minimally methylated population is not due to insufficient enzyme or time (Extended Data Fig. 1d,e). To determine whether this phenomenon is unique to HeLa S3 cells, we performed mtFiber-seq in U2-OS cells and in human skeletal muscle myoblasts (HSMM) (Extended Data Fig. 2b,c) at several stages of differentiation—a process associated with both changes in mitochondrial biogenesis and proliferation[39]. In all of these cell types, the accessible population of nucleoids remained in the minority (Fig. 1g). To assess whether this proportion of nucleoids is static, we treated HeLa cells with antimycin A—a Complex III inhibitor that stimulates oxidative stress. After treatment for 24 h, we saw an increase in the accessible nucleoid population compared with the vehicle control (Fig. 1h). Together, these results reveal that mitochondrial nucleoids undergo heterogeneous compaction, with most nucleoids existing in a minimally accessible, compacted state, and that this population can shift in response to perturbed mitochondrial function.

To orthogonally validate this finding and resolve whether the variability in mtDNA accessibility is driven by inter- or intracellular differences in mtDNA compaction, we adapted and performed ATAC-see for mitochondria. ATAC-see uses Tn5 transposase to ligate fluorescent oligonucleotides into accessible DNA, which are visualized alongside mtDNA by fluorescence microscopy[40] (Extended Data Figs. 3a–c and 4a–d). In bulk, ATAC-see signal at mtDNA puncta was distributed broadly but skewed towards lower intensities, with most puncta exhibiting a low signal, mirroring the compacted state we observed with mtFiber-seq (Extended Data Fig. 3d). At the single-cell level, the ATAC-see and DNA signal distributions were similar to those in our population-level analysis, revealing that nucleoid accessibility is also highly variable across individual nucleoids within a given cell (Extended Data Fig. 3e). Together, these results demonstrate that most nucleoids are inaccessible and indicate that this heterogeneity is present in each cell.

Single-molecule protein occupancy of accessible nucleoids

We next sought to determine whether accessible nucleoids exhibited features of mitochondrial transcription and replication—processes that are regulated by multiple DNA-binding proteins. To identify protein occupancy along each nucleoid, we developed a hidden Markov model (HMM)-based footprinting approach that controls for local methylation propensities (Fig. 2a). Application of this HMM to accessible nucleoids from six HeLa datasets revealed diverse single-molecule patterns of protected footprints (Fig. 2b). For example, within the mt-tRNA Leu(UUR) gene, 50% of highly accessible nucleoids contained a ~30 bp footprint overlapping the known binding site of the transcription termination factor MTERF1 (ref. 41) (Fig. 2c), consistent with occupancy by a single MTERF1 molecule (Fig. 2d). These results demonstrate that accessible nucleoids are preferentially bound by MTERF1 and highlight the power of mtFiber-seq to capture several molecular features on the same DNA molecules.

Single-molecule architecture of the mitochondrial D-loop

Seventy percent of highly accessible nucleoids exhibited protection of the D-loop region (Fig. 2e), with single-molecule footprints localizing to two conserved sequence elements of the D-loop: the termination associated sequence (TAS) and the conserved sequence box I (CSBI). These single-molecule footprints mirrored chromatin immunoprecipitation followed by sequencing (ChIP–seq) profiles from the DNA polymerase Polγ and the replicative helicase TWINKLE[42] (Fig. 2e, top), indicating that the footprint patterns were formed, in part, by the residence of these factors at the TAS and CSBI. Notably, most reads with a TAS or CSBI footprint also contained an MTERF1 footprint, indicating that these factors were preferentially co-occupying the same nucleoids (Extended Data Fig. 5a).

https://doi.org/10.1038/s41594-024-01225-6

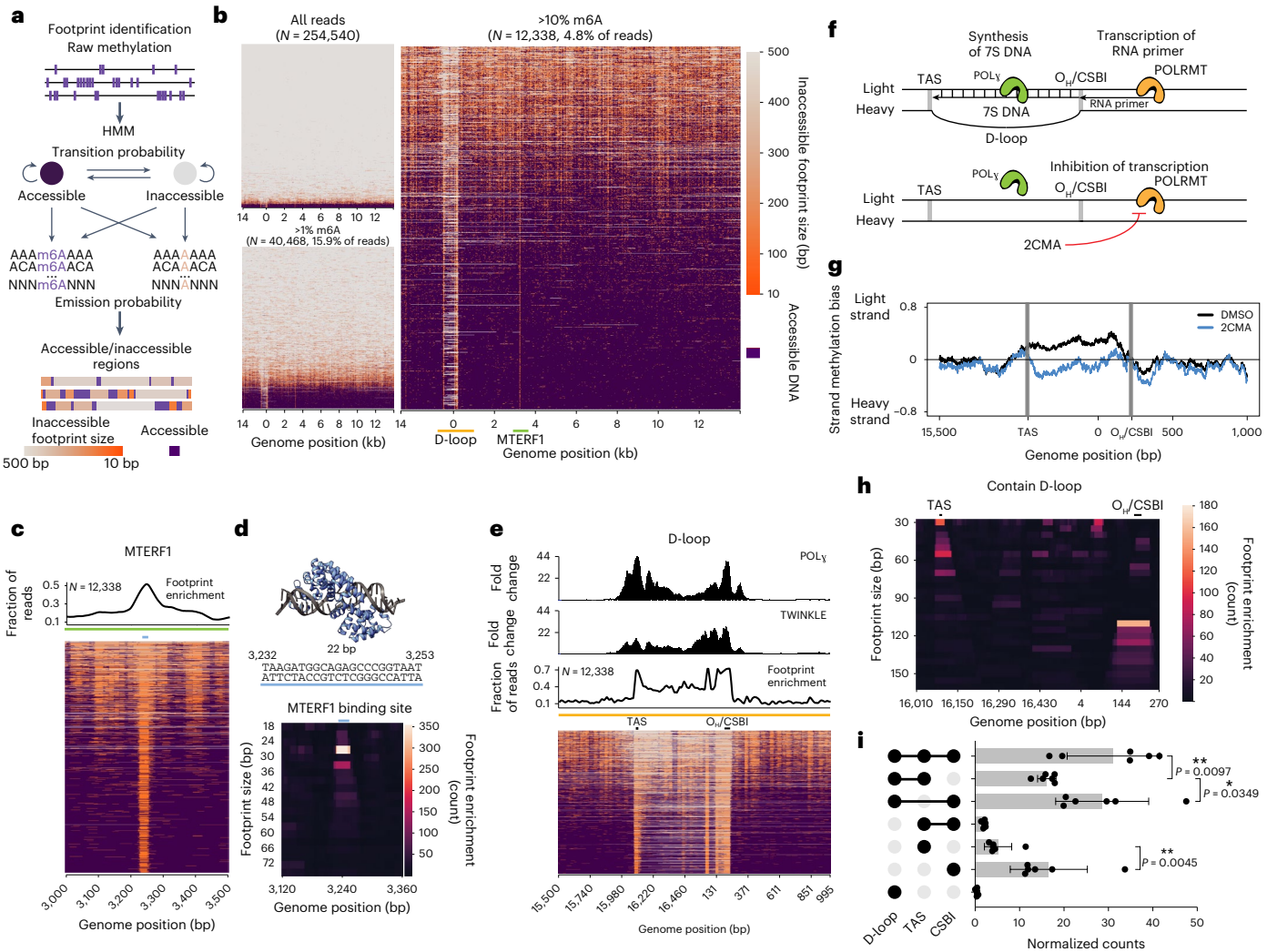


Fig. 2 | Accessibility patterns reveal mtDNA architecture. a, HMM used to identify accessible and inaccessible regions of the genome using probabilities of methylation in hexamer sequence contexts. **b**, Heatmaps showing the identified accessible and inaccessible regions of the genome from six biological replicates from HeLa cells. Each row represents an individual read. Accessible regions are colored purple, and protected regions are colored according to size. **c**, Top, metaplot of footprint enrichment showing the fraction of reads with >10% m6A. Bottom, zoom-in of the MTERF1 binding site for reads containing >10% m6A. The MTERF1 binding site is indicated by a blue line. **d**, Top, crystal structure of MTERF1 bound to its 22 bp binding site (PDB code 3MVA ref. 65). Bottom, heatmap of the footprint size enrichment at the MTERF1 site. The binding site is indicated by a blue line. **e**, Top, ChIP–seq tracks showing the fold change of signal over control for Polγ and TWINKLE[42]. Metaplot of the footprint enrichment showing the fraction of reads with >10% m6A. Bottom, zoom-in showing the region surrounding the D-loop. Heatmap shows reads containing >10% m6A. **f**, Cartoon depicting D-loop formation. POLRMT transcribes the RNA primers for POLγ to synthesize the 7S DNA, forming a D-loop. 2CMA inhibits POLRMT, resulting in no 7S DNA synthesis or D-loop formation. **g**, mtFiber-seq methylation strand bias at the NCR in HeLa cells. Bias is calculated as the number of methylations on the light and heavy strands, averaged over a 150-nt sliding window and normalized against the AT content of the region. **h**, Heatmap of the footprint size enrichment at the D-loop from HeLa cells in reads containing a D-loop. **i**, UpSet plot showing the frequency of nucleoids containing combinations of footprints at TAS and CSBI and containing a D-loop. Maximum footprint sizes of 60 bp and 140 bp were used for TAS and CSBI, respectively. Groups were compared by paired two-sided Student's *t*-test; not significant (NS), $P > 0.05$; *$P < 0.05$; **$P < 0.01$. $n = 6$ biological replicates. Error bars represent s.d. of the mean. The categories in the plot represent 2.10%, 2.31%, 1.89%, 1.01%, 2.81% and 2.67% of the total molecule population for replicates 1–6, respectively.

In some mtDNA molecules, the D-loop structure forms through 7S DNA hybridization to the light strand of the D-loop locus, displacing the heavy strand (Fig. 2f, top). The structure remains enigmatic as it has been challenging to isolate the D-loop containing nucleoid population that is typically in the vast minority. To quantify D-loop formation among individual nucleoids, we leveraged two important features unique to mtFiber-seq. First, the Hia5 MTase is >15-fold less active on single-stranded DNA than on double-stranded DNA (Extended Data Fig. 5b), so the methylation patterns on the light and heavy strands across the D-loop region are predicted to differ when the 7S DNA is present. Second, PacBio sequencing detects methylation on both strands, so strand-specific methylation within a region can be calculated by averaging the number of m6As identified on each strand, normalized against the local AT content. Using this approach, we found that, outside of the D-loop, both the light and heavy strands were methylated to a similar extent (Extended Data Fig. 5c,e). However, specifically between the TAS and CSBI of the D-loop region, we observed more methylation on the light strand (Fig. 2g and Extended Data Figs. 5d,e and 6a). If the methylation strand bias is due to the presence of a 7S DNA molecule forming a D-loop, then this strand bias should disappear when replication is inhibited (Fig. 2f, bottom). To test this, we treated cells with 2′-C-methyladenosine (2CMA)—a nucleoside analog that preferentially

inhibits mitochondrial transcription (Extended Data Fig. 5f) and thus blocks mtDNA replication by depleting the essential RNA primers[43–46]. Treatment of HeLa cells with 2CMA for 24 h led to a loss of strand bias within the D-loop region, suggesting that the bias was indeed caused by the D-loop structure (Fig. 2g and Extended Data Fig. 5c–e). In total, an average of 4% of nucleoids from HeLa cells contained a 7S DNA (Extended Data Fig. 5g), consistent with previously measured 7S DNA content in this cell type[47].

The single-molecule resolution of mtFiber-seq allows for the identification of regulatory features specifically associated with nucleoids containing a D-loop. Individual nucleoids with strand-specific methylation at the D-loop preferentially featured footprints that precisely overlapped the TAS and/or CSBI, indicating that mtDNA molecules with 7S DNA are probably actively engaged by replication machinery (Fig. 2h,i and Extended Data Fig. 6b,c). The pronounced footprint at the TAS is consistent with a paused or terminating Polγ, congruent with the proposed origin of the 7S DNA and frequently abortive nature of mtDNA replication[42,48]. Interestingly, we observed nucleoids lacking a D-loop that contained one, but not both, replication machinery-associated footprints. Separating these nucleoids into those protected at the TAS and CSBI revealed that most of these were protected only at CSBI, suggesting a poised, prereplication state (Fig. 2i). Together, these results demonstrate that nucleoids containing a 7S DNA are co-occupied with replication machinery at one or both ends of the D-loop. This suggests that D-loop-containing nucleoids are continuously cycling through replication initiation and premature termination, consistent with the high turnover of the 7S DNA and indicating that the D-loop structure is the natural byproduct of these cycles[6,7].

mtDNA accessibility decreases during myocyte differentiation

To understand how nucleoid accessibility changes in a biological context, we analyzed mtDNA packaging during HSMM differentiation, in which cells induce mitochondrial biogenesis (Extended Data Fig. 7a,b), exit the cell cycle and fuse together to form multinucleated myotubes[39,49]. We observed a decrease in the accessible nucleoid population during the first 3 days of differentiation that was maintained at the 6-day timepoint (Fig. 1g). Closer inspection of these nucleoids revealed changes to the D-loop region that indicate a reduction in mtDNA replication. We found a significant decrease in the methylation strand bias within the D-loop at 3 and 6 days, but not at other loci (Fig. 3a,b and Extended Data Fig. 7c), similar to what we observed with 2CMA treatment (Fig. 2g and Extended Data Fig. 5d,e), indicating lower levels of 7S DNA synthesis and mtDNA replication. We also subsampled the datasets to account for changes in the overall methylation distribution and classified each read as containing a D-loop or not. We found that the proportion of reads with D-loops decreased across differentiation (Fig. 3c). In sum, myoblast nucleoids decrease in accessibility and show signs of reduced replication as myoblasts withdraw from the cell cycle towards terminal differentiation, suggesting that a higher accessible nucleoid population supports the needs of active proliferation.

TFAM levels modulate mtDNA accessibility in cells

As TFAM is the primary constituent of mitochondrial nucleoids and compacts DNA in vitro[19–21], we hypothesized that this shift in the accessible population could be explained by an increase in the TFAM:mtDNA ratio during differentiation. Consistent with previous reports, we observed an increase in mtDNA copy number of ~1.3-fold by 6 days differentiation (Extended Data Fig. 7e). However, we saw an even larger increase in TFAM protein levels (Extended Data Fig. 7d,f), leading to a twofold increase in the TFAM:mtDNA ratio at day 3 and a 3.5-fold increase by day 6 (Fig. 3d). Together, these results support our hypothesis that the observed shift in the nucleoid population is due to a change in the TFAM:mtDNA ratio during myogenesis.

To determine whether an increase in TFAM levels alone is sufficient to shift nucleoid accessibility, we induced the expression of human influenza hemagglutinin (HA)-tagged TFAM that localized to mitochondria in HeLa S3 cells, yielding a 50% increase in TFAM levels (Fig. 3e,f and Extended Data Fig. 8a). We found that TFAM overexpression shifted the population towards more inaccessible nucleoids compared with dimethylsulfoxide (DMSO)-treated control cells (Fig. 3g and Extended Data Fig. 8b,c). In some systems, TFAM overexpression leads to increased mtDNA levels, but mtDNA levels were unchanged in our cells (Extended Data Fig. 8d), consistent with previous TFAM overexpression in HeLa cells[50]. Therefore, the increase in the TFAM:mtDNA ratio (Fig. 3h) explains the shift in the accessible population. Although the global fraction of accessible nucleoids decreased, the footprint occupancy patterns remained constant with higher levels of TFAM (Fig. 3i and Extended Data Fig. 8e–g). These findings indicate that TFAM levels predominantly impact mitochondrial nucleoid packaging by modulating the global fraction of accessible nucleoids without altering their architecture.

TFAM levels modulate mtDNA accessibility in vitro

To investigate the mechanism by which TFAM mediates mitochondrial nucleoid compaction, we performed in vitro MTase reactions on full-length mtDNA equilibrated with recombinant TFAM (Fig. 4a and Extended Data Fig. 9a,b). TFAM alone was sufficient to block the ability of Hia5 to methylate DNA in vitro in a concentration-dependent manner (Fig. 4b), with 30 μM TFAM virtually abolishing mtDNA methylation (Fig. 4b and Extended Data Fig. 9c). We then performed PacBio sequencing on these MTase-treated samples to measure single-molecule accessibility and TFAM occupancy patterns along each in vitro reconstituted nucleoid (Fig. 4c). Congruent with our bulk measurements (Fig. 4b), increasing TFAM concentrations was associated with decreasing levels of per-read methylation (Fig. 4c,d and Extended Data Fig. 9d).

The high resolution of mtFiber-seq permits a deeper analysis of mtDNA packaging over bulk measures of compaction. We extracted the footprints across the reconstituted nucleoids using our HMM and compared the in vitro architecture with our measurements in cells (Extended Data Fig. 9e). The in vitro reconstituted nucleoids lacked localized footprint densities at the D-loop and MTERF1 binding site, consistent with non-TFAM factors driving the formation of these features in cells (Fig. 4e). However, across the rest of the mitochondrial genome, the footprint patterns along in vitro reconstituted nucleoids mirrored those derived from HeLa cells (Fig. 4e). These results demonstrate that the protection patterns observed in cells are dominated by TFAM binding and indicate that the population of accessible nucleoids is driven by TFAM protein levels in vivo.

TFAM packages nucleoids via 'nucleation-and-spreading'

We next used the single-molecule footprint patterns on in vitro reconstituted nucleoids to elucidate the mechanism by which TFAM packages DNA. First, we calculated the relative binding affinity ($K_{1/2}$ constants) of TFAM for each base in the mitochondrial genome to ask whether certain sites are preferentially bound by TFAM. To compute these per-base $K_{1/2}$ constants, we calculated the fraction of reads with a footprint at each genomic position with each TFAM concentration and used a four-parameter logistics regression to determine the concentration of TFAM at which 50% of mtDNA molecules are protected at that position (Fig. 4f,g). While binding assays with short fragments of DNA show similar equilibrium dissociation constant (K_D) values irrespective of sequence[19], here we see a range of binding affinities, revealing preferential binding sites situated throughout the genome. This approach revealed 31 unique high-affinity TFAM binding sites along the mitochondrial genome that were preferentially bound even at low concentrations of TFAM. We were unable to identify strong motifs or common features that defined these sites. We next calculated the Hill coefficients of TFAM binding across the genome to determine the degree of cooperative TFAM binding. We found Hill coefficients exceeded 2 at the large majority of mtDNA loci (Extended Data

https://doi.org/10.1038/s41594-024-01225-6

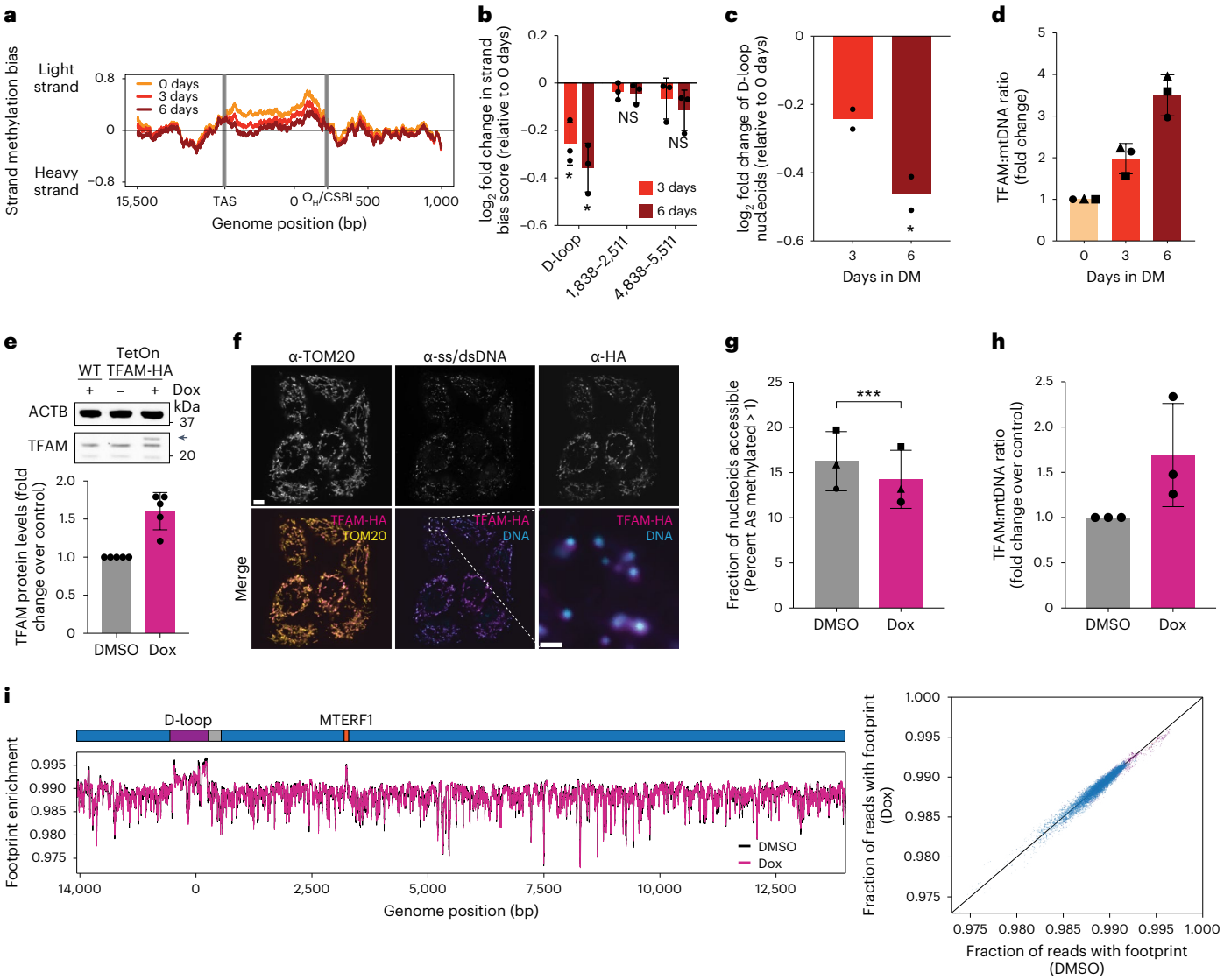


Fig. 3 | Altered TFAM levels shift the population of accessible nucleoids. **a**, Methylation strand bias at the NCR in HSMM throughout differentiation. **b**, The $\log_2$ fold change in the strand bias score after 3 and 6 days in differentiation media (DM) across three regions. Samples were compared with two-sided Student's *t*-tests: $^*P < 0.05$; NS, $P > 0.05$. Exact *P* values: 0.034 (3 days versus 0 days) and 0.022 (6 days versus 0 days); $n = 3$ biological replicates. **c**, The $\log_2$ fold change of D-loop containing nucleoids from HSMMs after 3 and 6 days in DM relative to undifferentiated myoblasts. Datasets were subsampled to match methylation distributions. Samples were compared by Fisher's exact tests: $^*P < 0.05$. Exact *P* values: 0.066 and 0.244 (3-day versus 0-day replicates) and 0.0008 and 0.027 (6-day versus 0-day replicates); $n = 2$ biological replicates. **d**, Barplot showing the TFAM:mtDNA ratio from HSMMs after 0, 3 and 6 days in DM; $n = 3$ biological replicates. **e**, Overexpression of TFAM. Wild type (WT) HeLa cells and TetOn-TFAM-HA HeLa cells were treated with DMSO or Dox for 48 h. Top, TFAM levels were assessed by western blot, with ACTB as a control. Treatment of TetOn-TFAM-HA cells with Dox results in the appearance of a band corresponding to the HA-tagged TFAM. Bottom, TFAM bands were quantified and normalized against ACTB; $n = 5$ biological replicates. Error bars represent s.d. of the mean. **f**, Confocal microscopy showing the TFAM-HA construct localized to mitochondria and nucleoids. Cells were treated with Dox for 48 h and labeled for TOM20, ss/dsDNA and HA. Scale bars, 5 µm and 1 µm (zoom-in). **g**, Barplot showing the fraction of accessible nucleoids (>1% m6A). Individual replicates were compared with two-sided chi-squared tests: $^{***}P < 0.001$. Exact *P* values: 2.265×10^{-9}, $<2.2 \times 10^{-16}$ and 8.022×10^{-13} for replicates 1–3; $n = 3$ biological replicates. **h**, Barplot showing the TFAM:mtDNA ratio after treatment with DMSO or Dox for 48 h; $n = 3$ biological replicates. **i**, Lineplot (left) and scatterplot (right) showing the footprint enrichment with TFAM overexpression relative to the control (Pearson's $r = 0.97$). Datasets were subsampled to match methylation distributions. Error bars represent s.d. of the mean for all plots.

Fig. 9f), consistent with previous reports of cooperative TFAM binding to nonspecific DNA in vitro[24,25]. Our results indicate that TFAM preferentially binds certain DNA sequences and has a strong propensity for cooperative binding genome-wide.

We next hypothesized that these properties of TFAM binding would lead to nucleoid compaction via a cooperative occupancy of TFAM adjacent to high-affinity TFAM-bound nucleation sites. Indeed, at lower concentrations of TFAM, high-affinity sites had footprint sizes that mirrored those expected from the occupancy of a single TFAM molecule (~30 bp) (Fig. 4h–j and Extended Data Figs. 9g,h and 10a–d). At increased TFAM concentrations, these footprints increased in size while remaining centered at the high-affinity site, consistent with cooperative binding of additional TFAM proteins rather than the appearance of additional single TFAM binding events. Low-affinity sites were substantially more accessible at lower TFAM concentrations and became protected by large footprints at higher concentrations,

https://doi.org/10.1038/s41594-024-01225-6

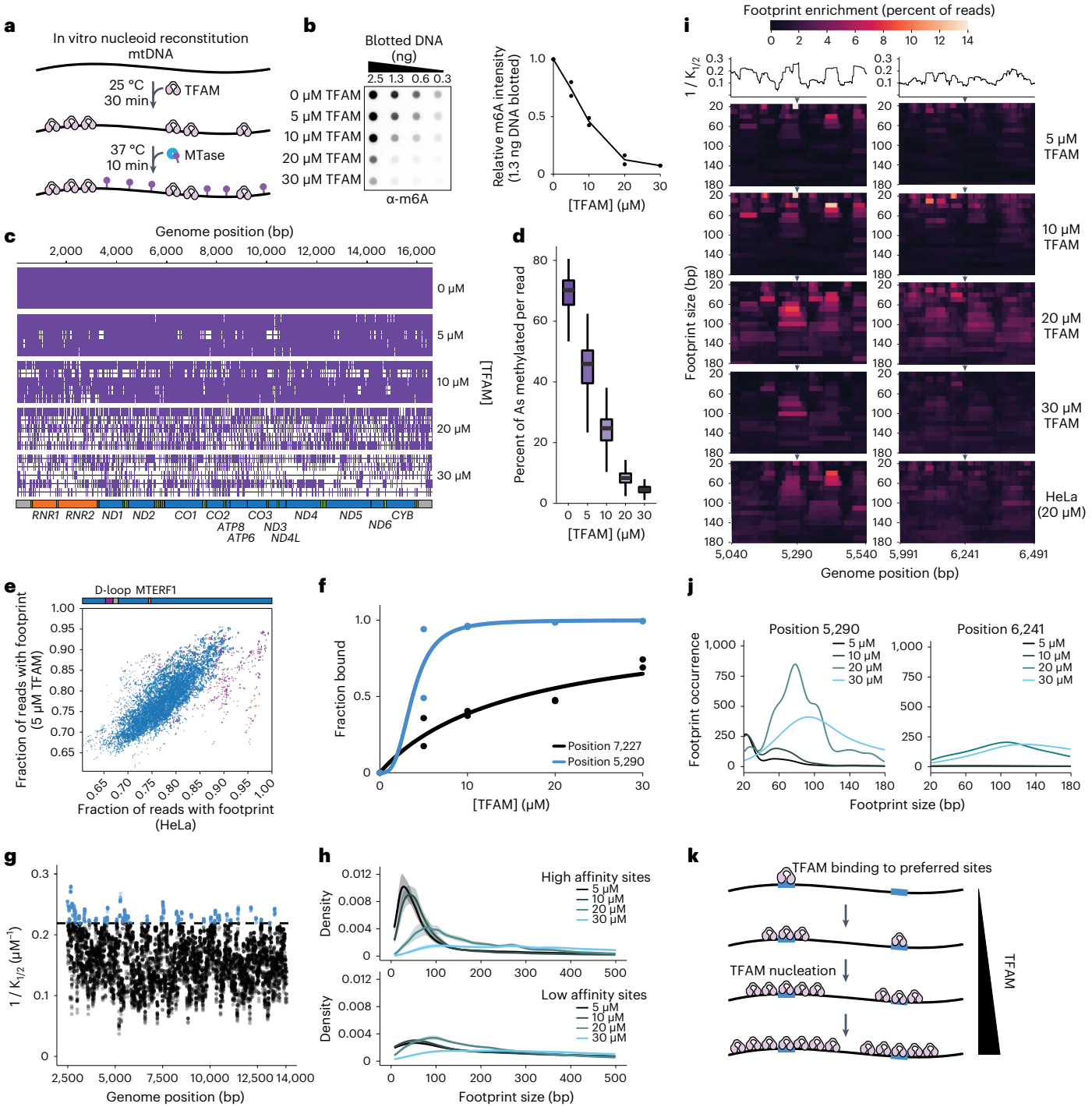


Fig. 4 | In vitro reconstituted nucleoids reveal preferential TFAM binding and nucleation sites throughout the genome. a, Nucleoids were reconstituted with PCR-amplified mtDNA and recombinant TFAM and methylated with Hia5. **b**, Left, dot blot assessing methylation with increasing TFAM. DNA was adsorbed onto a nitrocellulose membrane and detected with an anti-m6A antibody. Right, dot blot quantification: intensities were quantified and normalized against the 0 μM sample; $n = 2$ replicates. **c**, Five randomly sampled mtFiber-seq reads for each TFAM concentration. Reads are indicated by black lines and m6A bases are marked by purple dashes. **d**, Boxplot showing the percent of As methylated per read for each TFAM concentration. Boxplot elements are: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. $n = 29{,}799$, 47,300, 68,377, 80,562 and 82,809 reads for 0, 5, 10, 20 and 30 μM TFAM, respectively. **e**, Scatterplot showing the footprint enrichment across the genome in HeLa cells and in vitro with 5 μM TFAM for reads subsampled to match methylation distributions (Pearson's $r = 0.665$). **f**, Example binding curves from in vitro mtFiber-seq data. Reads were classified as bound at a site if a footprint of ≥20 bp overlapped the position. Data were fit to a four-parameter logistic regression; $n = 2$ replicates. **g**, Affinities were determined from position 2,500 to 14,000; $1/K_{1/2}$ constants are shown. Results are from the average of two replicates. **h**, Meta density plots of footprint sizes from (top) 26 high-affinity sites and (bottom) 64 low-affinity sites at each TFAM concentration. The center line represents the kernel density estimate with the 95% confidence intervals shaded. Distributions between high and low-affinity sites are significantly different at 5 μM TFAM (two-sample Kolmogorov–Smirnov (KS) test, $P < 2.2 \times 10^{-16}$, D = 0.32) and 10 μM TFAM (two-sample KS test, $P < 2.2 \times 10^{-16}$, D = 0.29). **i**, Heatmaps of footprint size enrichment at two loci from in vitro reconstituted nucleoids (top) and for HeLa nucleoids, subsampled to match the methylation distribution of the 20 μM TFAM dataset. Line plots indicating the $1/K_{1/2}$ are shown above each heatmap. **j**, Density plots showing the footprint size distribution at positions 5,290 and 6,241 as a function of TFAM concentration. **k**, TFAM binds and nucleates from preferred sites (blue) located throughout the genome.

indicative of TFAM association aided by the cooperative extension from a proximal high-affinity site (Fig. 4h–j and Extended Data Fig. 10e–h). Moreover, the in vivo data mirrored the in vitro results; we observed nearly identical footprint patterns for our in vivo HeLa S3 data when we subsampled the reads to match the methylation densities of the in vitro reads (Fig. 4i and Extended Data Fig. 10a,b,e,f). These TFAM association patterns are consistent with computational modeling of proteins binding cooperatively to DNA[51]. Taken together, these results imply a model for mtDNA compaction in cells in which TFAM binds higher-affinity sites throughout the genome that nucleate the compaction of mtDNA through TFAM spreading out from these sites (Fig. 4k).

Discussion

Our single-molecule analysis of the mitochondrial genome revealed substantial heterogeneity in the packaging of individual mitochondrial nucleoids—a feature previously obscured by the bulk averaging inherent to cleavage-based accessibility methods. These observations demonstrated that most mtDNA molecules are inaccessible, uncovering an unappreciated layer of mtDNA regulation. The remaining population shows heterogeneity in the extent and pattern of accessibility, which could reflect several genomic processes including progression of the RNA and DNA polymerases and stochastic dissociation/reassociation of TFAM. We found that TFAM levels correlate strongly with methylation levels detected by mtFiber-seq. Our data are consistent with a model whereby TFAM nucleates from preferred binding sites throughout the genome and then spreads to compact mtDNA through cooperative binding. Thus, TFAM levels directly control DNA accessibility in cells, which would impact the ability of the transcription and replication machinery to bind and initiate their respective reactions. Consistent with this, high TFAM levels inhibit transcription and replication in vitro and in vivo[52–55], inactive nucleoids show higher TFAM signal by STED microscopy[18], and only a fraction of nucleoids undergo transcription or replication[18,56]. Moreover, we found that, when mtDNA accessibility decreased, fewer hallmarks of active replication were observed within the reads (Fig. 3a–c and Extended Data Fig. 8g). Although accessibility and activity may not be coupled directly, we propose that the two are correlated, with inaccessible nucleoids having a higher likelihood of being inactive, and, conversely, accessible nucleoids having a higher potential to be active, as is indicated by the co-occupancy of these accessible nucleoids by transcription and replication components that are defining features of active mitochondrial nucleoids.

Why does the cell maintain a pool of inaccessible nucleoids? Two hypotheses could explain the phenomenon. First, inactive molecules could serve as a genetic reservoir. The maintenance of a healthy pool of mtDNA is critical, especially in long-lived cells such as neurons. mtDNA repair is inefficient, and mtDNA replication introduces mutations and deletions due to the intrinsic error rate of Polγ and from replication-fork stalling[57–61]. Thus, damaged molecules could be selectively degraded and replaced by undamaged copies that were inactive in the reservoir. Second, an inactive population could serve as excess capacity, enabling transcription and replication throughout the mitochondrial network precisely where they are needed. As levels of nuclear-encoded TFAM vary across tissues and disease states[62,63], the fraction of accessible and active nucleoids is likely to be dynamic as well, as we observed during differentiation (Fig. 1g), and may represent a crucial nuclear-controlled parameter in the balancing of oxidative phosphorylation production[64]. Our current data lack time resolution, so we are unable to comment on the dynamics and transitions between inaccessible and accessible states. Thus, addressing these hypotheses will require a time-resolved mtFiber-seq strategy, representing a priority for future studies.

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/s41594-024-01225-6.

References

1. Farge, G. & Falkenberg, M. Organization of DNA in mammalian mitochondria. *Int. J. Mol. Sci.* **20**, 2770 (2019).
2. Ojala, D., Crews, S., Montoya, J., Gelfand, R. & Attardi, G. A small polyadenylated RNA (7S RNA), containing a putative ribosome attachment site, maps near the origin of human mitochondrial DNA replication. *J. Mol. Biol.* **150**, 303–314 (1981).
3. Xu, B. & Clayton, D. A. RNA–DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA–DNA hybrids serving as primers. *EMBO J.* **15**, 3135–3143 (1996).
4. Di Re, M. et al. The accessory subunit of mitochondrial DNA polymerase gamma determines the DNA content of mitochondrial nucleoids in human cultured cells. *Nucleic Acids Res.* **37**, 5701–5713 (2009).
5. Clayton, D. A. Replication of animal mitochondrial DNA. *Cell* **28**, 693–705 (1982).
6. Bogenhagen, D. & Clayton, D. A. Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence. *J. Mol. Biol.* **119**, 49–68 (1978).
7. Robberson, D. L. & Clayton, D. A. Pulse-labeled components in the replication of mitochondrial deoxyribonucleic acid. *J. Biol. Chem.* **248**, 4512–4514 (1973).
8. Milenkovic, D. et al. TWINKLE is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication. *Hum. Mol. Genet.* **22**, 1983–1993 (2013).
9. Nicholls, T. J. & Minczuk, M. In D-loop: 40 years of mitochondrial 7S DNA. *Exp. Gerontol.* **56**, 175–181 (2014).
10. Suomalainen, A. & Battersby, B. J. Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat. Rev. Mol. Cell Biol.* **19**, 77–92 (2018).
11. Schon, E. A., DiMauro, S. & Hirano, M. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat. Rev. Genet.* **13**, 878–890 (2012).
12. Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).
13. Kim, M., Mahmood, M., Reznik, E. & Gammage, P. A. Mitochondrial DNA is a major source of driver mutations in cancer. *Trends Cancer Res.* **8**, P1046–P1059 (2022).
14. Gorman, G. S. et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann. Neurol.* **77**, 753–759 (2015).
15. Keogh, M. J. & Chinnery, P. F. Mitochondrial DNA mutations in neurodegeneration. *Biochim. Biophys. Acta* **1847**, 1401–1411 (2015).
16. Sanchez-Contreras, M. & Kennedy, S. R. The complicated nature of somatic mtDNA mutations in aging. *Front Aging* **2**, 805126 (2022).
17. Kukat, C. et al. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc. Natl Acad. Sci. USA* **108**, 13534–13539 (2011).
18. Brüser, C., Keller-Findeisen, J. & Jakobs, S. The TFAM-to-mtDNA ratio defines inner-cellular nucleoid populations with distinct activity levels. *Cell Rep.* **37**, 110000 (2021).
19. Ngo, H. B., Lovely, G. A., Phillips, R. & Chan, D. C. Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat. Commun.* **5**, 3077 (2014).
20. 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).

21. 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).
22. Lu, B. et al. Phosphorylation of human TFAM in mitochondria impairs DNA binding and promotes degradation by the AAA+ Lon protease. *Mol. Cell* **49**, 121–132 (2013).
23. Cuppari, A. et al. DNA specificities modulate the binding of human transcription factor A to mitochondrial DNA control region. *Nucleic Acids Res.* **47**, 6519–6537 (2019).
24. Farge, G. et al. Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat. Commun.* **3**, 1013 (2012).
25. King, G. A. et al. Acetylation and phosphorylation of human TFAM regulate TFAM-DNA interactions via contrasting mechanisms. *Nucleic Acids Res.* **46**, 3633–3642 (2018).
26. Ngo, H. B., Kaiser, J. T. & Chan, D. C. The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.* **18**, 1290–1296 (2011).
27. Rubio-Cosials, A. et al. Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.* **18**, 1281–1289 (2011).
28. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
29. Blumberg, A., Danko, C.G., Kundaje, A. & Mishmar, D. A common pattern of DNase-I footprinting throughout the human mtDNA unveils clues for a chromatin-like organization. *Genome Res.* **28**, 1158–1168 (2018).
30. Mercer, T. R. et al. The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
31. 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).
32. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
33. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).
34. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
35. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
36. Abdulhay, N. J. et al. Nucleosome density shapes kilobase-scale regulation by a mammalian chromatin remodeler. *Nat. Struct. Mol. Bio.* **30**, 1571–1581 (2023).
37. Sakamoto, Y. et al. Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing. *Nucleic Acids Res.* **49**, e81 (2021).
38. Chen, J. et al. Whole-genome long-read TAPS deciphers DNA methylation patterns at base resolution using PacBio SMRT sequencing technology. *Nucleic Acids Res.* **50**, e104 (2022).
39. Frangini, M. et al. Synthesis of mitochondrial DNA precursors during myogenesis, an analysis in purified C2C12 myotubes. *J. Biol. Chem.* **288**, 5624–5635 (2013).
40. Chen, X. et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
41. Terzioglu, M. et al. MTERF1 binds mtDNA to prevent transcriptional interference at the light-strand promoter but is dispensable for rRNA gene transcription regulation. *Cell Metab.* **17**, 618–626 (2013).
42. Jemt, E. et al. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res.* **43**, 9262–9275 (2015).
43. Arnold, J. J. et al. Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. *PLoS Pathog.* **8**, e1003030 (2012).
44. Pham, X. H. et al. Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J. Biol. Chem.* **281**, 24647–24652 (2006).
45. Chang, D. D. & Clayton, D. A. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc. Natl Acad. Sci. USA* **82**, 351–355 (1985).
46. Gillum, A. M. & Clayton, D. A. Mechanism of mitochondrial DNA replication in mouse L-cells: RNA priming during the initiation of heavy-strand synthesis. *J. Mol. Biol.* **135**, 353–368 (1979).
47. Brown, W. M., Shine, J. & Goodman, H. M. Human mitochondrial DNA: analysis of 7S DNA from the origin of replication. *Proc. Natl Acad. Sci. USA* **75**, 735–739 (1978).
48. Brown, T. A. & Clayton, D. A. Release of replication termination controls mitochondrial DNA copy number after depletion with 2′,3′-dideoxycytidine. *Nucleic Acids Res.* **30**, 2004–2010 (2002).
49. Shen, X. et al. Genome-wide examination of myoblast cell cycle withdrawal during differentiation. *Dev. Dyn.* **226**, 128–138 (2003).
50. Kozhukhar, N. & Alexeyev, M. F. Limited predictive value of TFAM in mitochondrial biogenesis. *Mitochondrion* **49**, 156–165 (2019).
51. van der Heijden, T. & Dekker, C. Monte Carlo simulations of protein assembly, disassembly, and linear motion on DNA. *Biophys. J.* **95**, 4560–4569 (2008).
52. Uchida, A. et al. Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter. *eLife* **6**, e27283 (2017).
53. Farge, G. et al. In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription. *Cell Rep.* **8**, 66–74 (2014).
54. Bonekamp, N. A. et al. High levels of TFAM repress mammalian mitochondrial DNA transcription in vivo. *Life Sci. Alliance* **4**, e202101034 (2021).
55. Shutt, T. E., Lodeiro, M. F., Cotney, J., Cameron, C. E. & Shadel, G. S. Core human mitochondrial transcription apparatus is a regulated two-component system in vitro. *Proc. Natl Acad. Sci. USA* **107**, 12133–12138 (2010).
56. Lewis, S. C., Uchiyama, L. F. & Nunnari, J. ER-mitochondria contacts couple mtDNA synthesis with mitochondrial division in human cells. *Science* **353**, aaf5549 (2016).
57. Phillips, A. F. et al. Single-molecule analysis of mtDNA replication uncovers the basis of the common deletion. *Mol. Cell* **65**, 527–538.e6 (2017).
58. Larsson, N.-G. Somatic mitochondrial DNA mutations in mammalian aging. *Annu. Rev. Biochem.* **79**, 683–706 (2010).
59. Falkenberg, M. & Gustafsson, C. M. Mammalian mitochondrial DNA replication and mechanisms of deletion formation. *Crit. Rev. Biochem. Mol. Biol.* **55**, 509–524 (2020).
60. Fontana, G. A. & Gahlon, H. L. Mechanisms of replication and repair in mitochondrial DNA deletion formation. *Nucleic Acids Res.* **48**, 11244–11258 (2020).
61. Lujan, S. A. et al. Ultrasensitive deletion detection links mitochondrial DNA replication, disease, and aging. *Genome Biol.* **21**, 248 (2020).
62. Kang, I., Chu, C. T. & Kaufman, B. A. The mitochondrial transcription factor TFAM in neurodegeneration: emerging evidence and mechanisms. *FEBS Lett.* **592**, 793–811 (2018).
63. Hsieh, Y.-T. et al. Mitochondrial genome and its regulator TFAM modulates head and neck tumourigenesis through intracellular metabolic reprogramming and activation of oncogenic effectors. *Cell Death Dis.* **12**, 961 (2021).

64. Soto, I. et al. Balanced mitochondrial and cytosolic translatomes underlie the biogenesis of human respiratory complexes. *Genome Biol.* **23**, 170 (2022).
65. Yakubovskaya, E., Mejia, E., Byrnes, J., Hambardjieva, E. & Garcia-Diaz, M. Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. *Cell* **141**, 982–993 (2010).

Methods

Cell lines and culture

HeLa S3 cells (ATCC CCL-2.2) were grown in DMEM containing glucose and pyruvate (Thermo Fisher Scientific, cat. no. 11995073) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, cat. no. 10437028). U2-OS cells (ATCC, cat. no. HTB-96) were grown in McCoy's 5A medium (ATCC, cat. no. 30-2007) supplemented with 10% fetal bovine serum. HSMM were from anonymous healthy control samples kindly provided by B. Battersby (Institute of Biotechnology, University of Helsinki). Myoblasts were grown in human skeletal muscle cell basal medium containing growth supplement (Cell Applications, cat. no. 151K-500). To differentiate into myocytes, myoblasts at 80% confluency were switched to DMEM containing glucose and pyruvate (Thermo Fisher Scientific, cat. no. 11995073) supplemented with 2% heat-inactivated horse serum (Thermo Fisher Scientific, cat. no. 26050-070) and 0.4 μg ml^{-1} dexamethasone. Medium was then replaced every 24 h, and cells were harvested at either 3 days or 6 days postdifferentiation. Dermal human fibroblasts were grown in DMEM containing glucose and pyruvate supplemented with 10% fetal bovine serum. Cell lines were tested for mycoplasma contamination and confirmed negative by PCR using a Universal Mycoplasma Detection Kit (ATCC, cat. no. 30-1012K). For 2CMA treatment, HeLa S3 cells at roughly 60% confluency were treated with either 100 μM 2CMA in DMSO or DMSO alone and harvested after 2, 4, 6 or 24 h. For AMA treatment, HeLa S3 cells at roughly 60% confluency were treated with either AMA (0.1 or 100 μM) in ethanol or ethanol alone and harvested after 24 h.

Constructs and cloning

For recombinant TFAM, cDNA corresponding to TFAM lacking the mitochondrial targeting sequence (amino acids 50–246) was cloned into pET30a using the *Bam*HI and *Not*I restriction sites. The construct contains an N-terminal 6×His tag and a TEV cleavage site just upstream of the TFAM coding region to allow for the generation of untagged protein. For the TFAM-HA overexpression HeLa cell line, full-length TFAM with a C-terminal HA-tag connected by an SGGS linker was cloned into pCW57.1 using *Nhe*I and *Bam*HI sites. Lentivirus was generated using HEK293T cells (ATCCTEV, cat. no. CRL-3216) using pCW57.1_TFAMHA, pRSV-REV, pMDLg and pMD2.G with Lipofectamine 3000 (Thermo Fisher Scientific, cat. no. L3000008) according to the manufacturer's instructions. Collected virus was added to HeLa S3 cells grown to 50% confluency in six-well tissue culture dishes. Then 1 ml virus was added along with 1 ml DMEM + 10% FBS and polybrene to a final concentration of 8 μg ml^{-1}. After 24 h, cells were selected with 2 μg ml^{-1} puromycin until all negative control cells had died.

Immunoblotting and antibodies

Lysates used for immunoblotting were prepared in RIPA buffer in the presence of protease inhibitors. Lysates were normalized using Qubit Protein Assay (Invitrogen, cat. no. Q33211), and standard SDS–PAGE and blotting protocols were used. The following antibodies were used in this study: N6-methyladenosine (Active Motif, cat. no. 61995), TFAM (Santa Cruz, cat. no. sc-376672), TFAM (Proteintech, cat. no. 22586-1-AP), TOMM40 (Proteintech, cat. no. 18409-1-AP), TOM20 (Santa Cruz, cat. no. sc-17764), HSP60 (Cell Signaling, cat. no. D307), MT-CO1/COX1 (Abcam, cat. no. ab14705), ACTB (Cell Signaling, cat. no. 3700), ds/ssDNA (PROGEN, cat. no. 61014), sodium potassium ATPase (Abcam, cat. no. ab76020), NDUFS1 (Abcam, cat. no. ab169540), ATP6 (Proteintech, cat. no. 55313-1-AP), GAPDH (Invitrogen, cat. no. AM4300), HA (Cell Signaling, cat. no. 3724), horseradish peroxidase (HRP)-conjugated anti-rabbit IgG (Cell Signaling, cat. no. 7074S), HRP-conjugated anti-mouse IgG (Cell Signaling, cat. no. 70765), goat anti-rabbit IgG Alexa Fluor 647 (Thermo Fisher, cat. no. A-21245), goat anti-mouse IgG2a Alexa Fluor 488 (Thermo Fisher, cat. no. A-21131), goat anti-mouse IgG2a Alexa Fluor 555 (Thermo Fisher, cat. no. A-21137), goat anti-mouse IgM Alexa Fluor 647 (Abcam, cat. no. ab150123), goat anti-mouse IgM Alexa Fluor 555 (Thermo Fisher, cat. no. A-21426) and goat anti-rabbit IgG Alexa Fluor 546 (Invitrogen, cat. no. A-11010).

Hia5 expression and purification

pHia5ET was expressed and purified as previously described[32]. pHia5ET was transformed into T7 Express lysY/Iq *Escherichia coli* cells (NEB, cat. no. C3013I). Overnight cultures were added to two 1 l cultures of Luria-Bertani (LB) medium supplemented with 50 μg ml^{-1} kanamycin and grown with shaking at 37 °C to an optical density measured at a wavelength of 600 nm (OD_{600}) of 0.8–1.0. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and protein was expressed for 4 h at 20 °C with shaking. Cells were pelleted at 4,220*g* for 10 min at 4 °C. The pellet was resuspended in 35 ml lysis buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 10% glycerol; 0.5% Triton X-100; 10 mM β-mercaptoethanol) supplemented with 2× Complete, EDTA-free protease inhibitor cocktail (Millipore Sigma, cat. no. 11873580001). Cells were lysed by probe sonication (Qsonica, cat. no. Q125) for 10 min on ice at 50% amplitude, 30 s on/off. Lysate was clarified by centrifuging for 1 h at 40,000*g*. Ni-NTA agarose (Qiagen, cat. no. 30210) was prepared by washing 5 ml slurry with 30 ml equilibration buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 20 mM imidazole) and centrifuged at 500*g* for 3 min, repeating once. The clarified lysate and Ni-NTA agarose were combined and rotated at 4 °C for 1 h. The lysate mixture was poured over a disposable gravity flow column (Bio-Rad, cat. no. 7321010) and washed with 20 ml Buffer 1 (50 mM HEPES, pH 7.5; 300 mM NaCl; 50 mM imidazole) and 15 ml Buffer 2 (50 mM HEPES, pH 7.5; 300 mM NaCl; 70 mM imidazole). Protein was eluted with 15 ml elution buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 250 mM imidazole). SnakeSkin dialysis tubing (6–8K MWCO; Spectrum Laboratories 132650) was prewet in dialysis buffer (50 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM dithiothreitol (DTT)) and the eluate was added to the tubing and dialyzed against 2 l dialysis buffer overnight at 4 °C. Dialyzed sample was concentrated using a 10K Amicon Ultra-15 spin concentrator (Millipore Sigma, cat. no. UFC901008) and centrifuged at 3,220*g* to concentrate to a volume of less than 1 ml. The concentrated sample was injected onto Tandem HiTrap Q HP (Cytiva, cat. no. 17115301) and HiTrap SP HP (Cytiva, cat. no. 17115101) columns equilibrated with FPLC Buffer A (50 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM DTT). Columns were washed with five column volumes of FPLC Buffer A. The Q column was removed and the sample eluted from the SP column using a linear gradient over 20 column volumes of 0 to 100% FPLC Buffer B (50 mM Tris-HCl, pH 8; 1 M NaCl; 1 mM DTT). Peak fractions were collected and concentrated to 250 μl using a 10K Amicon Ultra-15 spin concentrator. Protein was supplemented with glycerol to a final concentration of 10%, frozen and stored at −80 °C. Protein purity was assessed by SDS–PAGE, and its activity measured by an in vitro methylation assay.

In vitro MTase activity assay

Hia5 activity was quantified as previously described[32] with minor modifications. Substrate DNA was prepared by PCR of the pHia5ET plasmid using T7 forward and reverse primers. The PCR product was purified using a Monarch PCR & Cleanup Kit (NEB, cat. no. T1030S) according to manufacturer's instructions. A series of 11 60 μl MTase reactions were prepared with 1 μg substrate DNA with alternating twofold and fivefold enzyme dilutions (10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005 and 0.0001 μl MTase) in MTase buffer (15 mM Tris, pH 8.0; 15 mM NaCl; 60 mM KCl; 1 mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0; 0.5 mM Spermidine) supplemented with 0.8 mM S-adenosyl-methionine (NEB, cat. no. B9003S). A negative control was prepared without MTase. The reactions were mixed by gentle flicking before a 1 h incubation at 37 °C. Reactions were quenched (Monarch PCR & DNA Cleanup Kit) and the purified DNA eluted in 20 μl elution buffer. Twelve restriction

enzyme digests were prepared by combining 15 µl of each purified DNA sample with 1 µl *Dpn*I (NEB, cat. no. R0176S) and 4 µl 10× CutSmart Buffer (NEB) in a 40 µl reaction. The reactions were mixed by flicking then incubated at 37 °C for 1.5 h; 1 µl of each reaction was combined with 2 µl of 6× Purple Gel Loading Dye (NEB, cat. no. B7024S) and 11 µl H_2O and run on a 1.2% agarose gel containing 1× GelGreen Nucleic Acid Stain (Biotium, cat. no. 41005) at 130 V for 1.5 h. The gel was imaged on an Azure C200 Gel Imager. MTase activity was determined by the highest MTase dilution that methylates 1 µg of DNA substrate leading to no fully intact DNA molecules after *Dpn*I digestion.

Mitochondrial isolation

HeLa and U2-OS cells were grown in 150 mm plates to a confluency of 70%, and myoblasts to a confluency of 80–90%. Cells were pelleted by spinning at 150*g* for 5 min at 4 °C. Medium was removed and cells were resuspended in 4 ml cell lysis buffer (10 mM Tris, pH 7.5; 10 mM NaCl; 1.5 mM $MgCl_2$). Cells were allowed to swell on ice for 7.5 min. Cells were then lysed by douncing in a 7 ml dounce with 20 strokes and 2M Sucrose T10E20 buffer (10 mM Tris, pH 7.6; 1 mM EDTA, pH 8.0; 2 M sucrose) was added to bring the final sucrose concentration to 250 mM. Cell debris was pelleted by centrifuging lysates at 1,300*g* for 3 min at 4 °C. Supernatants were transferred to fresh tubes and spun again at 1,300*g* for 3 min at 4 °C. Supernatants were again transferred to fresh tubes, and mitochondria were pelleted by spinning at 18,000*g* for 15 min at 4 °C.

Proteinase K protection assay

HeLa S3 cells were grown in 150 mm plates to a confluency of 70%. Mitochondria were isolated as described above. The mitochondria pellet was resuspended in 100 µl permeabilization buffer (20 mM Tris, pH 7.4; 70 mM potassium acetate; 250 mM sucrose; 0%/0.1%/0.25%/0.5% Tween-20; 0%/0.1%/0.25%/0.5% NP-40/Igepal-630). Mitochondria were permeabilized on ice for 10 min and then pelleted by spinning at 18,000*g* for 15 min at 4 °C. Mitochondrial pellets were resuspended in 80 µl proteinase K reaction buffer (20 mM Tris, pH 7.5; 70 mM potassium acetate; 250 mM sucrose). Samples were split in two and either 5 µl buffer or 5 µl 1 µg $µl^{-1}$ proteinase K (Millipore Sigma, cat. no. 3115887001) was added. Samples were mixed by gently flicking and incubated for 20 min at 37 °C. Reactions were quenched by adding 5 µl 100 mM phenylmethylsulfonyl fluoride (PMSF) and heat inactivated for 10 min at 95 °C. Proteins were assessed by western analysis using anti-TOMM40 1:2,000 (Proteintech, cat. no. 18409-1-AP), anti-COX1 1:1,000 (Abcam, cat. no. 14705) and anti-HSP60 1:1,000 (Cell Signaling, cat. no. D307).

Mitochondrial ATAC-seq

ATAC-seq was performed as previously described[28] using TDE1 (Illumina, cat. no. 20034197) with modifications. Mitochondria were first isolated as described above. Mitochondria were resuspended in 50 µl cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20) and immediately centrifuged for 10 min at 4 °C at 18,000*g*. Mitochondrial pellets were resuspended in 50 µl transposition reaction mix (25 µl TD, 2.5 µl TDE1, 22.5 µl H_2O) and incubated at 37 °C for 30 min. Following transposition, samples were immediately purified using the Qiagen MinElute PCR Purification kit (Qiagen, cat. no. 28004). DNA was minimally amplified with one cycle of 5 min at 72 °C and 30 s at 98 °C followed by five cycles of 10 s at 98 °C, 30 s at 63 °C and 1 min at 72 °C. A 5 µl sample of partially amplified DNA was used in quantitative PCR reactions, and the number of additional cycles to amplify was determined as the number of cycles corresponding to one third of the maximum fluorescence intensity signal by quantitative PCR. The samples were then amplified with one cycle of 30 s at 98 °C followed by the determined number of cycles for 10 s at 98 °C, 30 s at 63 °C and 1 min at 72 °C. Following amplification, samples were assessed on an Agilent BioAnalyzer and subjected to paired-end Illumina sequencing.

ATAC-seq alignment and processing

Paired-end reads were aligned using bowtie2 (v.2.2.9) either to Hg38 or to just chrM of Hg38. The sam file output was converted to a bam file using samtools (v.1.3.1) and sorted. Duplicate reads were filtered using PICARD (v.2.8.0). Aligned reads were then filtered for those having a mapping quality score greater than 30 using samtools (-q 30). Read coverage was determined using the bamCoverage command of deeptools (v.3.0.2) with a binSize of 1, normalized using counts per million, and extending the reads by 100. Data were visualized using Integrative Genomics Viewer (v.2.4.9).

Mitochondrial Fiber-seq

Mitochondria were first isolated as described above. Mitochondrial pellets were resuspended in permeabilization buffer containing 0.25% Tween-20 and 0.25% NP-40/Igepal-630 unless otherwise indicated. After permeabilization, mitochondria were pelleted by centrifugation at 18,000*g* for 15 min at 4 °C. Supernatant was removed and pellets were resuspended in 56 µl mtFiber-seq reaction buffer (20 mM Tris, pH 7.4; 70 mM potassium acetate; 250 mM sucrose). Then, 1.5 µl 32 mM S-adenosylmethionine was added to a final concentration of 0.8 mM. Tubes were transferred to a thermocycler prewarmed to 37 °C. Reactions were started by adding 500 U Hia5, unless otherwise indicated, and mixing. The methylation reaction was allowed to proceed for 10 min at 37 °C, unless otherwise indicated. For the 120 min sample, an additional 1.5 µl of 32 mM SAM was spiked in after 60 min. Reactions were quenched by adding 3 µl 20% SDS and mixing. Sample volume was then increased to 200 µl by adding 7 µl 20% SDS and buffer. Protein was degraded by adding 2 µl 18.2 mg ml^{-1} proteinase K and incubating at 55 °C for 1 h. Phenol:chloroform:isoamyl alcohol extractions were performed to extract DNA, and DNA was precipitated by standard ethanol precipitation protocols using 0.1 volumes of sodium acetate and 2.5 volumes of ethanol. DNA was pelleted and washed with 70% ethanol. DNA was air-dried for 10 min at room temperature before resuspending in 81 µl H_2O. NEB CutSmart Buffer (10 µl) was then added along with 3 µl RNase A (Thermo Fisher Scientific, cat. no. AM2270) and restriction enzymes. For all samples, 3 µl *Xma*I (NEB, cat. no. R0180S) was added to fragment chromatin. An additional 3 µl of *Bam*HI-HF (NEB, cat. no. R3136L) was added to linearize mtDNA from all cell lines with the exception of HSMM, for which 3 µl of *Eag*I-HF (NEB, cat. no. R3505) was added due to the mitochondrial genome in this cell line having a single nucleotide polymorphism resulting in a second *Bam*HI cut site. Reactions were incubated at 37 °C for 1 h. A phenol:chloroform:isoamyl extraction was performed followed by a chloroform:isoamyl extraction to remove any remaining phenol. DNA was precipitated as before. DNA was pelleted and washed with 70% ethanol. After air-drying, DNA was resuspended in 75 µL Buffer EB (Qiagen cat. no. 19086) and quantified by Qubit hsDNA (Thermo Fisher Scientific, cat. no. Q32851). Samples were then subjected to PacBio library preparation protocols according to the manufacturer's instructions.

Mapping single-molecule mtFiber-seq reads and m6A identification

Using the raw subread bam files and modified chrM reference genomes from Hg38, we ran a custom pipeline combined with Fibertools[66] on a SLURM managed cluster (see 'Data and code availability'). Modified chrM genomes were used as the reference depending on how the particular library was constructed: (1) linearized with *Bam*HI (hg38_chrM_BamHI.fa), (2) linearized with *Eag*I (hg38_chrM_EagI.fa) or (3) synthesized by long-range PCR (LRPCR) (hg38_chrM_lrpcr.fa). This pipeline takes the raw PacBio subread bam files and converts them into a 12-column BED file containing mapped m6A positions.

Specifically, subread bam files are first split into 60 chunks using the split_subreads.sh and bamseive_batch.slurm scripts. Circular consensus sequencing (CCS) reads were then generated for each chunk using *ccs* (v.6.4.0) with the –hifi-kinetics flag to include averaged kinetic

information. Next, chunks were recombined into a single bam file using merge_bam.sh. Fibertools (v.0.1.3) predict-m6a was then run on the combined bam file with the -s flag. Reads were then aligned to the appropriate reference genome with *pbmm2* (v.1.9.0) using the -preset CCS, -sort, -sort-memory 1 G, -unmapped and -log-level INFO flags. Following this, Fibertools extract was run using the -r and -min-ml-score. To account for differing PacBio chemistries, a range of -min-ml-score thresholds from 232 to 255 were used to generate BED files at each cutoff.

Tn5 expression and purification

Tn5 was expressed as previously described[67] with modifications. pTXB1-Tn5 (Addgene, cat. no. 60240) was transformed into T7 Express LysY/Iq (NEB, cat. no. C3013). Two 1 l cultures of LB supplemented with 100 μg ml^{-1} ampicillin were inoculated with 10 ml of overnight culture. Cells were grown at 37 °C with shaking until the OD_{600} reached 0.55. Protein expression was induced with 0.2 mM IPTG, the temperature was reduced to 18 °C and cells were allowed to express for 18 h. Cells were spun down at 2,700*g* at 4 °C for 20 min. The cell pellet was resuspended in 80 ml of Tn5 lysis/wash buffer (20 mM HEPES, pH 7.2; 0.8 M NaCl; 1 mM EDTA; 10% glycerol; 0.2% Triton X-100) supplemented with 1× complete, EDTA-free protease inhibitor cocktail (Millipore Sigma, cat. no. 11873580001). Cells were lysed by probe sonication (Qsonica Q125) for 10 min on ice at 50% amplitude, 30 s on/off for a total of 20 min. The lysate was clarified by centrifugation at 30,596*g* for 30 min at 4 °C. To the supernatant, 2.1 ml 10% polyethyleneimine was added dropwise on a magnetic stirrer. The precipitate was removed by centrifugation at 17,210*g* for 10 min at 4 °C. Clarified lysate was added to 10 ml chitin resin (NEB, cat. no. S6651S) pre-equilibrated with Tn5 lysis/wash buffer and rocked for 1 h at 4 °C. The resin was washed with 300 ml Tn5 lysis/wash buffer. Protein was eluted by adding Tn5 chitin elution buffer (20 mM HEPES, pH 7.2; 0.8 M NaCl; 1 mM EDTA; 10% glycerol; 0.2% Triton X-100; 100 mM DTT). The first 11 ml were collected and discarded. The resin was kept at 4 °C for 48 h to allow the intein fusion to cleave and elute the protein from the Chitin resin. After 48 h, 1 ml fractions were collected and the fractions containing protein were determined using detergent compatible Bradford assay (Thermo Scientific, cat. no. 23246). Fractions were dialyzed against 1 l of Tn5 size exclusion chromatography (SEC) buffer (50 mM Tris, pH 7.5; 800 mM NaCl; 0.2 mM EDTA; 2 mM DTT; 10% glycerol) overnight at 4 °C. The sample was injected and run on an ENrich SEC 650 10 × 300 24 ml size exclusion column (Bio-Rad, cat. no. 7801650) using Tn5 SEC Buffer. Peak fractions were collected and dialyzed into 2× Tn5 dialysis buffer (100 mM HEPES, pH 7.2; 0.2 M NaCl; 0.2 mM EDTA; 2 mM DTT; 0.2% Triton X-100; 20% glycerol). Protein was concentrated using a 10K spin concentrator (Millipore Sigma, cat. no. UFC801008) and the concentration was determined using detergent compatible Bradford assay. Glycerol was added to a final concentration of 55% and protein was stored at −80 °C. Protein purity was confirmed by SDS–PAGE.

Tn5 transposome assembly

The transposome assembly was performed as previously described[40] with modifications. In brief, adapter oligonucleotides were ordered from Integrated DNA Technologies, the reverse oligonucleotide Tn5MErev (5′-[phos]CTGTCTCTTATACACATCT-3′), Tn5ME-A-ATTO488 (5′-/ATTO488N/TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′) and Tn5ME-B-ATTO488 (5′-/ATTO488N/GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAG-3′); 100 μM Tn5MErev was mixed in equal amounts either with 100 μM Tn5ME-A-ATTO488 or 100 μM Tn5ME-B-ATTO488. The oligonucleotide mixtures were denatured for 5 min at 95 °C in a thermocycler. The cycler was shut off and the oligonucleotides left inside the thermomixer until they reached room temperature. The transposome was assembled in a mixture of mixed oligonucleotides (Tn5MErev/Tn5ME-A-ATTO488 and Tn5MErev/Tn5ME-B-ATTO488) (final concentration 12.5 μM for each mix), 31.75% glycerol (final concentration reached 47.9 %), 0.24× Tn5 Dialysis Buffer and 5 μM Tn5. The mixture was mixed by pipetting up and down carefully then incubated at room temperature for 60 min. The assembled transposome was stored at −20 °C.

Tn5 in vitro activity assay

To determine optimal buffer conditions for the assembled transposome, an in vitro tagmentation of a plasmid (pET30a-6xHis-ΔN-TFAM) was performed. Plasmid (50 ng) was incubated with or without 100 nM preassembled Tn5. Four reaction buffers were tested; Buffer 1 (2× TD-Buffer (10 mM KCl, 75 μM $Na_2HPO_4{\cdot}7H_2O$, 274 mM NaCl, 50 mM Tris, pH 7.4 (ref. 68)), Buffer 2 (2× TD-Buffer (33 mM Tris, pH 7.8, 66 mM potassium acetate, 11 mM magnesium acetate and 16% N,N-dimethylformamide[69]), Buffer 3 (2× TD-Buffer (20 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 20% N,N-dimethylformamide[70]) and Buffer 4 (5× TAPS-PEG (50 mM TAPS-NaOH, pH 8.5, 25 mM $MgCl_2$, 40% PEG 8000 (ref. 67)). Reactions were performed in 1× concentration of the respective buffer. The reaction mixtures were incubated for 30 min at 37 °C. Reactions were quenched by the addition of 2× concentrated washing buffer (100 mM EDTA, 0.02% SDS, 2× PBS) and incubated for 15 min at 55 °C. Samples were brought to a volume of 200 μl, and DNA was precipitated by the addition of 2 mM $MgCl_2$, 60 mM sodium acetate pH 5.5 and 80% v/v ethanol. The samples were incubated for 2 h on ice and the DNA was pelleted for 15 min at full speed in a tabletop centrifuge at 4 °C. The pellet was washed with 70% ethanol and dried at room temperature. The DNA was resuspended in water and the concentration measured by Nanodrop. To analyze the tagmentation process, the samples were analyzed using the Agilent TapeStation D1000.

Mitochondrial ATAC-see

ATAC-see was performed as previously described[40,69] with modifications. U2-OS cells were grown in μ-Slide eight-well glass bottom chambers (no. 1.5H glass bottom, ibidi, cat. no. 80827) until 60–70% confluent. Cells were washed three times with 1× PBS and fixed for 10 min with 1% methanol-free formaldehyde (Thermo Fisher, cat. no. 28906) in 1× PBS at room temperature. After three washes with 1× PBS, cells were permeabilized with 1× PBS, 0.25% TX-100 for 20 min. Cells were washed three times with 1× PBS and the chamber slide was incubated in a hybridization oven (Boekel Scientific, RapidFISH) prewarmed to 37 °C. Tn5 (100 nM) was used in tagmentation buffer (16.5 mM Tris, pH 7.8, 33 mM potassium acetate, 5.5 mM magnesium acetate and 8% N,N-dimethylformamide). For the inactive Tn5 control, the mixture included 50 mM EDTA. After mixing by pipetting, the transposase was centrifuged for either 10 (EFD 3) or 20 min (EFD 4 C) at maximum speed at 4 °C in a tabletop centrifuge to reduce aggregates in the sample. The cells were incubated with the transposase for 60 min at 37 °C in the hybridization oven in the dark. The transposase was inactivated and washed away in three washes for 15 min at 37 °C with prewarmed washing buffer (1× PBS, 50 mM EDTA, 0.1% TX-100). Afterwards, the cells were rinsed three times with 1× PBS at room temperature.

Immunofluorescence and 4′,6-diamidin-2-phenylindol staining after mitochondrial ATAC-see

For immunofluorescence, ATAC-see samples were incubated for 1 h at room temperature protected from light with blocking buffer (1× PBS, 0.1% TX-100 and 5% normal goat serum (Vector Laboratories, cat. no. S-1000-20). ds/ssDNA mouse monoclonal IgM antibody clone AC-30-10 (PROGEN, cat. no. 61014) was used in a 1:100 dilution in blocking buffer. To be able to segment single cells, the rabbit monoclonal anti-sodium potassium ATPase antibody (EP1845Y) (Abcam, cat. no. ab76020) was used in a 1:250 dilution in blocking buffer. The antibody mix was added to the cells, and they were incubated at 4 °C protected from light overnight or for 1 h at room temperature. The samples were washed three times for 5 min each at room temperature with 1× PBS and 0.1% TX-100 followed by an incubation for 1 h at room temperature protected from light with the secondary antibodies. The goat anti-mouse IgM mu chain

(Alexa Fluor 647) (Abcam, cat. no. ab150123) and the goat anti-rabbit IgG (H+L) cross-adsorbed secondary antibody (Alexa Fluor 546) (Invitrogen, cat. no. A-11010) were used in a 1:1,000 dilution in blocking buffer. After the 1 h incubation, the samples were washed three times for 5 min at room temperature protected from light. After three rinses with 1× PBS, 5 µg ml^{-1} 4′,6-diamidin-2-phenylindol (DAPI) in 1× PBS was added. The samples were incubated for 10 min protected from light and washed again three times with 1× PBS. Glycerol mounting media (80% glycerol, 1× PBS, 20 mM Tris, pH 8, 2.5 g ml^{-1} n-propyl-gallate) was added to all wells.

Immunofluorescence and DAPI staining

To test the detection of mtDNA, U2-OS cells were grown in µ-Slide eight-well glass bottom chambers until 80% confluent. Cells were washed once with 1× PBS and fixed using 4% formaldehyde in 1× PBS for 1 h at room temperature. After three washes with 1× PBS, cells were permeabilized for 20 min with 1× PBS and 0.1% TX-100 followed by three washes using 1× PBS. For the TFAM overexpression, HeLa S3 cells were grown in µ-Slide eight-well glass bottom chambers with high walls (cat. no. 80807) and TFAM-HA expression was induced by addition of 100 ng ml^{-1} doxycycline (Dox) for 48 h. DMSO (diluted 1:1,000) served as a control. TFAM-HA was expressed for 48 h. Cells were rinsed once with 1× PBS and fixed for 20 min with 4% formaldehyde in 1× PBS. Washing and permeabilization were performed as described before. The immunofluorescence was performed the same for both experiments.

After permeabilization and washing, the buffer was replaced by blocking buffer and the samples were incubated for 1 h at room temperature. The primary antibodies were diluted in blocking buffer, added to the samples and incubated overnight at 4 °C. The antibodies were used in the following dilutions: for the DNA detection in the mitochondrial network, the ds/ssDNA mouse monoclonal IgM antibody was used in a 1:250 dilution and in the TFAM-HA overexpression experiment in a 1:162 dilution. The TOM20 (F-10) (Santa Cruz, cat. no. sc-17764) and the HA-Tag (C29F4) Rabbit mAb (Cell Signaling, cat. no. 3724) were used in a 1:500 dilution. On the next day samples were washed three times for 5 min using 1× PBS and 0.1% TX-100. The secondary antibodies were diluted 1:1,000 in blocking buffer. For Fig. 3b and Extended Data Figs. 2b and 3b,c mtDNA anti-mouse IgM Alexa Fluor 647 and Goat anti-Mouse IgG2a cross-adsorbed secondary antibody, Alexa Fluor 555 (Thermo Fisher, cat. no. A-21137) were used. For TFAM overexpression (Fig. 3b) the following antibodies were used: goat anti-mouse IgM (heavy chain) cross-adsorbed secondary antibody, Alexa Fluor 555 (Thermo Fisher, cat. no. A-21426), goat anti-rabbit IgG (H+L) highly cross-adsorbed secondary antibody, Alexa Fluor 647 (Thermo Fisher, cat. no. A-21245), goat anti-mouse IgG2a cross-adsorbed secondary antibody, Alexa Fluor 488 (Thermo Fisher, cat. no. A-21131). The samples were incubated for 1 h at room temperature and in the dark. Afterwards, the cells were washed again three times for 5 min with 1× PBS and 0.1% TX-100. After three rinses with 1× PBS, 5 µg ml^{-1} DAPI in 1× PBS was added. The samples were incubated for 10 min in the dark and washed again three times with 1× PBS. No DAPI stain was used for the TFAM-HA overexpression experiment. Glycerol mounting media (80% glycerol, 1× PBS, 20 mM Tris, pH 8, 2.5 g ml^{-1} n-Propyl-Gallate) was added to all wells.

Microscopy

As controls for imaging, one well was empty and was used to image the background for flatfield correction. To be able to perform collar correction with the silicon oil objective and to test the point spread function as well as chromatic aberrations, fluorescent blue/green/orange/dark red TetraSpec Microspheres size 0.2 µm (Thermo Fisher Scientific, cat. no. T7280) were used. The spheres were first vortexed to allow for well resuspension; 1 µl spheres were mixed with 9 µl 100% ethanol and vortexed. The mix was spread onto the well with a nipped tip and allowed to dry for 30 min. Glycerol mounting medium was added. The spheres were always prepared on the same day as the samples to allow the addition of the mounting medium at the same time.

For imaging, a Nikon Ti2 inverted microscope with a W1 Yokogawa Spinning disk with a 50 µm pinhole disk, laser lines 405, 488, 561, 640 (Nikon Instruments, Laser Unit model Lun-F), emission filters (455/50, 480/40, 525/36, 605/52, 630/75, 705/72), a Nikon motorized stage, a Physik Instrument Piezo Z motor, Andor Zyla 4.2 Plus sCMOS monochrome VSC-04833 camera and a Plan Apo λS SR HP ×100C/1.45 Silicon differential interference contrast (DIC) objective were used. The Nikon Elements Acquisition Software AR v.5.02 was used. Slides were cleaned first with Sparkle, water and finally with 100% ethanol before imaging to remove any salt and oil residues. A silicon immersion oil 30cc (Nikon, cat. no. MXA22179) was used in combination with the silicon oil objective. The laser power was adjusted using no-EDTA samples. The same exposure times (green channel, 700 ms; all others 500 ms) were used for the biological replicates. No binning was used. The 16-bit dual gain ¼ setting was used. The final image had a pixel size of 0.065 µm × 0.065 µm; 30 fields of view (FoVs) and 27 *z*-planes were imaged on each session of an empty well to perform averaging of the FoVs and use for flatfield correction. To assess the noise detected by the detector, 100 frames with lasers off were taken and averaged for background subtraction of samples and the flatfield image. *z*-Stacks of 29 or 27 steps of 200 nm were imaged using the NIDAQ piezo stage. The shutter was off during image acquisition and all fluorescence channels one after the other (starting with green, red, far red and, finally, blue) were taken for each *z*-plane before moving to the next plane. All images were acquired at room temperature and saved as .ND2 files. The immunofluorescence of DNA and TOM20 was imaged with the same microscope as described above. Again, the silicon oil objective was used. A *z*-stack of 19 planes with each plane sized 300 nm was imaged. The 16-bit dual gain ¼ setting was used. The final image had a pixel size of 0.065 µm × 0.065 µm. The exposure time for each channel was 500 ms and again all channels were first imaged before moving to the next *z*-plane. For Fig. 3b, the Plan Apo λ ×100/1.45 oil differential interference contrast objective in combination with immersion oil (Cargille, nondrying immersion oil Type 37, cat. no. 16237) was used. Again, no binning and 16-bit dual gain ¼ setting was used. A *z*-stack of 38 planes with a plane size of 300 nm was imaged. The final image had a pixel size of 0.065 µm × 0.065 µm. Again the exposure times were 500 ms.

Image analysis

For image analysis, Fiji, Arivis and R were used. Images were exported as ome.tiff files for analysis. First, sample images were flatfield corrected. The FoVs of the empty well were averaged using a custom macro in Fiji. Next, the camera noise was subtracted from the averaged flatfield correction image. Since the background subtraction of the camera noise leads to negative pixel values, an offset of 200 was added to the sample images as well as the corrected flatfield image. Next, the camera noise was subtracted from the sample image and the result of this was divided by the flatfield corrected image.

$$((\text{sample} + 200) - \text{camera noise})/((\text{flatfield} - \text{camera noise}) + 200)$$

For replicate 1, the first five images had 29 *z*-planes. Two *z*-planes were deleted to gain the same number as the other images. *z*-Planes without signals were chosen to be removed. The corrected images could be used then for segmentation in Arivis. First, nuclei were segmented based on the DAPI stain and masked from the image. Therefore, the denoising module was used with the following settings: bilateral, diameter was set to 1 pixel, sensitivity was set to 24.3. Next, an intensity threshold segmenter was used with the settings, 'simple,' 'bright' and the 'visible range' of 0.789–25. The threshold was set to 1. The segment morphology module was used to erode objects for 11 pixels to reduce detected mtDNA. The erode function was performed planewise. To detect the full nuclei size, the objects were dilated again planewise using the segment morphology filter by 11 pixels. The resulting objects

were used to mask the nuclei in the next segmentation step. mtDNA was segmented based on the ds/ssDNA-antibody signal. A morphology filter was used to preserve bright objects with a size of 80 pixels. The option 'spheres' was chosen. Afterwards, the nuclei were masked using the already created objects. The blob finder function was used to segment the nucleoids with the following settings: the probability threshold was set to 5% and the split sensitivity to 60%. The diameter was set to 1 pixel. The blob finder function segments objects in three-dimensional space. Objects smaller than five voxels were filtered out, since they are considered artifacts. The object features were extracted. Single cells were segmented by hand based on the signal of the anti-sodium potassium ATPase antibody. These masks were used to assign the previously segmented nucleoids to single cells using the compartment function of Arivis. A total of 30 cells were segmented for no-EDTA (seven FoVs) and EDTA (five FoVs) for replicate 1 and 29 cells for no-EDTA (seven FoVs) and 30 cells (nine FoVs) for EDTA for replicate 2. The object features of those objects were extracted and used for analysis in R. In the case of cells where the nuclei mask did not cover the full signal and the mtDNA segmentation pipeline detected objects, those objects were deleted by hand to avoid artifacts. The extracted intensities were either analyzed in bulk, considering only objects assigned to single cells, or as single cells. For figures, brightness and contrast were adapted in Fiji. For the TFAM overexpression cell line (Fig. 3b), a median filter of pixel size 2 was used. For Extended Data Fig. 3d screenshots of the different segmentation steps were taken in Arivis. Brightness and contrast had been adjusted before to represent the signals.

mtFiber-seq footprint identification

To identify footprints, a HMM was applied with two hidden states: accessible and inaccessible. To account for any sequence biases, the HMM observations were considered to be a methylation (m6A) or a lack of methylation (A) observed at every possible hexamer sequence context (−3 nucleotides, +3 nucleotides). For example, AAAm6AAAA was treated as a separate observation from TAAm6AAA. The emission probabilities for these hexamer sequence context observations were calculated based on experimental data. For the probability of methylation given an accessible state, data from mtFiber-seq performed on mtDNA generated from LRPCR was used. For the probability of methylation given an inaccessible state, and to account for biases in the methylation caller or low levels of endogenous methylation, data from mtFiber-seq performed on isolated mitochondria without Hia5 (MTase) were used. Observations at a position with a C or G were given a methylation emission probability of 0. The transition and starting probabilities were trained using 1% of HeLa mtFiber-seq data 20 times, with the reads shuffled each time and different starting parameters used. The starting parameters for each training were generated by sampling the Dirichlet distribution with all parameters set to 1. Each training converged to almost identical starting and transition probabilities. The model was applied to both experimental data and to a sampled portion of the LRPCR product data. A minimum mean posterior probability of 0.95 was identified as a threshold that provided a 1% false discovery rate for footprints in the accessible data. Footprint enrichment was calculated as the fraction of reads, either total or with a defined level of methylation, protected at a particular base. Although developed independently, a similar approach to calling footprints has been successful in other contexts[36]. Footprint size enrichment at genomic loci were visualized using matplotlib[71] along with seaborn[72] in Python.

Strand specificity of methylation

To calculate methylation strand bias, A (for light strand methylation) and T (for heavy strand methylation) were represented as the values 1 and 0, respectively. Using a 150 nucleotide sliding window, methylations (1 or 0) across all reads were averaged. Each window was required to have at least 2,250 methylations across all reads combined or was assigned a value of 'NA'. Values >0.5 thus represent a light strand methylation bias, and vice versa. The reference sequence A/T content was then evaluated using an identical approach and methylation A/T values were normalized by the reference A/T values and $\log_2$-transformed. D-loop boundaries are shown at the coordinates 16,080–16,100 and 210–230. To calculate an overall strand bias score, all methylations across the region (for example, 16,100–210) for all reads were recorded as 1 or 0, then the mean was calculated.

7S DNA D-loop identification

Reads with a high ratio of light strand methylations to heavy strand methylations in the D-loop region were determined and called as probably containing a D-Loop. This ratio was defined as $(mA_{light} + 1)/(mA_{heavy} + 1)$. A minimum count of seven total methylations in the region was used to call a D-loop and a Gaussian mixture model (GMM) was applied to the distribution of the log of the ratios. A ratio threshold of 3.01 was identified to call a D-loop, based on a GMM posterior probability of 0.99.

Footprint co-occurrence

Footprint co-occurrence on a read was tallied using the following filters: minimum size for all, 20 nucleotides; maximum size across the left D-loop boundary (16,080:16,100), 60 nucleotides; maximum size across the right D-loop boundary (210:230), 140 nucleotides; maximum size across the MTERF binding site (3,232:3,253), 35 nucleotides. Minimum overlap of footprint with site: six nucleotides. D-loop presence in a read was determined by the ratio of light to heavy strand methylations as described above. Counts were normalized by the total number of reads in the categories considered and shown as a percentage.

In vitro MTase substrate specificity assay

Hia5 MTase substrate specificity was measured using the MTase-Glo methyltransferase assay (Promega) according to manufacturer's instructions. Oligonucleotides 27 nucleotides in length were used as substrate: 5′-TGACATGAACACAGGTGCTCAGATAGC-3′ (A) and 5′-GCTATCTGAGCACCTGTGTTCATGTCA-3′ (B). Oligonucleotides were resuspended in duplex buffer (100 mM potassium acetate; 30 mM HEPES, pH 7.5) to a final concentration of 100 µM. Oligonucleotides A and B were mixed at a ratio of 1:1, heated to 94 °C for 2 min, and cooled slowly to room temperature to anneal the dsDNA substrate. Oligonucleotide A was used as the ssDNA substrate. Optimal enzyme concentration was determined by varying Hia5 with 1 µM dsDNA. Reactions were allowed to proceed for 30 min at room temperature before quenching, developing and measuring luminescence using a Tecan Infinite 200 FPlex plate reader. To measure substrate specificity, 0.02 U μl^{-1} Hia5 was used with increasing concentrations of dsDNA or ssDNA. Reactions were allowed to proceed for 30 min before quenching with 0.5% trifluoroacetic acid, developing and measuring luminescence.

mtRNA NanoString quantification

HeLa S3 cells were treated with either DMSO or 2CMA dissolved in DMSO (100 µM final concentration) for 2, 4, 6 or 24 h. For each timepoint, cells were harvested and counted using a Countess II Cell Counter (Thermo Fisher Scientific). Cells were diluted to a concentration of 2,000 cells μl^{-1} and pelleted by spinning at 300*g* for 5 min at 4 °C. Cells were resuspended in 1 ml RLT buffer (Qiagen, cat. no. 79216) and vortexed for 1 min. Cells were then diluted in RLT buffer to make 200 µl aliquots at 1,000 cells μl^{-1} and 100 µl aliquots at 500 cells μl^{-1}. Cells were flash frozen in liquid nitrogen and stored at −80 °C. To hybridize NanoString probes (modified from MitoString profiling probe set[73]), frozen cells were thawed on ice; 4 µl probe stocks A and B were each diluted with 29 µl TE Tween (TE + 0.1% Tween-20). Then 70 µl hybridization buffer was added to the probe set, 7 µl of diluted probe stock A was added to the diluted probe set, mixed by inverting, and spun down for 2–3 s in a tabletop centrifuge and 7 µl of diluted probe stock B and 84 µl H_2O were added, mixed by inverting, and spun down for 2–3 s in a

tabletop centrifuge. A 14 µl sample of this master mix was transferred to each required PCR tube; 1 µl of thawed lysate was added to each tube, the sample was mixed by flicking and briefly spun down using a tabletop centrifuge. Samples were then hybridized by incubating at 67 °C for 16 h.

Relative mtDNA content quantification

Relative mtDNA content was quantified as previously described[74] with modifications. Total DNA was isolated from cell culture using the Qiagen QIAamp DNA mini kit. DNA concentrations were determined by Nanodrop. PCR reactions were assembled by combining 10 µl SsoFast EvaGreen Supermix (Bio-Rad, cat. no. 1725201), 0.8 µl of 10 µM forward primer, 0.8 µl of 10 µM reverse primer, 3 µl of 20 ng µl^{-1} template DNA and 5.4 µl H_2O. Samples were cycled as follows: 50 °C for 2 min, 95 °C for 10 min, 40 cycles of 95 °C for 15 s and 62 °C for 60 s. Samples were performed in technical replicate for each PCR target. C_q values for the three technical replicates were averaged. The relative mtDNA content was calculated as $2 \times 2^{\Delta CT}$ where ΔCT is defined as $Nuclear_{CT} - Mitochondrial_{CT}$. For nuclear DNA, primers targeting the µglobulin gene (forward: 5′-TGCTGTCTCCATGTTTGATGTATCT-3′, reverse: 5′-TCTCTGCTCCCCACCTCTAAGT-3′) were used, and for mtDNA primers targeting the Leu tRNA gene were used (forward: 5′-CACCCAAGA ACAGGGTTTGT-3′, reverse: 5′-TGGCCATGGGTATGTTGTTA-3′).

RNA-sequencing

RNA-seq libraries from differentiating HSMM were prepared using the SMARTer Stranded Total RNA HI Mammalian Kit (Takara, cat. no. 634873) in biological duplicates (days 0, 3 and 6). Libraries were sequenced on an Illumina NovaSeq with paired-end 100 nucleotide reads at the Harvard Medical School's Biopolymers Facility. Reads were trimmed with cutadapt v.1.14 and aligned using STAR v.2.5.4a with default parameters to the reference hg38 genome. Rsubread's featureCounts was used to count uniquely mapping reads across coding sequences and differential expression analysis was performed using DESeq2 (ref. 75). The genes *GAPDH*, *HPRT1* and *RPS12* were used in the controlGenes parameter of the DESeq2 estimateSizeFactors function[76].

TFAM overexpression

For TFAM overexpression, TFAM-HA TetOn HeLa S3 cells were grown on 150 mm plates to a confluency of 50%. Cells were treated with either DMSO or 100 ng ml^{-1} Dox for 48 h. Medium containing either DMSO or 100 ng ml^{-1} Dox was replaced after 24 h. Following treatment, mtFiber-seq was performed according to the above protocol using equal cell numbers, and TFAM overexpression levels were confirmed by western blot analysis using an anti-TFAM antibody (Santa Cruz, cat. no. sc-376672) and normalized with an anti-β-Actin antibody (Cell Signaling, cat. no. 3700).

Subsampling of datasets

To directly compare the footprinting patterns of datasets with often widely varying levels of methylation, a subsampling approach to match their overall methylation distributions was employed using sample_bed.py. The overall number of methylations were counted either across the entire read for the overexpression datasets, or in the coding region (Genome positions 4,000–14,000) to avoid the NCR and mTERF1 binding site, for the in vitro datasets. The distribution of methylation counts were then equalized in both datasets by downsampling the set of reads with a given methylation count in a dataset to match the count in the other dataset. These samplings were repeated ten times with different random seeds to make sure that any conclusions derived from them were stable.

TEV protease expression and purification

TEV protease was purified as previously described[77] with modifications. MBP-TEV plasmid was transformed into T7 Express LysY/Iq cells. Overnight cultures were grown and four 1 l cultures of 2× LB supplemented with 100 µg ml^{-1} ampicillin were inoculated with 10 ml overnight culture. Cultures were grown at 37 °C with shaking until the OD_{600} reached 0.4–0.5. Cultures were transferred to 18 °C and grown until the OD_{600} reached 0.6–0.8. Expression was induced by adding 400 µl of 1 M IPTG per 1 L of culture. Protein was expressed for 18 h at 18 °C with shaking. Cells were spun down at 4,220*g* for 25 min. Pellets were resuspended in 60 ml TEV lysis/wash buffer (1× PBS, pH 7.5; 300 mM KCl; 10% glycerol; 7.5 mM imidazole). Cells were lysed by probe sonication (Qsonica, cat. no. Q125) for 10 min on ice at 50% amplitude, 30 s on/off. Lysate was clarified by centrifuging for 20 min at 25,000*g* at 4 °C. Supernatant was transferred to new tubes and spun at 40,000*g* for 25 min at 4 °C. Clarified lysate was added to 6 ml Ni-NTA agarose resin pre-equilibrated in TEV lysis/wash buffer. Lysate mixture was incubated for 1 h at 4 °C with rocking. Resin was spun down in 50 ml conicals at 1,000*g* for 3 min at 4 °C. Resin was washed with 150 ml TEV lysis/wash buffer. Protein was eluted with 20 ml TEV elution buffer (25 mM HEPES, pH 7.5; 100 mM KCl; 500 mM imidazole). Protein was concentrated using a 3 kDa cutoff spin concentrator (Millipore Sigma, cat. no. UFC9003) to a volume of 2 ml. Protein was injected and run on a HiLoad 16/60 S200 column (Cytiva, cat. no. 28989335) pre-equilibrated in TEV storage buffer (25 mM HEPES, pH 7.5; 300 mM KCl; 10% glycerol; 2 mM DTT). Fractions were analyzed by SDS–PAGE, and pure fractions were pooled. Concentration was determined by the absorbance at a wavelength of 280 nm (A_{280}) using an extinction coefficient of 36,130 M^{-1} cm^{-1}. TEV was stored at 3 mg ml^{-1} at −80 °C.

TFAM expression and purification

The pET30a_ΔN_TFAM plasmid was transformed into C43 (DE3) *E. coli* cells (BioSearch Technologies, cat. no. 60345-1), and overnight cultures were grown in LB supplemented with 25 µg ml^{-1} kanamycin. Four 1 l cultures of LB supplemented with 25 µg ml^{-1} kanamycin were inoculated with 15 ml each of overnight culture. Cultures were grown at 37 °C with shaking until the OD_{600} reached 0.6. Protein expression was induced with 1 mM IPTG and cells were induced for 24 h at 25 °C with shaking. Cells were harvested by spinning at 4,220*g* for 20 min. Pellets were resuspended in 80 ml TFAM lysis buffer (50 mM Tris, pH 7.4; 300 mM NaCl; 5 mM imidazole; 1 mM β-mercaptoethanol) supplemented with 1× Complete, EDTA-free protease inhibitor cocktail. Cells were lysed by probe sonication (Qsonica, cat. no. Q125) for 10 min on ice at 50% amplitude, 30 s on/off. Lysate was supplemented with 5 µl benzonase to reduce *E. coli* genomic DNA, and lysate was clarified by centrifugation in Oak Ridge tubes at 30,596*g* for 30 min. Lysate was combined with Cobalt TALON resin pre-equilibrated with TFAM lysis buffer and incubated for 1 h at 4 °C with rocking. Resin was washed with 120 ml TFAM lysis buffer, and protein was eluted with 10 ml TFAM elution buffer (50 mM Tris, pH 7.4; 300 mM NaCl; 500 mM imidazole; 1 mM β-mercaptoethanol). Elution was concentrated to 2 ml using a 10 kDa cutoff spin concentrator (Millipore Sigma, cat. no. UFC9010) and supplemented with 100 µl 3 mg ml^{-1} TEV protease. Sample was dialyzed overnight against heparin load/wash buffer (25 mM Tris, pH 7.4; 100 mM NaCl; 5 mM DTT; 10% glycerol). Sample was injected on Heparin HP column and eluted with a 0 to 100% gradient with heparin load/wash buffer and heparin elution buffer (25 mM Tris, pH 7.4; 1 M NaCl; 5 mM DTT; 10% glycerol) over 20 column volumes. Peak fractions were concentrated to 1 ml and injected onto a HiLoad 16/60 S200 column pre-equilibrated with TFAM size exclusion buffer (25 mM Tris, pH 7.4; 150 mM NaCl; 5 mM DTT; 10% glycerol). Peak fractions were pooled and concentrated, the purity determined by SDS–PAGE and the concentration determined by NanoDrop using the absorbance at 280 nm with 35,410 M^{-1} cm^{-1}.

TFAM in vitro DNA binding

TFAM binding activity was confirmed using a fluorescence polarization-based method. DNA oligonucleotides corresponding

to the heavy strand promoter were annealed by heating to 95 °C and slowly cooling to room temperature. The forward oligonucleotide (5′-GGTTGGTTCGGGGTATGGGGTTAGCAGC-3′) contained a 5′ FAM fluorophore. DNA (100 nM) was mixed with protein at a volume of 20 µl in TFAM binding buffer (25 mM Tris, pH 7.4; 150 mM NaCl; 1 mM DTT; 0.01% NP-40) in a black Corning 384-well plate (Corning, cat. no. 3575). Samples were equilibrated for 30 min at room temperature. Polarization was measured using a Tecan Infinite 200 Pro FPlex plate reader using 485 nm/20 nm bandwidth excitation filters and 530 nm/25 nm bandwidth emission filters fitted with polarizers. Data were fit using FPobs = [Protein] × FPmax + K_D × FPmin / [Protein] + K_D.

mtDNA LRPCR amplification

LRPCR was performed to generate linear full-length mtDNA. mtDNA was amplified from HeLa genomic DNA that was isolated from HeLa S3 cells using the Qiagen QiAmp DNA Mini Kit according to the manufacturer's instructions. Takara LA HS Polymerase (Takara, cat. no. RR042) was used with 16426 Forward (5′-CCGCACAAGAGTGCTACTCTCCTCGCTC-3′) and 16425 Reverse (5′-GATATTGATTTCACGGAGGATGGTGGTCAAGG GACC-3′) primers using 100 ng template in 50 µl reactions. DNA was amplified using a two-step amplification protocol: 95 °C for 2 min, 30 cycles of 95 °C for 20 s and 68 °C for 13 min and a final extension at 68 °C for 18 min. DNA was purified using a Zymo Genomic DNA Clean & Concentrator 10 kit (Zymo Research, cat. no. D4011), pooling five 50 µl reactions per column. DNA was eluted with 50 µl elution buffer prewarmed to 50 °C by applying to the column, incubating for 5 min, and spinning for 1 min at 16,000*g*. Elution was reapplied to the column, incubated for 5 min and spun for 1 min at 16,000*g*. Several elutions were pooled and concentrated by SpeedVac. DNA concentration was determined using Qubit hsDNA kit (Thermo Fisher Scientific, cat. no. Q32851) according to the manufacturer's instructions. DNA length and purity was confirmed using a 0.5% agarose gel as well as by TapeStation using a gDNA tape.

In vitro mtFiber-seq

Methylation of mtDNA generated by LRPCR was performed using 1 µg template DNA in a final volume of 60 µl with a range of recombinant TFAM concentrations. DNA was mixed with TFAM in TFAM binding buffer (25 mM Tris, pH 7.4; 150 mM NaCl; 1 mM DTT) and equilibrated for 30 min at room temperature. Samples were supplemented with 0.8 mM S-adenosylmethionine and reactions were started by adding 200 U Hia5 and mixing. Samples were incubated for 10 min at 37 °C before quenching using a Zymo Genomic DNA Clean & Concentrator 10 kit according to the manufacturer's instructions. Samples were eluted with 50 µl elution buffer prewarmed to 50 °C. Elution buffer was incubated on the column membrane for 5 min before spinning. Elution was reapplied to the membrane, incubated for 5 min and then centrifuged. DNA concentration was determined by Qubit hsDNA assay. Integrity of the DNA was confirmed by TapeStation using a gDNA tape, and methylation was measured by DNA dot blot. DNA was subjected to PacBio library construction and sequencing.

DNA methylation dot blot

DNA was diluted in 20× SSC (3 M NaCl; 300 mM trisodium citrate) buffer in a 96-well plate and denatured at 95 °C for 10 min. Nitrocellulose membrane (Bio-Rad, cat. no. 1620112) was wetted in 20× SSC buffer and secured in a Bio-Dot Microfiltration Apparatus (Bio-Rad, cat. no. 1706545) according to the manufacturer's instructions. The membrane was washed with 100 µl 20× SSC per well and the vacuum applied until all liquid was pulled through. Then, 150 µl sample was applied to each well (150 µl 20× SSC was applied to unused wells) and pulled through by vacuum. Membrane was placed face up on dry Whatman filter paper and crosslinked with 125 mJ in a GS Gene Linker UV Chamber (Bio-Rad) using the C-L setting. The membrane was washed briefly with 1× TBST (10 mM Tris, pH 7.5; 0.25 mM EDTA; 150 mM NaCl; 0.1% TWEEN-20) and blocked in 1× TBST + 5% nonfat dry milk for 1 h at room temperature. Rabbit polyclonal anti-N6-methyladenosine antibody (Active Motif, cat. no. 61995) was diluted to 2 µg ml^{-1} in 1× TBST + 5% nonfat dry milk and incubated overnight at 4 °C. The blot was washed three times in 1× TBST for 15 min each. The anti-rabbit IgG, HRP-linked secondary (Cell Signaling, cat. no. 7074S) was diluted 1:5,000 in 1× TBST + 5% nonfat dry milk and incubated with the blot for 1 h at room temperature. Three washes were repeated and the blot was developed with ECL substrate (Cytiva, cat. no. RPN2106) and imaged using a Bio-Rad ChemiDoc MP system.

Genome-wide TFAM $K_{1/2}$ determination and $GN_{10}G$ enrichment

Footprints for in vitro mtFiber-seq datasets were identified as described above. For each read, each genome position was determined as being bound if it overlapped with a footprint at least 20 bp in size. The fraction bound was determined at each position as the number of reads protected at a site with a footprint at least 20 bp in size divided by the total number of reads. The fraction bound was performed at each genomic position for each TFAM concentration: 0, 5, 10, 20 and 30 µM. The $K_{1/2}$ was determined using a four-parameter logistics regression, with the minimum and maximum forced to be 0 and 1, respectively. The strongest affinity sites throughout the genome were identified as those with the top 5% of $1/K_{1/2}$ values. The center of each stretch of high-affinity position was determined, resulting in the identification of 35 highest affinity sites. A site was determined to have a $GN_{10}G$ motif if one was present within a ±15 bp window of the high-affinity site.

Reporting summary

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

Data availability

Raw and processed sequencing data are available from the Gene Expression Omnibus (GEO) at accession number GSE212611. PolG and TWINKLE ChIP–seq data were provided by X. Zhu. All microscopy data are available on Open Microscopy Environment (OMERO) at https://omero.hms.harvard.edu/webclient/?show=project-10105. Custom genome fasta files used for alignment based on the Hg38 reference genome (https://ncbi.nlm.nih.gov/grc/human) are available at https://www.github.com/churchmanlab/mtFiberseq. Source data are provided with this paper.

Code availability

Code for analysis of PacBio sequencing data available on GitHub: https://github.com/churchmanlab/mtFiberseq.

References

66. Jha, A. et al. DNA-m6A calling and integrated long-read epigenetic and genetic analysis with fibertools. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.20.537673 (2023).
67. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
68. Cold Spring Harbor Protocols. TD buffer. *Cold Spring Harb. Protoc.* **2010**, db.rec12138 (2010).
69. Yan, K. et al. Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Sci. Adv.* **6**, eaax0021 (2020).
70. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
71. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
72. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

73. Wolf, A. R. & Mootha, V. K. Functional genomic analysis of human mitochondrial RNA processing. *Cell Rep.* **7**, 918–931 (2014).
74. Rooney, J. P. et al. PCR based determination of mitochondrial DNA copy number in multiple species. *Methods Mol. Biol.* **1241**, 23–38 (2015).
75. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
76. Masilamani, T. J., Loiselle, J. J. & Sutherland, L. C. Assessment of reference genes for real-time quantitative PCR gene expression normalization during C2C12 and H9c2 skeletal muscle differentiation.*Mol. Biotechnol.* **56**, 329–339 (2014).
77. Shih, Y.-P., Wu, H.-C., Hu, S.-M., Wang, T.-F. & Wang, A. H.-J. Self-cleavage of fusion protein in vivo using TEV protease to yield native protein. *Protein Sci.* **14**, 936–941 (2005).

Acknowledgements

We thank H. Merens and additional members of the Churchman laboratory for helpful discussions and assistance; K. Choquet, N. Kramer and B. Smalec for critical reading of the manuscript; L. Tallon, L. Sadzewicz and the University of Maryland School of Medicine Genomics Resource Center for PacBio sequencing; X. Zhu (University of Gothenburg) for Polɣ and TWINKLE ChIP–seq data from HeLa S3 cells; B. Battersby (Institute of Biotechnology, University of Helsinki) for human myoblasts from anonymous healthy control samples; P. Montero Llopis and P.V. Anekal of the MicRon facility (Harvard Medical School) for providing the microscope and consultation for imaging and analysis; A. Bergen and M. Burdyniuk (Arivis) for help with analysis; the Harvard University Bauer Core Facility for Illumina sequencing services and the IDDRC Molecular Genetics Core at Boston Children's Hospital for NanoString services. This work was supported by the National Institutes of Health (R01-GM123002 to L.S.C., DP5-OD029630 to A.B.S. and F32-GM130028 to R.S.I.) and a Helen Hay Whitney Foundation Fellowship (F-1240 to K.G.H.). A.B.S. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew Biomedical Scholar.

Author contributions

R.S.I. and L.S.C. conceptualized the study. R.S.I. developed the methodology, with input from K.G.H, T.W.T., A.B.S. and L.S.C. R.S.I. led the investigation, with assistance from K.G.H. R.S.I., K.G.H., T.W.T., M.C. and D.D. performed formal analyses. R.S.I. wrote the original draft of the manuscript. R.S.I., K.G.H., T.W.T., M.C., A.B.S. and L.S.C. reviewed and edited the manuscript. R.S.I., K.G.H., A.B.S. and L.S.C. acquired funding. A.B.S. and L.S.C. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

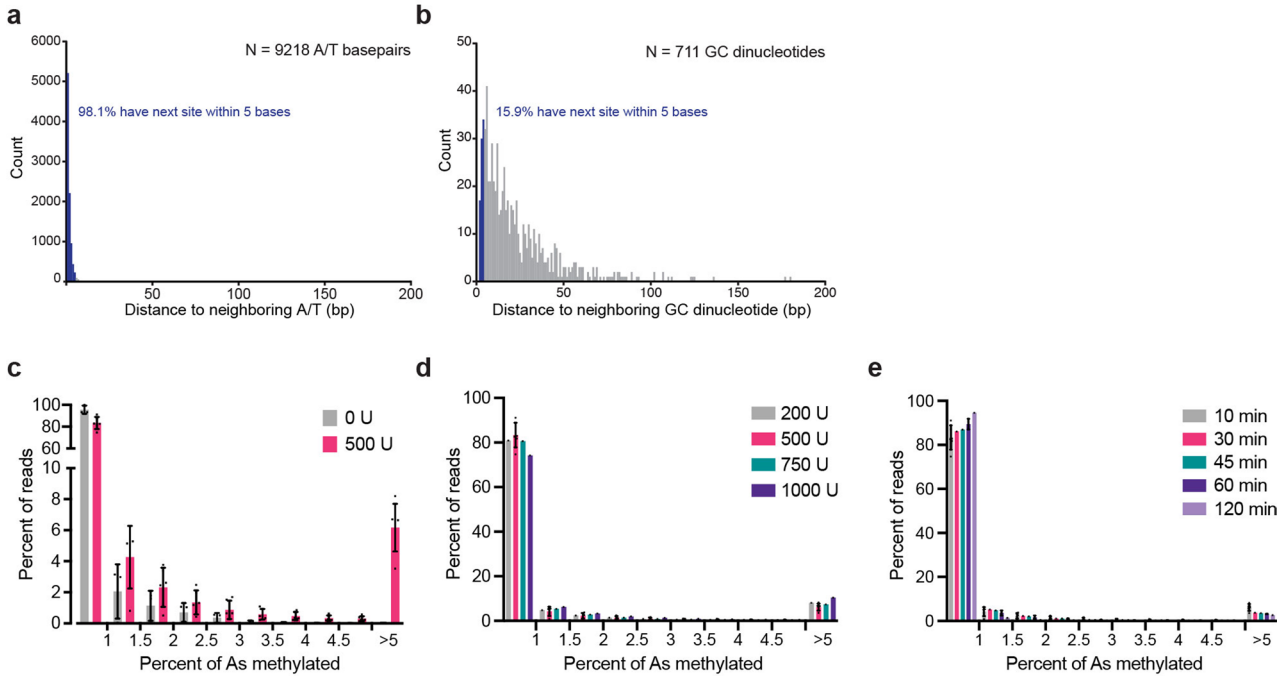
Extended data is available for this paper at https://doi.org/10.1038/s41594-024-01225-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-024-01225-6.

Correspondence and requests for materials should be addressed to Andrew B. Stergachis or L. Stirling Churchman.

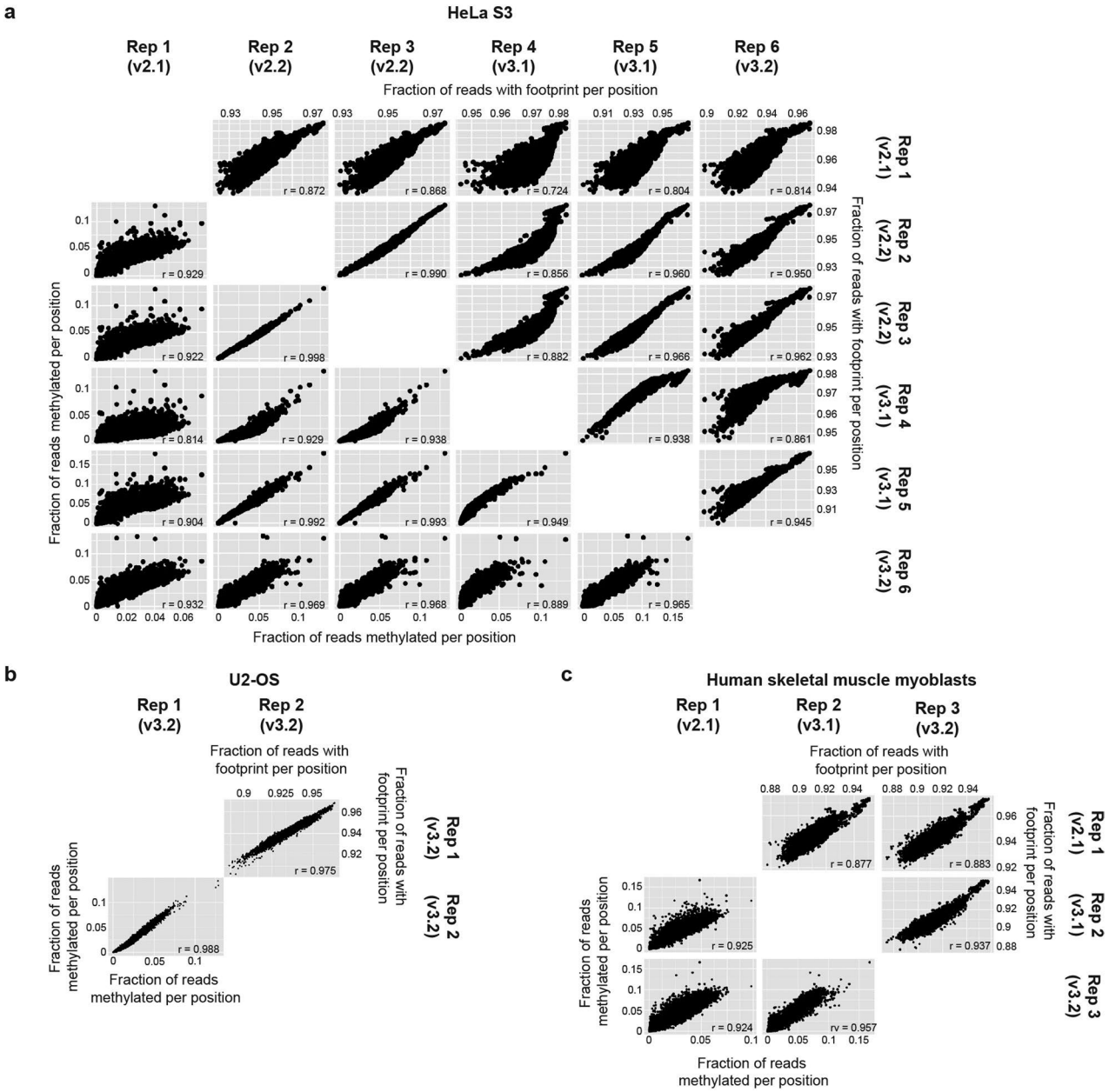
Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Sara Osman was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

https://doi.org/10.1038/s41594-024-01225-6



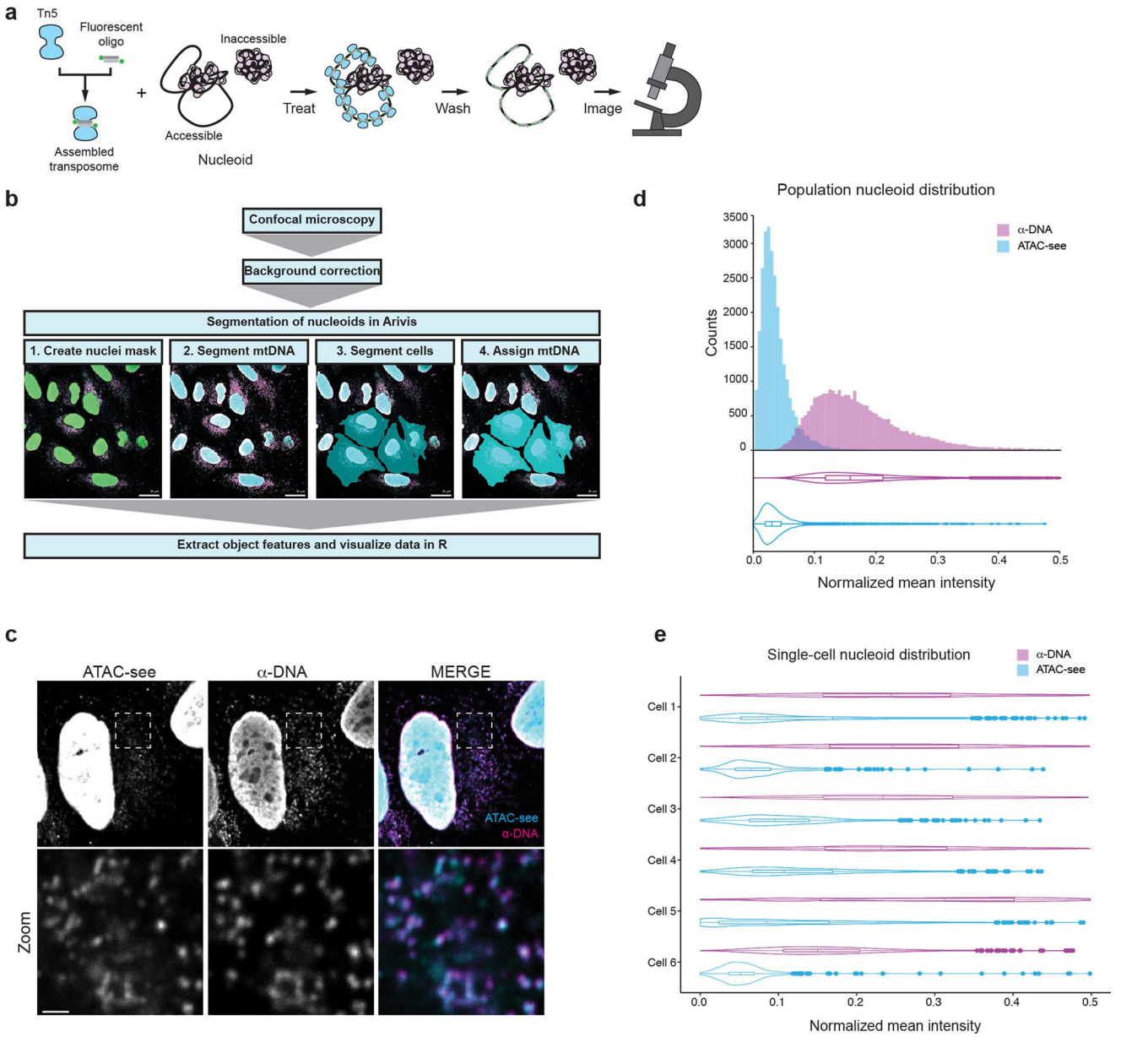
Extended Data Fig. 1 | Resolution of adenine methyltransferases with mtDNA. (**a**) Histogram showing the distance to the next neighboring A/T nucleotide from the 9,218 A/T nucleotides present in the human mitochondrial genome. (**b**) Histogram showing the distance to the next GC dinucleotide from the 711 present in the human mitochondrial genome. (**c**) Bar plot showing the percent of reads binned by the number of m6A modifications per read comparing untreated samples and those treated with 500 U of the m6A-MTase Hia5. Individual dots represent three and six biological replicates for 0 U Hia5 and 500 U Hia5, respectively. Error bars represent s.d. of the mean. (**d**) Bar plot showing the percent of reads binned by the number of m6A modifications per read for samples treated with 200, 500, 750, and 1000 U of the m6A-MTase Hia5. Individual dots represent six biological replicates for 500 U Hia5. Error bars represent s.d. of the mean. (**e**) Bar plot showing the percent of reads binned by the number of m6A modifications per read for samples treated with 500 U of the m6A-MTase Hia5 for 10, 30, 45, 60, and 120 minutes. The 120 minute sample received an additional SAM spike-in after 60 minutes. Individual dots represent 6 biological replicates for 10 minutes and 2 biological replicates for 60 minutes. Error bars represent s.d. of the mean.

https://doi.org/10.1038/s41594-024-01225-6



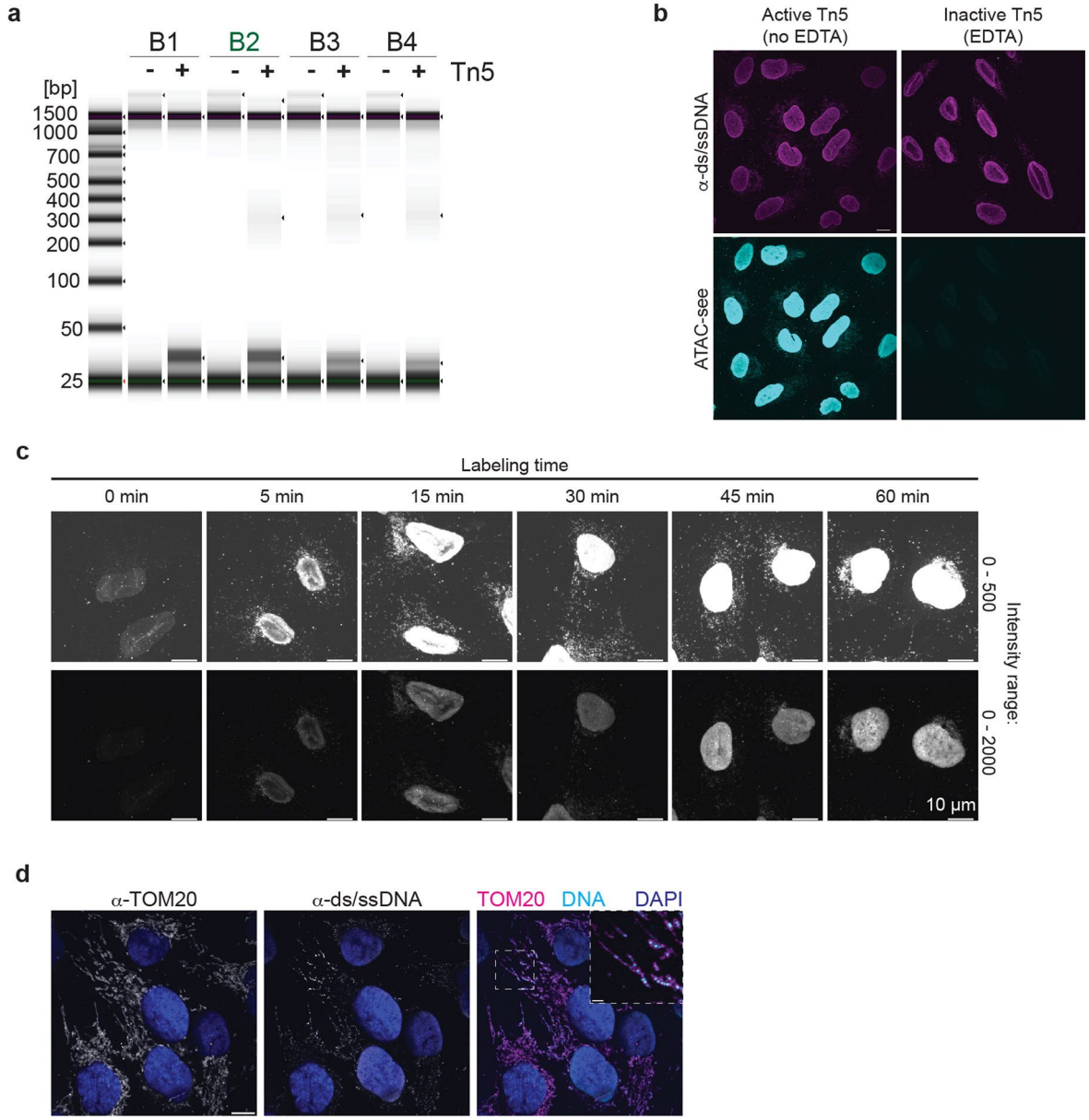
Extended Data Fig. 2 | Reproducibility of mtFiber-seq and HMM footprint calling. (**a**) Correlation scatter plots comparing six mtFiber-seq samples from HeLa S3 cells for (bottom left) fraction of reads methylated at each adenine and (top right) fraction of reads with a footprint at each genomic position. PacBio chemistry version is indicated for each replicate. Pearson's correlation coefficient is shown for each correlation. (**b**) Correlation scatter plots comparing two mtFiber-seq samples from U2-OS cells for (bottom left) fraction of reads methylated at each adenine and (top right) fraction of reads with a footprint at each genomic position. PacBio chemistry version is indicated for each replicate. Pearson's correlation coefficient is shown for each correlation. (**c**) Correlation scatter plots comparing three mtFiber-seq samples from undifferentiated human skeletal muscle myoblasts for (bottom left) fraction of reads methylated at each adenine and (top right) fraction of reads with a footprint at each genomic position. PacBio chemistry version is indicated for each replicate. Pearson's correlation coefficient is shown for each correlation.

https://doi.org/10.1038/s41594-024-01225-6



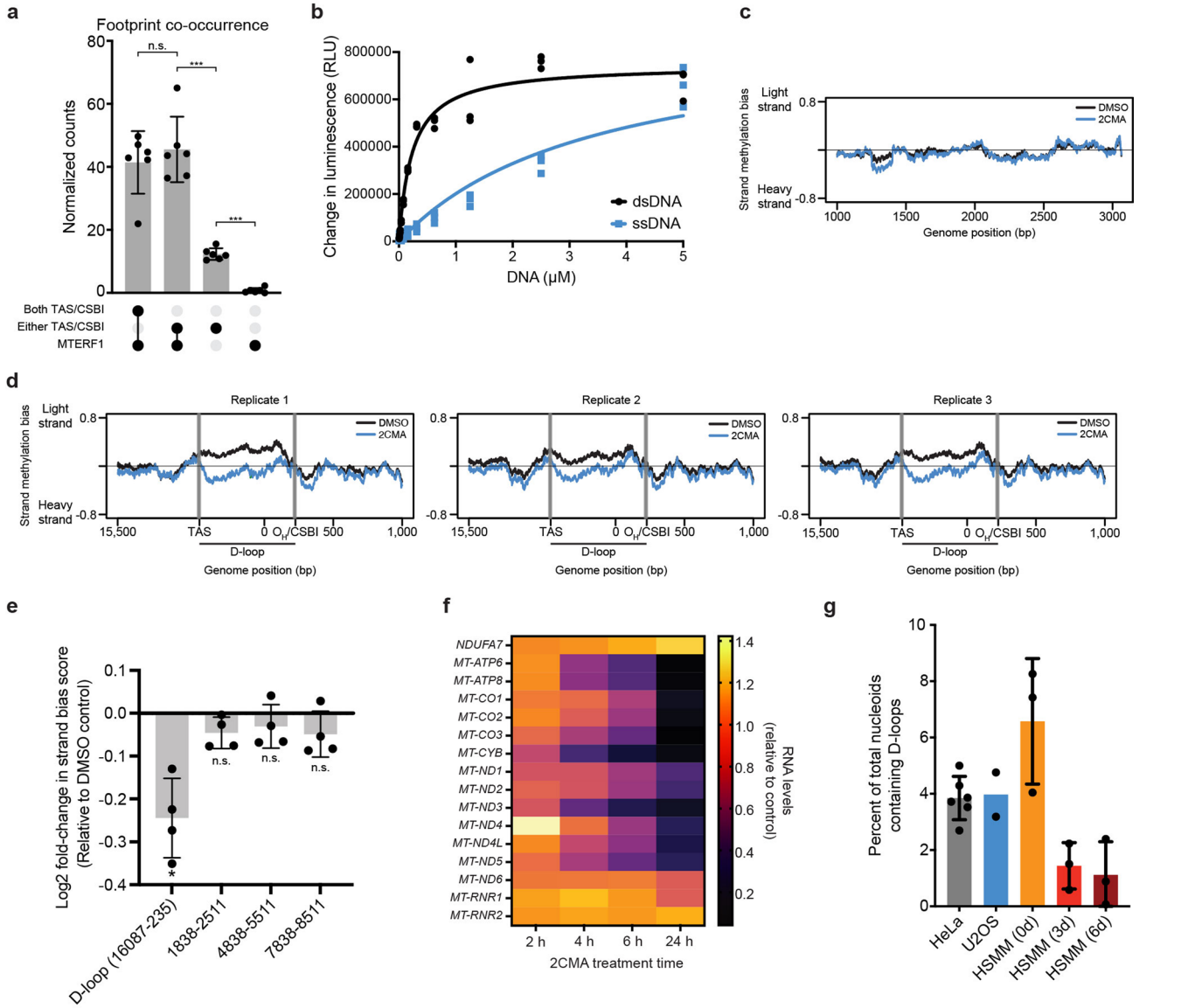
Extended Data Fig. 3 | ATAC-see reveals intracellular nucleoid heterogeneity. (**a**) Schematic depicting experimental design. Tn5 was loaded with ATTO488-labeled oligo nucleotides to form active transposomes. U2-OS cells were treated with transposome and imaged by confocal fluorescence microscopy (**b**) Schematic depicting the ATAC-see segmentation and analysis pipeline. Background corrected images were masked and segmented in Arivis. Features of assigned objects were extracted and analyzed. (**c**) Representative image of a U2-OS cell showing ATAC-see and DNA signals. DNA was labeled with an anti-ss/dsDNA antibody that shows preferential labeling of mtDNA (Scale bars, 5 µm for single cell, 1 µm for zoom) (**d**) Histogram and violin plot showing the distribution of ATAC-see and DNA signal. Shown are the min-max normalized mean intensities from 27,079 segmented objects. Box-plot elements are as follows: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points represent outliers. N = 54,459 segmented objects from 57 cells across two biological replicates. (**e**) Distribution of ATAC-see and DNA signal from 6 individual U2-OS cells. Shown are the min-max normalized mean intensities from each segmented object. Box-plot elements are as follows: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points represent outliers. N = 1,394, 792, 1,890, 1,558, 3,440, and 1,254 segmented objects from Cells 1, 2, 3, 4, 5, and 6, respectively.

https://doi.org/10.1038/s41594-024-01225-6



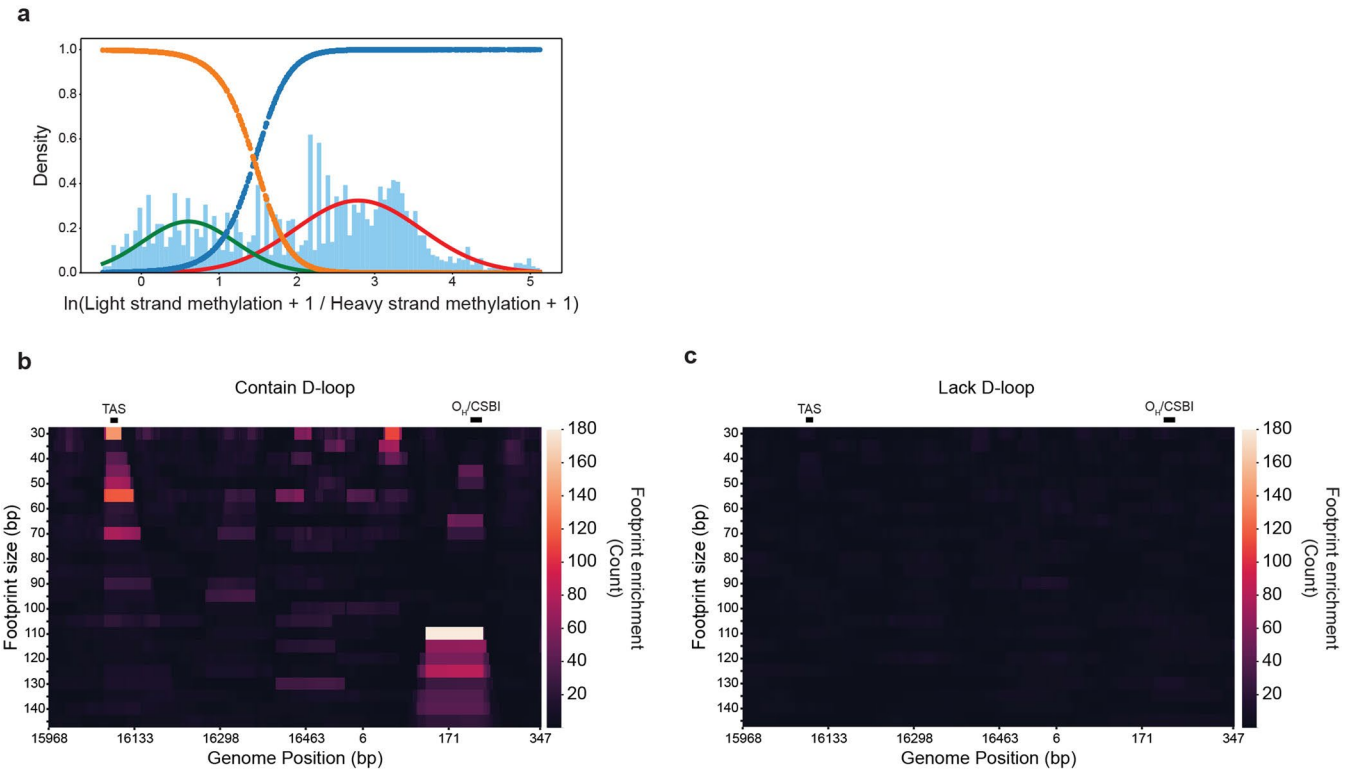
Extended Data Fig. 4 | Comparison of ATAC-see reaction conditions. (**a**) Tn5 *in vitro* activity in four different reaction buffers measured by DNA fragment analysis. Assembled transposome was mixed with 50 ng plasmid DNA for 30 minutes at 37 °C and DNA fragments were assessed by Agilent TapeStation D1000. Tn5 activity fragments the plasmid DNA, resulting in the appearance of smaller (< 500 bp) bands. Four buffers were tested: B1 (50 mM Tris, pH 7.4, 10 mM potassium chloride, 75 µM disodium phosphate, 274 mM sodium chloride), B2 (33 mM Tris, pH 7.8, 66 mM potassium acetate, 11 mM magnesium acetate, 16% N,N-dimethylformamide), B3 (20 mM Tris, pH 7.6, 10 mM magnesium chloride, 20% N,N-dimethylformamide), and B4 (50 mM TAPS, pH 8.5, 25 mM magnesium chloride, 40% PEG8000). Buffer B2 was used for ATAC-see reactions performed in Extended Data Figs. 3 and 4. (**b**) Z-projection of the max intensities of a background corrected field-of-view of U2-OS cells treated with Tn5 transposomes in the presence and absence of EDTA. DNA was labeled with an α-ss/dsDNA antibody. (Scale bar, 10 µm) Representative image of four independently repeated experiments. (**c**) U2-OS cells sowing ATAC-see signal after treatment with Tn5 over a 60 minute time course. Two intensity ranges are shown to highlight the nuclear and mitochondrial signals.(Scale bar, 10 µm). Experiment performed once to determine optimal labeling time for subsequent experiments. (**d**) Confocal fluorescence microscopy showing mtDNA labeling throughout the mitochondrial network. A single Z-plane (0.3 µm) is shown. The mitochondrial network is labeled with an α-TOM20 antibody, mtDNA with an α-ss/dsDNA antibody, and chromatin with DAPI (Scale bars, 10 µm, 2 µm for zoom). Experiment performed once to confirm mtDNA signal within the mitochondrial network.

https://doi.org/10.1038/s41594-024-01225-6



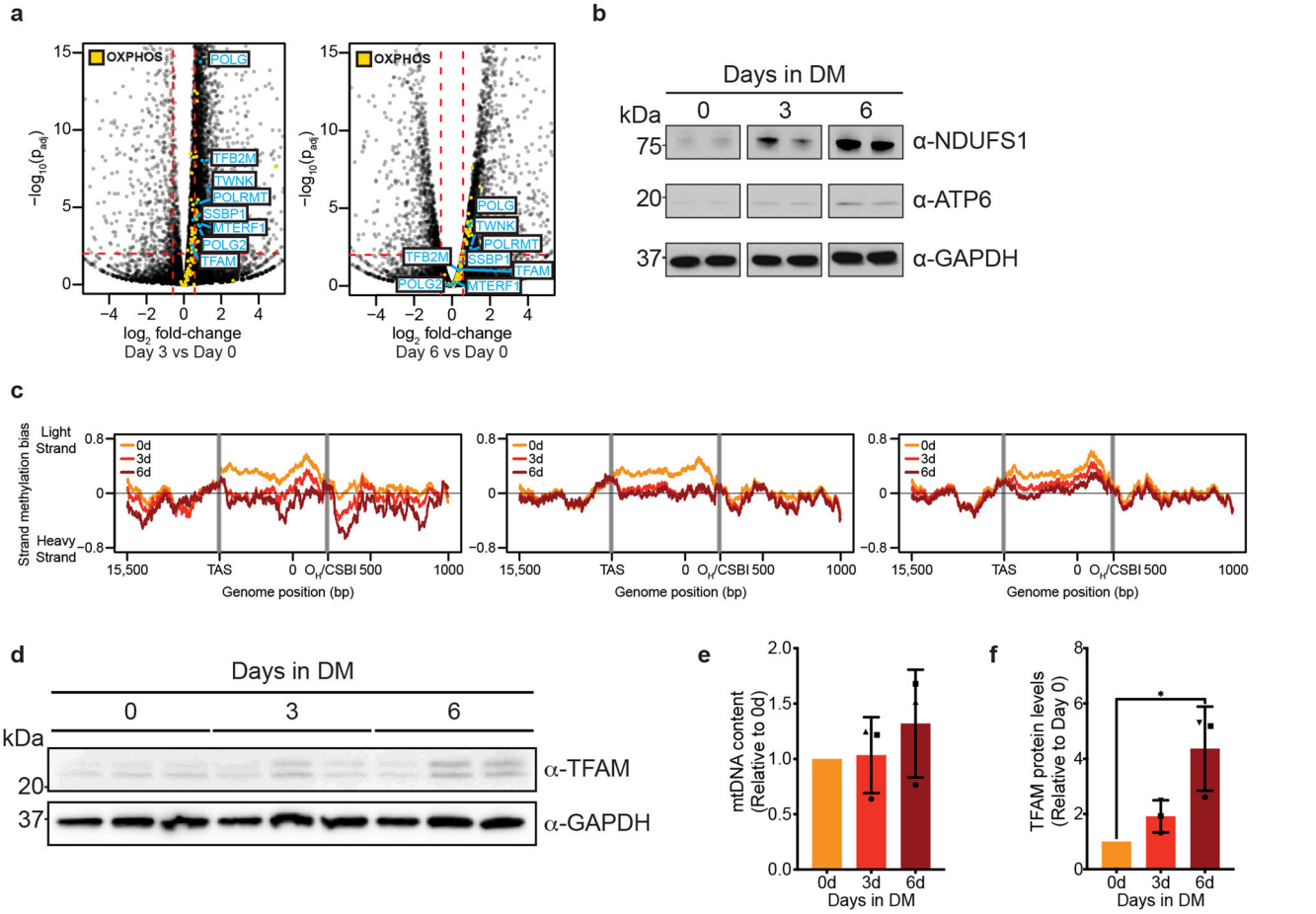
Extended Data Fig. 5 | mtFiber-seq exposes unique D-loop features. (**a**) UpSet plot showing the co-occurence of footprints at the TAS, CSBI, and MTERF1 binding site. Maximum footprint sizes were set based on the footprint size distributions: 60 bp (TAS), 140 bp (CSBI), and 35 bp (MTERF1). Reads with larger footprints were omitted. N = 6 biologically independent samples. Paired two-sided Student t-tests were used to compare the frequency of footprint co-occurrence: n.s. $P > 0.05$; ***$P < 0.001$. From left to right, specific p-values are 0.0008 (Category 2 vs Category 3) and <0.0001 (Category 3 vs Category 4). The categories in the plot represent 1.74%, 1.94%, 3.75%, 1.66%, 4.05%, and 2.08% of the total molecule population for replicates 1-6. (**b**) Hia5 activity on ssDNA and dsDNA. The K_m for dsDNA is 0.233 μM. A lower limit of 3.48 μM was set for the K_m for ssDNA as the reaction never saturated. N = 3 replicates. (**c**) Methylation strand bias from positions 1,000 to 3,000 in untreated HeLa cells and cells treated with 2CMA. Methylation bias is calculated as the number of methylations on the light and heavy strands, averaged using a 150 nt window and normalized against the region's AT content. (**d**) Same as (**c**) but for the NCR for three biological replicates. (**e**) Log_2 fold change in the methylation strand bias score between 2CMA treated and control samples at four loci. N = 4 biological replicates. Samples were compared with a two-sided Student's t-test: *$P < 0.05$; n.s. $P > 0.05$. Exactly p-value is 0.015 for D-loop region 2CMA vs DMSO. (**f**) Heatmap showing RNA levels in HeLa cells after treatment with 2CMA as measured by NanoString. Counts were internally normalized to *GAPDH* and calculated relative to the DMSO control. All RNAs shown are mitochondrially encoded except for *NDUFA7*. N = 3 biological replicates. (**g**) Bar plot showing the percent of total nucleoids containing D-loops. A nucleoid was defined as having a D-loop if it had at ≥7 methylations within the region and a strand methylation ratio ≥3.01. N = 6 biological replicates for HeLa S3, 2 biological replicates for U2-OS, and 3 biological replicates for HSMM. For all plots, error bars represent s.d. of the mean.

https://doi.org/10.1038/s41594-024-01225-6



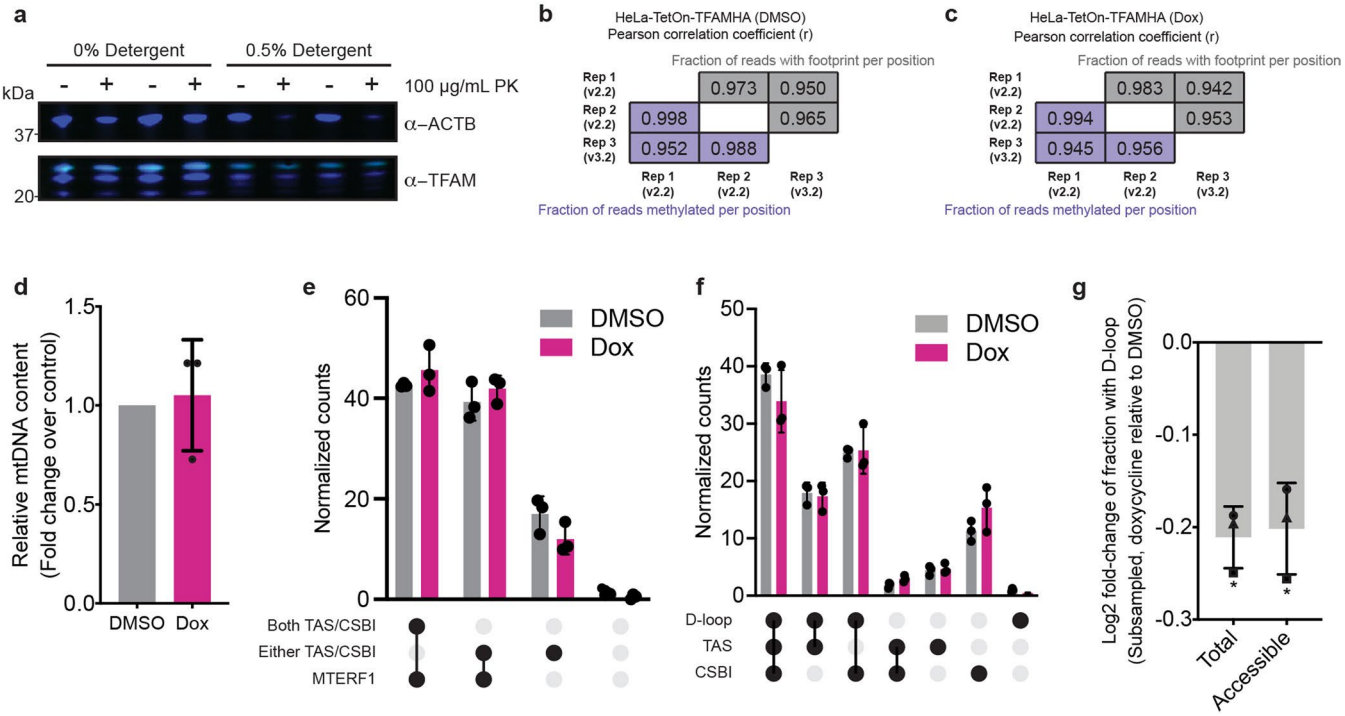
Extended Data Fig. 6 | TAS/CSBI footprints are associated with D-loop containing nucleoids. (**a**) Histogram of the natural log of the ratio of light strand methylation to heavy strand methylation for reads with a minimum of 7 methylations in the D-loop. A Gaussian mixture model was applied and a threshold was identified based on a GMM posterior probability of 0.99. Red and green lines indicate each Gaussian fit. The blue and orange lines indicate the posterior probability of each population. A threshold of 3.01 was determined and used to identify reads with D-loop as shown in main Fig. 2h,i and Extended Data Figure 6b, c (**b, c**) Heatmap of the footprint size enrichment at the D-loop region in (b) reads containing a D-loop and in (c) reads lacking a D-loop in HeLa cells. Each row represents a footprint size, and each column shows a position in the genome. Presence of a D-loop was calculated using a GMM with a threshold of 3.01 from the log distribution of the ratio of Light Strand and Heavy Strand methylation.

https://doi.org/10.1038/s41594-024-01225-6



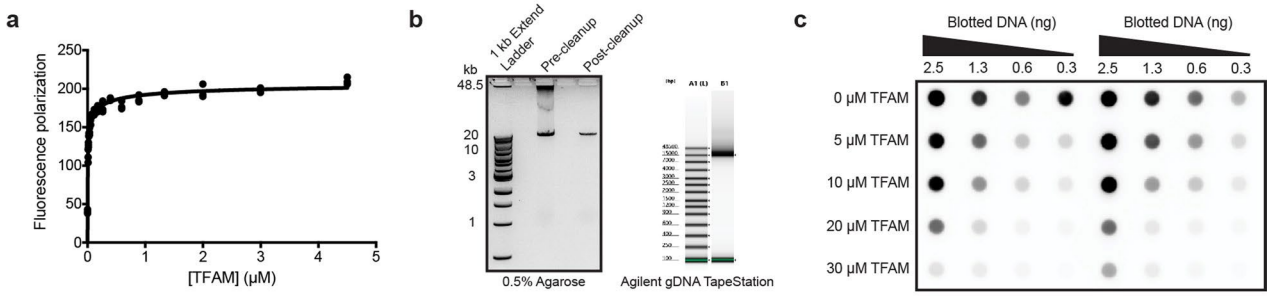
Extended Data Fig. 7 | Changes in OXPHOS levels and TFAM:mtDNA ratio during myoblast differentiation. (**a**) Volcano plot showing differential expression analysis of human skeletal muscle myoblasts in differentiation media for 3 days compared to 0 days (left) and 6 days compared to 0 days (right). Red dotted lines are shown for a p_{adj} value of 0.01 and a fold-change value of 1.5. OXPHOS genes are shown in yellow and key nucleoid-associated proteins are labeled in light blue. P-values were obtained using the Wald test and corrected for multiple testing using the Benjamini and Hochberg method. (**b**) Western blot of nuclear-encoded (NDUFS1) and mitochondrial-encoded (ATP6) OXPHOS subunits from human skeletal muscle myoblasts in differentiation media for 0, 3, and 6 days. GAPDH was used as a loading control. Shown are two biological replicates. (**c**) mtFiber-seq methylation strand bias at the NCR from three biological replicates of human skeletal muscle myoblasts after 0, 3, and 6 days in differentiation media. Methylation bias is calculated as the number of methylations on the light and heavy strands, averaged over a 150 nt sliding window and normalized against the region's AT content. Each window was required to have at least 2,250 methylations across all reads combined. (**d**) Western blot of TFAM levels from human skeletal muscle myoblasts in differentiation media for 0, 3, and 6 days. GAPDH was used as a loading control. Shown are three biological replicates. (**e**) Quantification of relative mtDNA levels by qPCR from human skeletal muscle myoblasts in differentiation media for 0, 3, and 6 days. Shown are mtDNA levels relative to day 0 from three biological replicates. Separate replicates are indicated by circle, triangle, and square shapes. Error bars represent s.d. of the mean. (**f**) Quantification of TFAM levels by western blot. TFAM bands were quantified and normalized against GAPDH. Shown are the TFAM levels relative to day 0. Samples were compared with a one-sided Student's t-test. Exact p-value 0.031 for 6d vs 0d. Shown are three biological replicates. Separate replicates are indicated by circle, triangle, and square shapes. Error bars represent s.d. of the mean.

https://doi.org/10.1038/s41594-024-01225-6



Extended Data Fig. 8 | Increased TFAM levels decrease the accessible population without changing mtDNA levels in HeLa cells. (**a**) Proteinase K protection assay to validate TFAM-HA localization to mitochondria. TFAM-HA appears as a third upper band. Representative blot of two independently repeated experiments (**b, c**) Pearson correlation coefficients from three mtFiber-seq replicates from HeLa S3 TetOn-TFAM-HA cells treated with (b) DMSO or (c) doxycycline comparing (purple) fraction of reads methylated at each adenine and (grey) fraction of reads with a footprint at each position. PacBio chemistry version is indicated for each. Pearson's correlation coefficients are shown. (**d**) Relative mtDNA levels from HeLa S3 TetOn-TFAM-HA cells treated with DMSO or doxycycline. N = 3 biological replicates. (**e**) UpSet plot showing the co-occurrence of footprints at TAS, CSBI, and MTERF1 from HeLa S3 TetOn-TFAM-HA cells treated with DMSO or doxycycline. Maximum footprint sizes based on the footprint size distributions: 60 bp (TAS), 140 bp (CSBI), and 35 bp (MTERF1). A maximum footprint size was set at 170 bp. Reads with larger footprints at these loci were omitted. N = 3 biological replicates. The categories represent 0.59%, 0.98%, and 1.67% of the total molecule population for DMSO replicates 1-3 and 0.44%, 0.54% and 1.31% of the total molecule population for doxycycline replicates 1-3. (**f**) UpSet plot showing the co-occurrence of footprints at the TAS and CSBI with containing a D-loop in HeLa S3 TetOn-TFAM-HA cells treated with DMSO or doxycycline. Maximum footprint sizes: 60 bp (TAS) and 140 bp (CSBI). N = 3 biological replicates. The categories represent 1.23%, 1.85%, and 3.37% of the total molecule population for DMSO replicates 1-3 and 1.00%, 1.14%, and 2.61% of the total molecule population for doxycycline replicates 1-3. (**g**) Log_2 fold change in the fraction of reads containing a D-loop from total and accessible nucleoids (> 1% m6A) in HeLa S3 TetOn-TFAM-HA cells treated with doxycycline relative to DMSO. Data were subsampled to match methylation distributions. Samples were compared with two-sided chi-squared tests for each replicate: *P < 0.05. Exact p-values were 0.011, $5.321*10^{-6}$, and $2.037*10^{-7}$ for doxycycline vs DMSO for total reads, and 0.026, $9.487*10^{-6}$, and $5.823*10^{-9}$ for doxycycline vs DMSO for accessible reads. N = 3 biological replicates for each condition. For all plots, error bars represent s.d. of the mean.

https://doi.org/10.1038/s41594-024-01225-6



d

Pearson correlation coefficient (r)

Fraction of reads methylated per position (Rep 2)

	0 μM	5 μM	10 μM	20 μM	30 μM
30 μM	0.510	0.827	0.888	0.989	0.999
20 μM	0.540	0.874	0.933	1.000	0.987
10 μM	0.677	0.990	0.998	0.917	0.857
5 μM	0.862	0.937	0.866	0.732	0.679
0 μM	1.000	0.735	0.658	0.543	0.507

Fraction of reads methylated per position (Rep 1)

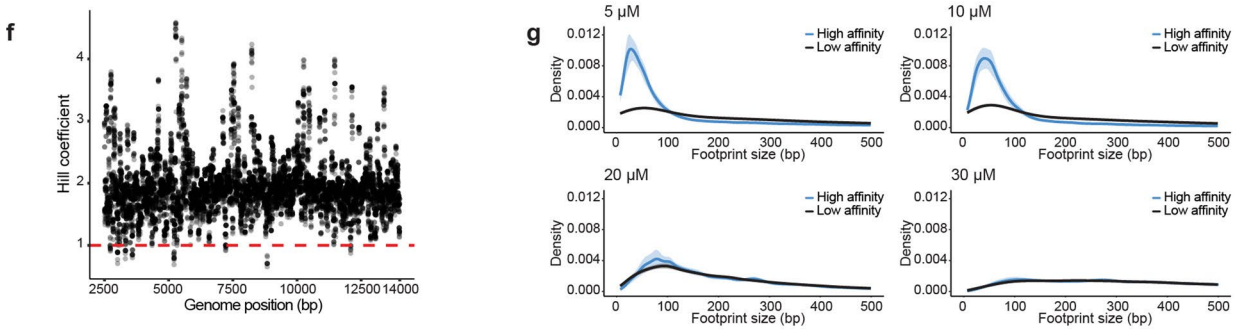
e

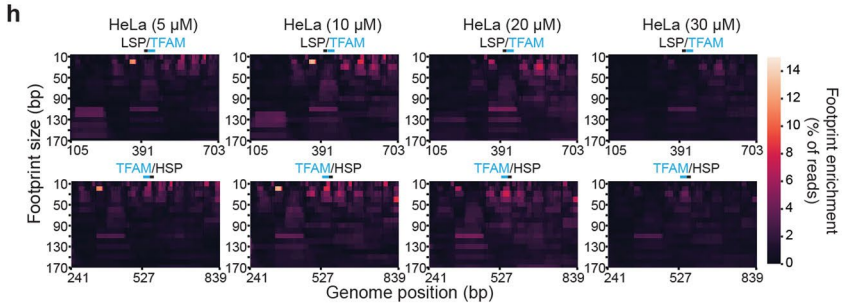
Pearson correlation coefficient (r)

Fraction of reads with footprint per position (Rep 2)

	0 μM	5 μM	10 μM	20 μM	30 μM
30 μM	0.067	0.694	0.769	0.977	0.998
20 μM	0.093	0.782	0.856	1.000	0.967
10 μM	0.301	0.993	0.996	0.822	0.705
5 μM	0.568	0.903	0.822	0.566	0.469
0 μM	0.987	0.353	0.268	0.097	0.065

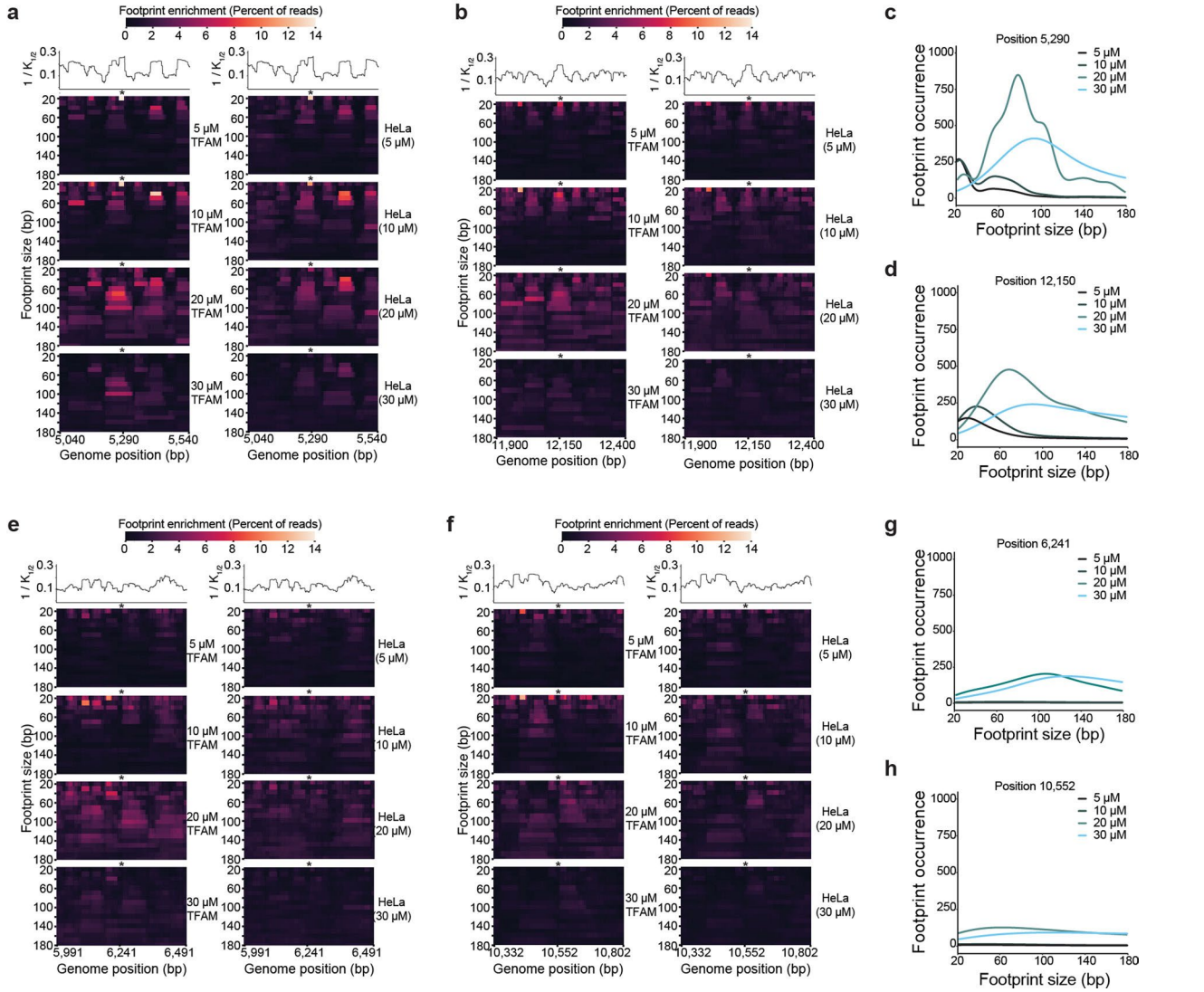
Fraction of reads with footprint per position (Rep 1)





Extended Data Fig. 9 | *In vitro* reconstituted nucleoids reveal TFAM protection patterns. (**a**) TFAM binding to a 28mer corresponding to the HSP binding site measured by fluorescence polarization. The K_D was determined to be 6.2 nM. Results shown are the mean with s.d. from three replicates. (**b**) (Left) 0.5% agarose gel showing mtDNA LR-PCR product before and after column cleanup. (Right) Genomic DNA ScreenTape analysis showing LR-PCR product. Representative gel of five independently repeated experiments (**c**) Two replicates of a dot blot assessing methylation of mtDNA with increasing concentrations of TFAM. A dilution series of DNA amounts were adsorbed onto a nitrocellulose membrane and detected with an anti-m6A antibody. (**d, e**) Tables of Pearson correlation coefficients between replicates and TFAM concentrations for (d) the fraction of reads methylated at each adenine and (e) the fraction of reads with a footprint at each genomic position from two replicates with each TFAM concentration. (**f**) Hill coefficients calculated from genomic position 2,500 to 14,000 from a four parameter logistics regression based on the fraction of reads protected with a footprint at least 20 bp long. Results are from the average of two replicates. (**g**) Meta density plots of footprint sizes from (top) 26 high affinity sites and (bottom) 64 low affinity sites at each TFAM concentration. The center line represents the kernel density estimate with the 95% confidence intervals shaded. Distributions between high and low affinity sites are significantly different at 5 μM TFAM (two-sample KS test, $P < 2.2*10^{-16}$, $D = 0.32$) and 10 μM TFAM (two-sample KS test, $P < 2.2*10^{-16}$, $D = 0.29$). (**h**) Heatmaps of footprint size enrichment at the light strand promoter (LSP) (top) and heavy strand promoter (HSP) (bottom) from HeLa cells subsampled to each TFAM concentration. Each heatmap row represents a footprint size, and each column shows a position in the genome.

https://doi.org/10.1038/s41594-024-01225-6



Extended Data Fig. 10 | TFAM nucleates from preferred binding sites *in vitro* and in cells. (**a, b**) Heatmaps of the footprint size enrichment from two high affinity sites for each concentration of TFAM and from HeLa cells subsampled to each concentration: (a) from position 5,040 to 5,540 and (b) from position 11,900 to 12,400. Each heatmap row represents a footprint size, and each column shows a position in the genome. Line plots indicating the $1/K_{1/2}$ across these loci are shown above the heatmaps. (**c, d**) Density plots showing the footprint size distribution between 20 and 180 bp at (c) position 5,290 and (d) position 12,150 as a function of TFAM concentration. (**e, f**) Heatmaps of the footprint size enrich from two low affinity sites for each concentration of TFAM and from HeLa cells subsampled to each concentration: (a) from position 5,991 to 6,491 and (b) from position 10,332 to 10,802. Each heatmap row represents a footprint size, and each column shows a position in the genome. Line plots indicating the $1/K_{1/2}$ across these loci are shown above the heatmaps. (**g, h**) Density plots showing the footprint size distribution between 20 and 180 bp at (g) position 6,241 and (h) position 10,552 as a function of TFAM concentration.

nature portfolio

Corresponding author(s): Andrew B. Stergachis, L. Stirling Churchman

Last updated by author(s): Nov 10, 2023

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Nikon Elements Acquisition Software AR 5.02 for all imaging
Data analysis	ATAC-seq data processed using bowtie2 (v2.2.9), samtools (v.1.3.1), PICARD (v2.8.0) and visualized with deeptools (v3.0.2) and IGV (v2.4.9). Image analysis performed with arivis (v3.6.1) and Fiji (v1.53). mtFiber-seq data processed using Fibertools (v0.1.3), ccs (v6.4.0), pbmm2 (v1.9.0). RNA-seq data processed using cutadapt(v.1.14), star (v2.7.9a), and DESeq2 (v3.17). Custom scripts available at www.github.com/churchmanlab/mtFiberseq

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

asdfRaw and processed sequencing data are available from the Gene Expression Omnibus at accession number GSE212611 (reviewer token: ahctguuedxgvpkh).

PolG and TWINKLE ChIP-seq data provided by X. Zhu. All microscopy data are available on the Open Microscopy Environment (OMERO) at https://omero.hms.harvard.edu/webclient/?show=project-10105. Custom genomes used for alignment based on the Hg38 reference genome (ncbi.nlm.nih.gov/grc/human) are available at www.github.com/churchmanlab/mtFiberseq

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	This is not relevant to the present study; human participants, their data, or biological material were not used.
Reporting on race, ethnicity, or other socially relevant groupings	This is not relevant to the present study; human participants, their data, or biological material were not used.
Population characteristics	This is not relevant to the present study; human participants, their data, or biological material were not used.
Recruitment	This is not relevant to the present study; human participants, their data, or biological material were not used.
Ethics oversight	This is not relevant to the present study; human participants, their data, or biological material were not used.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	For PacBio sequencing of methylation optimization experiments (e.g. varying Hia5 enzyme or reaction time), single biological replicates were sequenced on the PacBio Sequel IIe due to the high cost of this experiments. High coverage was obtained and supported that our reaction conditions are saturating. Determination of ATAC-see labeling time was performed once, and the optimal time determined was used for all subsequent ATAC-see experiments. For other PacBio sequencing experiments, qPCR, binding assays, ATACsee, Nanostring, and western analysis, 2 or more biological replicates were performed per condition, as is common in the field, and as replicates showed reproducible results.
Data exclusions	No data excluded from analyses. Criteria used for filtering reads are described in the Methods section.
Replication	Experiments were performed in at least 2 biological replicates to verify reproducibility. For mtFiber-seq data, Pearson correlations were performed and reported to confirm reproducibility of both per position methylation and footprinting. Statistical testing of comparisons performed and reported throughout. All attempts at replication for reported data were successful.
Randomization	Randomization is not relevant to this study. Experiments were performed in cultured cells in which known perturbations were compared to controls in the same cell type. Perturbations were always performed in parallel with the appropriate control samples
Blinding	No blinding was used. All analyses performed in this study are based on computational software and are not affected by knowledge of sample conditions.

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

m6A (Active Motif 61995,1:500), IgG HRP secondary (Cell Signaling 7074S,1:1000), TFAM (Santa Cruz, sc-376672,1:1000), TFAM (Proteintech (22586-1-AP, 1:1000), TOMM40 (Proteintech 18409-1-AP,1:8000), TOM20 (Santa Cruz, sc-17764,1:500), HSP60 (Cell Signaling, D307, 1:1000), MTCO1 (Abcam 14705, 1:1000), ACTB (Cell Signaling 3700,1:5000), ds/ssDNA (PROGEN 61014, 1:100), Sodium Potassium ATPase (Abcam ab76020, 1:250), HA tag (Cell Signaling 3724, 1:500), anti Rabbit IgG Alexa Fluor 647 (Thermo A-21245, 1:1000), mouse IgG2a Alexa Fluor 488 (Thermo A-21131, 1:1000), mouse IgG2a Alexa Fluor 555 (Thermo A-21137, 1:1000), mouse IgM Alexa Fluor 647 (Abcam ab150123, 1:1000), mouse IgM Alxa Fluor 555 (Thermo A-21426, 1:1000), rabbit IgG Alexa Fluor 546 (Invitrogen A-11010, 1:1000), NDUFS1 (Abcam ab169540,1:2000), ATP6 (Proteintech 55313-1-AP,1:1000), GAPDH (Invitrogen AM4300,1:10000), mouse IgG HRP secondary (Cell Signaling 70765, 1:1000)

Validation

In our study, TFAM primary antibody validated for human targets by knockdown and overexpression. We detect m6A on DNA after treatment with an m6A methyltransferase in vitro. We detect mitochondrial markers including TOMM40, MTCO1, NDUFS1, HSP60, ATP6, and TFAM enriched after mitochondrial isolation.

All antibodies are commercially available. All other antibodies used as indicated according to validation data provided by the manufacturer on their respective websites.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

HeLa S3 (Female) (ATCC CCL-2.2), U2-OS (female) (ATCC HTB-96), HEK293T (female) (ATCC CRL-3216). Myoblasts provided by Dr. Brendan Battersby (Institute of Biotechnology, University of Helsinki)

Authentication

None of the cell lines were authenticated

Mycoplasma contamination

All cell lines tested negative for myoplasma species using the Universal Mycoplasma Detection Kit (ATCC 30-1012K)

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.