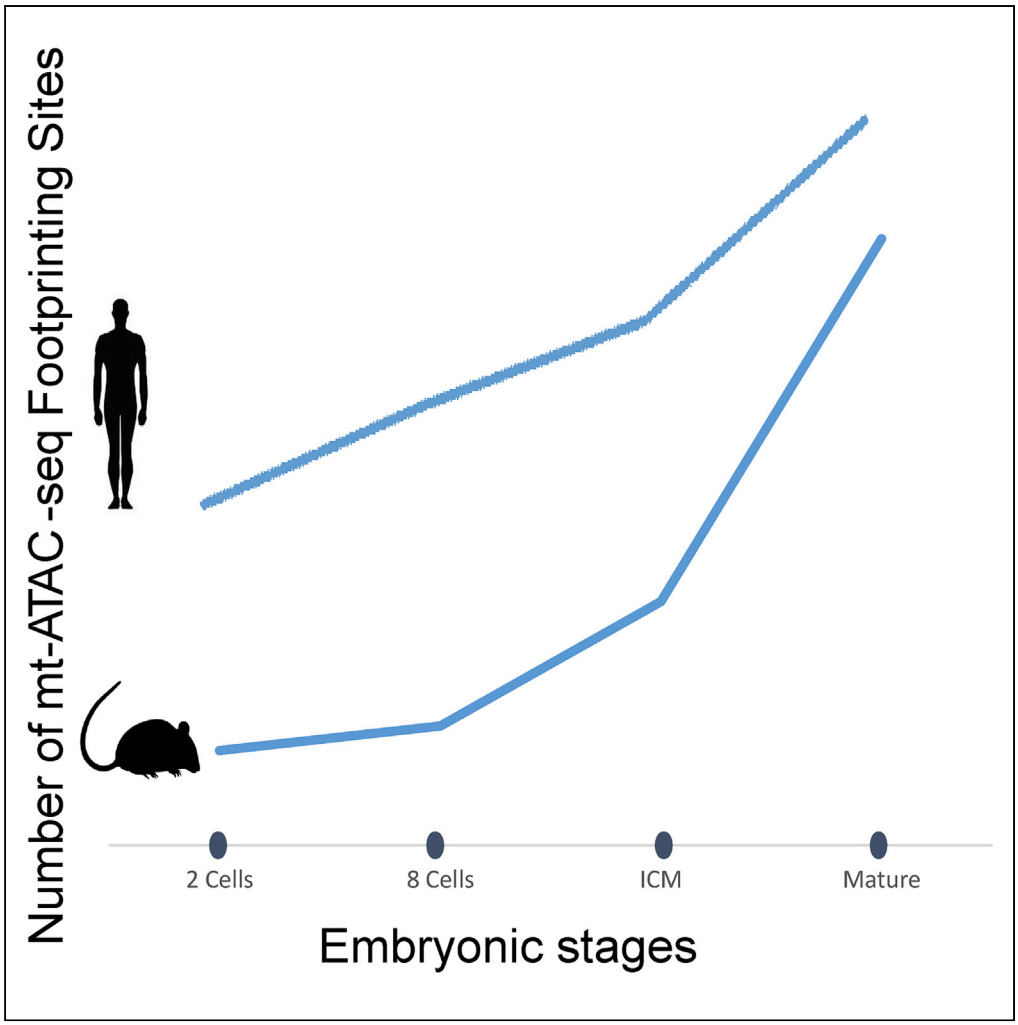
iScience

CellPress

Article

mtDNA Chromatin-like Organization Is Gradually Established during Mammalian Embryogenesis



Shani Marom,
Amit Blumberg,
Anshul Kundaje,
Dan Mishmar

dmishmar@bgu.ac.il

HIGHLIGHTS

Mouse and human mtDNA ATAC-seq footprinting patterns are formed during embryogenesis

mtDNA footprinting sites were either occupied in preimplantation or appeared later

mtDNA footprinting associates with regulatory elements and protein-binding sites

The mtDNA footprinting sites tend to harbor secondary structures

Marom et al., iScience *12*, 141–151
February 22, 2019
https://doi.org/10.1016/j.isci.2018.12.032

iScience

CellPress

Article

mtDNA Chromatin-like Organization Is Gradually Established during Mammalian Embryogenesis

Shani Marom,[1] Amit Blumberg,[1] Anshul Kundaje,[2,3] and Dan Mishmar[1,4,*]

SUMMARY

Unlike the nuclear genome, the mammalian mitochondrial genome (mtDNA) is thought to be coated solely by mitochondrial transcription factor A (TFAM), whose binding sequence preferences are debated. Therefore, higher-order mtDNA organization is considered much less regulated than both the bacterial nucleoid and the nuclear chromatin. However, our recently identified conserved DNase footprinting pattern in human mtDNA, which co-localizes with regulatory elements and responds to physiological conditions, likely reflects a structured higher-order mtDNA organization. We hypothesized that this pattern emerges during embryogenesis. To test this hypothesis, we analyzed assay for transposase-accessible chromatin sequencing (ATAC-seq) results collected during the course of mouse and human early embryogenesis. Our results reveal, for the first time, a gradual and dynamic emergence of the adult mtDNA footprinting pattern during embryogenesis of both mammals. Taken together, our findings suggest that the structured adult chromatin-like mtDNA organization is gradually formed during mammalian embryogenesis.

INTRODUCTION

During metazoan embryogenesis, the transition to zygotic gene expression occurs during the blastocyst stage. This stage precedes the transformation of totipotent cells to detrimental cell fate, which was shown to depend on the metabolic switch from glycolysis to oxidative phosphorylation (OXPHOS) (reviewed by Folmes et al., 2012). As OXPHOS occurs within the mitochondria, the major player in cellular metabolism, it is not surprising that the transition toward differentiating cells is also marked by changes in mitochondrial subcellular distribution (Folmes et al., 2011), mitochondrial morphology, and numbers (Hom et al., 2011), as well as in their regulation of transcription and replication (Wellen and Thompson, 2012). Therefore, as the transition to zygotic gene expression is also accompanied by nuclear genome (nDNA) chromatin remodeling, one may expect it will also affect the higher-order organization of the mitochondrial genome (mtDNA).

Unlike the nDNA, the mtDNA is present in multiple cellular copies that may differ in number (Akiyama and Okada, 1992) and sequence (Avital et al., 2012) between individuals and tissues. Mammalian mtDNA is mainly composed of coding sequences (~93% in humans and mice), including 13 genes for protein subunits of four of the five OXPHOS protein complexes (i.e., OXPHOS complexes I, III–V), two rRNA genes (12S and 16S), and 22 tRNA genes. The non-coding mtDNA regions (the D loop and the light-strand origin of replication [OriL]) harbor most known regulatory elements of mtDNA transcription and replication (Figure 1). During the early 1970s, strand-specific mtDNA polycistron transcripts have been identified in human cells (Aloni and Attardi, 1971), which subsequently were also determined in the mouse mtDNA (Barshad et al., 2018; Battey and Clayton, 1978; Blumberg et al., 2017). This bacterial-like pattern of transcription urged many to isolate the core elements of mtDNA transcriptional regulation, namely, mitochondrial transcription factor A (TFAM) (Fisher et al., 1987; Parisi and Clayton, 1991), the mitochondrial RNA polymerase (POLRMT) (Reid and Parsons, 1971; Ringel et al., 2011), mitochondrial transcription factor B2 (Falkenberg et al., 2002; Gustafsson et al., 2016; Morozov et al., 2014), and the transcription termination factor, mTERF (Daga et al., 1993). It was suggested that one of the key factors in mtDNA transcriptional regulation, TFAM, regulates mtDNA transcription and replication at low cellular concentrations, yet in higher concentrations it serves as the main (and possibly sole) mtDNA-coating protein (Ekstrand et al., 2004; Takamatsu et al., 2002), thus forming the bacterial-like DNA-protein structure, the nucleoid (Iborra et al., 2004; Kukat et al., 2015; Legros et al., 2004). Although it has been suggested that TFAM lacks binding sequence specificity, some researchers proposed mtDNA binding site preferences (Uchida et al., 2017),

[1]Department of Life Sciences, Faculty of Natural Sciences, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel

[2]Department of Genetics, Stanford University, Stanford, CA, USA

[3]Department of Computer Science, Stanford University, Stanford, CA, USA

[4]Lead Contact

*Correspondence: dmishmar@bgu.ac.il

https://doi.org/10.1016/j.isci.2018.12.032

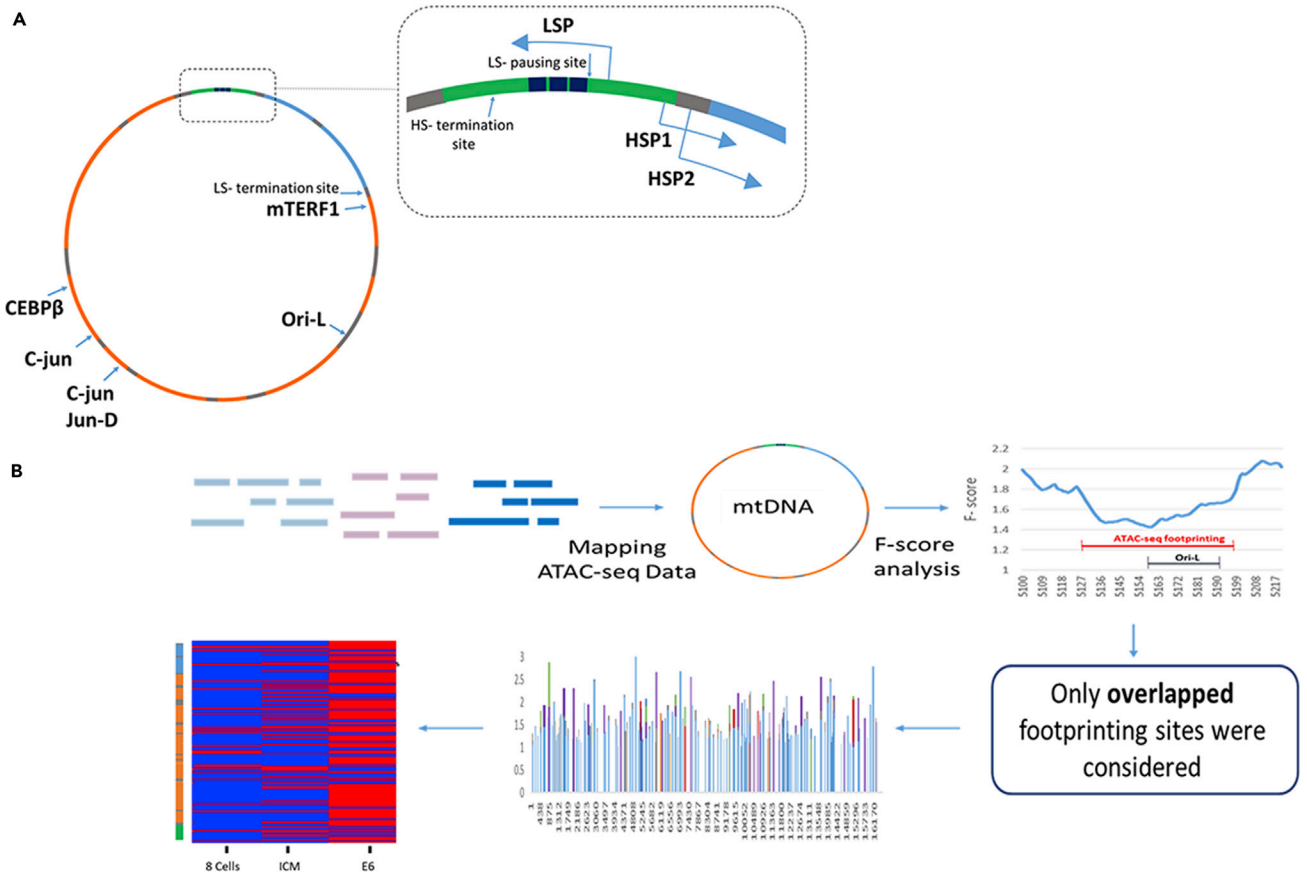


Figure 1. Workflow of mtDNA ATAC-Seq Data Analysis

(A) Human mtDNA map. Light blue panels represent rRNA. Orange panels represent protein-coding genes. Gray panels represent tRNA genes. Green panel represent the D loop. Blue boxes represent the three conserved sequence blocks (CSBs). The dashed box highlights the D loop and known mtDNA promotors.

(B) Workflow describing the identification of mtDNA ATAC-seq footprinting (ASFP) sites. Available ATAC-seq data from mouse embryos were downloaded from the Gene Expression Omnibus (GEO) database (see Transparent Methods). ATAC-seq reads were mapped onto the mouse mtDNA sequence, and read coverage per base was calculated.

especially at regions that tend to adopt G-quadruplex structures (GQP) *in vitro* (Lyonnais et al., 2017) but not *in vivo* (Blumberg et al., 2018). Apart from TFAM, accumulating evidence reveals that known modulators of chromatin structure, such as MOF (Chatterjee et al., 2016), STAT3 (Macias et al., 2014), and SIRT1 (Aquilano et al., 2010), as well as other known regulators of nuclear gene transcription (Blumberg et al., 2014; She et al., 2011) (such as c-Jun, JunD, CEBPB, and MEF2D) are also imported into the mitochondria, bind the mtDNA, and regulate transcription. This raises the possibility that mtDNA higher-order organization is more structured and hence more regulated than once thought. Thus, we hypothesized that during the past ~2 billion years of endosymbiosis, the mitochondria has retained some bacteria-like characteristics of genome packaging, yet additionally has adapted (at least in part) to the host higher-order DNA-protein organization (Barshad et al., 2018). By analyzing DNase-seq experiments from 324 human cell type samples, we recently found a conserved DNase genomic footprinting (DGF) pattern, with 29 mtDNA DGF (mt-DGF) sites common to >90% of the samples analyzed (Blumberg et al., 2018). Our analysis of publicly available TFAM chromatin immunoprecipitation sequencing experiments (ChIP-seq) in HeLa cells (Wang et al., 2013) suggest that these mt-DGFs are poor in TFAM binding, suggesting a more complex and regulated higher-order organization of the mtDNA (Blumberg et al., 2018). Taken together, these findings led us to hypothesize that the human mtDNA is subjected to higher-order protein-DNA organization, which is likely under tight regulation and is physiologically important. Nevertheless, it is unclear when during development such mtDNA organization is formed, and whether it is dynamic and associated with mitochondrial activity.

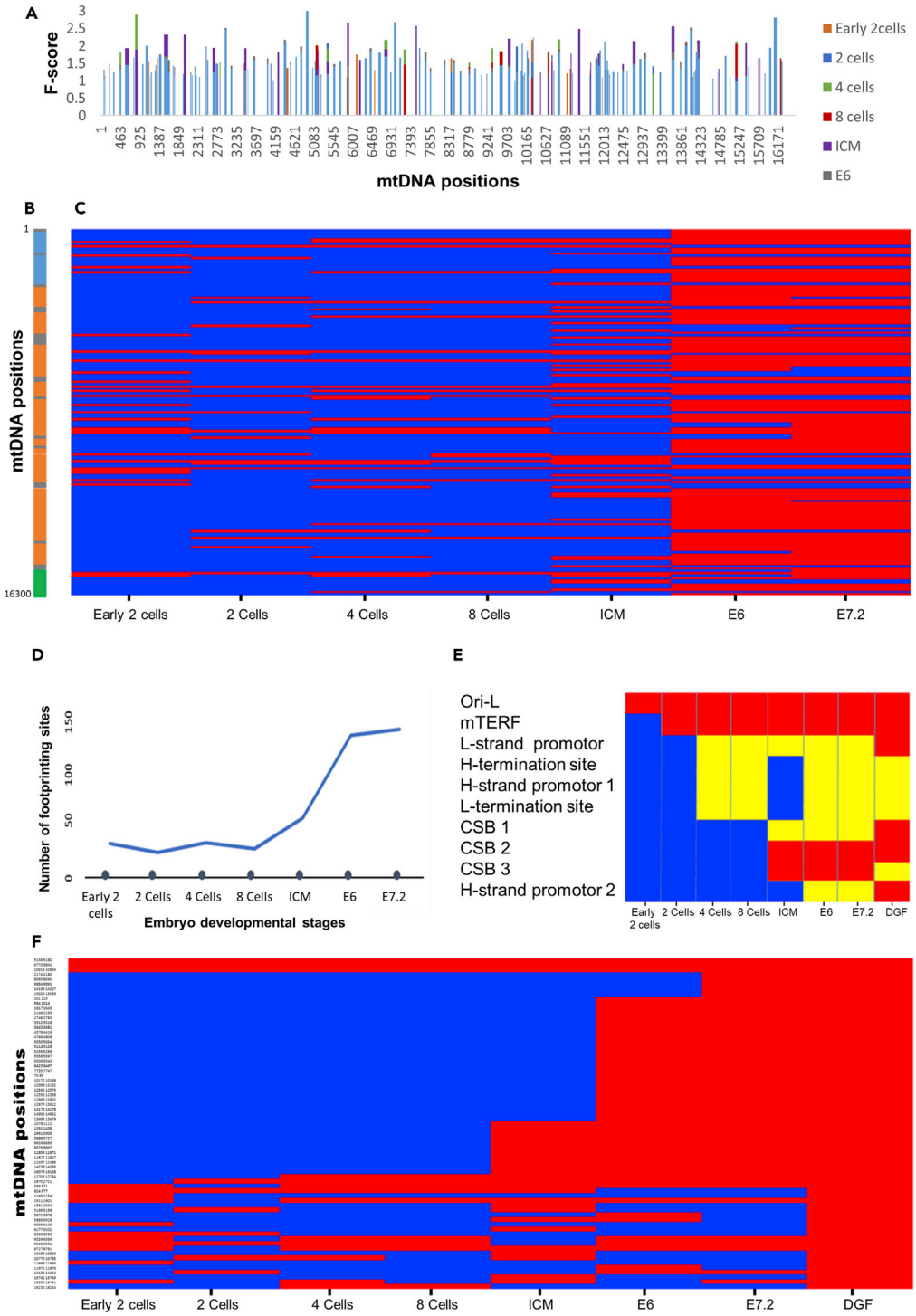


Figure 2. mt-ASFP Site Dynamics during Mouse Embryogenesis

(A) A summary of overall mt-ASFP site distribution across the mouse mtDNA. Bars indicate mt-ASFP sites identified during each of the indicated developmental stages (see color code to the left of the panel). x axis, mtDNA positions.

(B) Linear map of the mouse mtDNA (*Mus musculus*). Blue boxes represent rRNA genes. Orange boxes represent protein-coding genes. Gray boxes represent tRNA genes. Green panel represents the D loop.

(C) Gradual increase in mtDNA-ASFP site distribution during the indicated embryonic developmental stage. Red, occupied sites; blue, unoccupied sites relative to previously analyzed stages. Notice that sites that remained unoccupied during all developmental stages (see A) were excluded. y axis, relative mtDNA position; x axis, developmental stage.

Figure 2. *Continued*

(D) A graph demonstrating mt-ASFP site dynamics during mouse embryogenesis.

(E) Heatmap demonstrating mt-ASFP site association with known mtDNA regulatory elements during mouse embryogenesis. Red, positive association; blue, negative association; yellow, an mt-ASFP site was identified < 40 bp from the indicated mtDNA regulatory site.

(F) Heatmap representing all mt-ASFP sites that co-localize with mt-DGF sites identified in at least 4 out of 43 (>10%) analyzed mouse cell lines. y axis, mtDNA positions; x axis, developmental stage.

Recent analysis of experiments generated using assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) revealed footprinting site pattern, reflecting the dynamics of genome-wide chromatin accessibility during mouse preimplantation embryo development (Wu et al., 2016). This study encouraged us to extend such approach to investigate the accessibility pattern within the mtDNA during mammalian development.

Here, by analyzing ATAC-seq experimental data we found a dynamic mtDNA footprinting pattern during mouse and human embryogenesis. Specifically, we noticed progressive accumulation in the density of mtDNA footprinting sites during the course of mouse and human embryogenesis. Notably, whereas some ATAC-seq footprinting sites (ASFPs) occurred in multiple stages, including those that overlapped known regulatory elements, other sites where stage specific. Taken together our study reveals for the first time a dynamic chromatin-like mtDNA organization during the course of early mammalian embryogenesis, which associates with regulatory sites of mtDNA transcription and replication.

RESULTS

Identifying mtDNA ATAC-Seq Footprinting Patterns during Mouse Embryogenesis

As a first step toward characterizing mtDNA protein-DNA organization during embryogenesis, we analyzed publicly available experimental ATAC-seq data from mouse embryos (Wu et al., 2016). To identify ASFPs in the mtDNA (mt-ASFPs), we slightly modified our previously described approach (Blumberg et al., 2018), also used for earlier analysis of mtDNA footprinting sites (Mercer et al., 2011). In brief, we calculated an F-score for each mtDNA position in sliding windows of variable sizes with a maximum of 124 bases and a minimum of 18 bases (see Transparent Methods for details, and Figure 1). To control for possible Tn5 digestion bias we used a recently published tool (Martins et al., 2018) (also see Transparent Methods and Table S1). In brief, we screened for possible sequence bias in ATAC-seq reads, in a six-nucleotide window size (K-mer), and did not identify any digestion bias. To reduce noise and ensure that reproducible data were being used, we restricted our analysis to those samples for which experimental duplicates were reported. Each of the duplicates was separately analyzed, and only mt-ASFPs shared by the duplicates were considered in subsequent analyses. To increase the stringency of our comparison of mt-ASFP dynamics during mouse development, sites that overlapped at least in one nucleotide position were merged. Such a stringent approach was applied to avoid false identification of mt-ASFP site changes and dynamics, thus reducing potential noise (see Transparent Methods for details). This approach was applied to ATAC-seq experimental data collected from mouse pre-implanted embryos at the following developmental stages: early 2-cell, 2-cell, 4-cell, 8-cell, and inner cell mass (ICM) stages (Wu et al., 2016), as well as from mouse embryonic day 6 (E6) and E7.2 embryos from another dataset (Neijts et al., 2016). Our results revealed increased density of the mt-ASFP site pattern during the course of embryogenesis (Figures 2 and S1, Table S2). The total number of mt-ASFP sites in all the tested mouse embryonic stages was 156, with 28 sites identified in early 2 cells, 21 sites during 2 cells, 24 sites in 4 cells, 29 sites in 8 cells, 49 sites in ICM, 117 sites in E6, and 122 sites in E7.2. Interestingly, in pre-implanted embryos (early 2-cell stage, ICM), the mt-ASFP pattern was relatively scarce, with both gain and loss of sites. Dramatic increases in the density of mt-ASFP sites were observed in both post-implanted stages (E6, E7.2), which also shared many such sites among each other, i.e., many sites were deemed common to both stages (Figures 2 and S1). Closer inspection revealed several types of mt-ASFP sites: first, a core set of mt-ASFP sites was shared by all embryonic stages (3% of mt-ASFP). Second, we identified a group of mt-ASFP sites that appeared only after several cell divisions, yet persisted in all subsequent developmental stages (64% of mt-ASFP). Third, we noticed mt-ASFP sites that alternately appeared during the course of embryogenesis (13% of mt-ASFP), and finally, there was a group of mt-ASFP sites that appeared only during certain pre-implantation stages (20% of mt-ASFP) (Figure S1A). Notably, although the ATAC-seq experiments of pre- and postimplantation mouse embryos were generated by two different laboratories, they used virtually the same experimental protocol. Second, the persistence of certain mt-ASFP sites during all tested

stages attests for the negligible impact of CRISPR-based mtDNA depletion on the analysis. Nevertheless, we cannot exclude some fluctuations that could be resolved only upon analysis of ATAC-seq data generated by a single source.

In addition to changes in mt-ASFP site density, we also observed that the distribution of the mt-ASFP sites throughout the mtDNA varied across the embryonic stages. Although there were no significant differences in the distribution of sites within the protein-coding rRNA and tRNA genes during pre-implantation, the presence of mt-ASFP sites in the D loop was detected only during the 4-cell and 8-cell stages, as well as during post-implantation stages E6 and E7.2 (Figure S1B). This implies that certain regulatory elements become occupied during development after zygote formation. Next, we analyzed mt-ASFP site gain or loss in each of the embryonic mouse stages. Although there was a gradual increase in unique non-redundant sites during mouse development, there was no unique pattern of mt-ASFP site loss (Figure S1C). These results are consistent with the progressive establishment of a distinct mt-ASFP site pattern relatively early during mouse embryogenesis. Notably, it would be of interest to assess mt-ASFP site pattern in the oocyte, which will shed light on the dynamics of the pattern of mtDNA footprinting pre-fertilization.

Recently, we showed that the identification of footprinting sites is comparable between ATAC-seq and DNase-seq experiments (Blumberg et al., 2018). Hence, to validate the results of ATAC-seq data analysis by an alternative approach, we analyzed available DNase-seq data from early mouse embryos (2-cell, 4-cell, and 8-cell stages and the morula) (Lu et al., 2016). Such analysis revealed that all mt-ASFP sites that overlapped with known regulatory elements were also identified as mt-DGF (Table S3 and Table S4, Figure S1D). Notably, similar to the ATAC-seq analysis, a gradual increase in the number of mt-DGF sites was observed during the course of embryogenesis (Figure S1D), thus supporting the robustness of our observation and further attesting to the negligible impact of CRISPR-based mtDNA depletion on our analysis (see also Figure S2 for ATAC-seq read coverage). Finally, we excluded possible digestion bias in both mt-ASFP and mt-DGF data using a recently developed tool (Martins et al., 2018, and see Transparent Methods). Thus, the observed progressive mtDNA occupation during the course of embryogenesis is robust, regardless of the analysis method used.

Nuclear Mitochondrial DNA Fragments Are Not Enriched in Mouse mt-ASFP Sites

It is plausible that the mtDNA ATAC-seq reads used in the current study are contaminated by mtDNA fragments that were transferred into the nucleus during the course of evolution, known as nuclear mitochondrial pseudogenes (nuclear mitochondrial DNA fragments [NUMTs]) (Hazkani-Covo et al., 2003; Mishmar et al., 2004). Notably, as mt-ASFP sites are defined by reduced number of reads at a given site, these might be affected by excess of reads that were mapped to both the nucleus and to active mtDNA. We, therefore, conducted a comprehensive screen to assess the number of NUMT-associated reads per sample per mtDNA position in the ATAC-seq data. To facilitate such a screen, we used a previously published collection of NUMT variants recorded over the entire mouse mtDNA (Calabrese et al., 2012). In general, the proportion of reads harboring NUMT variants in mouse comprised an average of only 0.195% of the reads (SD = 0.48%). Furthermore, the proportion of NUMT reads was not statistically different between mt-ASFP sites and non-ASFP sites across the entire human mtDNA (ASFP sites = 0.18%, SD = 1.57%; non-ASFP sites = 0.263%, SD = 0.54%; Figure S1E). We thus conclude that NUMT reads had only a negligible impact on our ATAC-seq analysis.

Common ATAC-Seq Footprinting Sites Co-localize in Part with Known mtDNA Regulatory Elements

As an initial step toward assessing the functional importance of mt-ASFP sites during mouse embryogenesis, we screened for association between the mt-ASFP sites and mtDNA elements of regulatory importance (Blumberg et al., 2017; Chang and Clayton, 1985) (Figure 2A). First, we noted that the above-mentioned core set of mt-ASFP sites, detected at all embryonic stages, includes the light-strand OriL (Figure 2E). Second, we found that the group of mt-ASFP sites that appeared and persisted after several cell divisions included most known regulatory elements, other than the light-strand promoter (Figure 2). Specifically, the three conserved sequence blocks (CSBs) overlapped mt-ASFP sites that appeared during the ICM stage and persisted in subsequently analyzed stages. In addition, the binding sites for mTERF associated with an mt-ASFP site that appeared as early as during the 2-cell stage, and persisted in all subsequent stages. The mt-ASFP sites that overlapped H-strand promoter 2 (HSP2) appeared only during post-implantation (i.e., the E6 and E7.2 stages). Finally, we identified a group of mt-ASFP sites that displayed alternate

appearance pattern. These included those sites that overlapped the HSP1, light strand (L strand) transcription termination site, and heavy strand (H strand) transcription termination site, which appeared during the 4- to 8-cell stages, were not apparent during the ICM, but re-appeared during post-implantation (Figure 2E). Taken together, the findings show that whereas certain sites that overlap with regulatory sites became occupied even during pre-implantation, others became occupied only later (Figures 2E and 2F).

Secondary Structures Coincide with Common ATAC-Seq Footprinting Sites

It has been suggested that guanine-rich sequences in human mtDNA tend to form GQPs, which potentially affect mtDNA replication initiation and genome stability (Chen et al., 2008; Dong et al., 2014). Recently, by analyzing DNase-seq experiments in 324 human cell types, we found that >90% of the samples shared 29 mt-DGFs, which associated with GQP-forming sequences (Blumberg et al., 2018). This finding encouraged us to assess the association of GQP-forming sequences with the identified mouse mt-ASFP sites. We found that of the 81 sequences with the propensity to adopt GQPs, 25 co-localized with mt-ASFP sites, a value statistically different from random ($p < 0.01$, chi-square test). To further characterize the association of mt-ASFP sites with sequences that potentially form secondary structures, we also considered the association of mt-ASFP sites with tRNA sequences. Such analysis revealed that 21 of the 22 mtDNA-encoded tRNA genes co-localized with mt-ASFP sites, a value that also differed from what would be expected by chance ($p < 0.001$, chi-square test). As secondary structures associate with a variety of regulatory features, the significant co-localization of a subset of mt-ASFP sites with secondary structures we observed here further supports the functional importance of such sites.

Mouse Embryonic mt-ASFP Sites Are Also Occupied in Adult Cells

As mtDNA becomes increasingly coated during the course of mouse embryogenesis, we asked whether any of these apparently occupied sites in the mtDNA persist until adulthood. To address this question, we initially examined publicly available DNase-seq experimental results from 43 different mouse adult cell lines (ENCODE consortium) (see Transparent Methods). This revealed a total of 178 mt-DGF sites, of which more than 65% (N = 122) were shared by at least 10% of the cell lines (Table S5; Blumberg et al., 2018). We thus conclude that whereas certain mt-DGF sites are common to many cell types, others are cell line specific. A comparison of the precise mtDNA positions of these mt-DGF sites (in adult tissues) to the overall set of mt-ASFP sites identified during embryogenesis revealed that 69 of 122 "adult" mt-DGF sites (found in at least 10% of the analyzed adult samples) overlapped with "embryonic" mt-ASFP sites. Notably, 82% of these sites were identified during the post-implantation E6 and E7.2 stages (Figure 2, Table S2), suggesting that at least some of the mtDNA sites that were occupied during embryogenesis remained so until adulthood. Strikingly, the mt-DGF sites in adult cells that overlapped with mt-ASFP sites identified during embryogenesis harbored all known regulatory elements (Figures 2E and 2F). These findings suggest that establishment of the adult mt-ASFP sites pattern begins during differentiation.

mtDNA ATAC-Seq Footprinting Dynamics during Human Embryogenesis

We next asked whether the dynamic mt-ASFP sites pattern seen during mouse embryogenesis also exists in other mammals. To address this question, we analyzed the results of recently published ATAC-seq experiments in human pre-implantation embryos (Wu et al., 2018). ATAC-seq experimental data were collected from human 2-cell- and 8-cell-stage embryos, as well as from the ICM stage, and the analysis was performed using the same protocol as used for mouse embryos mentioned above (see Transparent Methods). The total amount of mt-ASFP sites during all the tested human embryogenesis stages was 143, with 82 sites identified in 2 cells, 109 sites in 8 cells, and 130 sites in ICM. Therefore, our results revealed similar dynamics of the human mt-ASFP sites pattern as observed during mouse mtDNA embryogenesis, namely, increased density of the mt-ASFP landscape (i.e., including both overlapping sites among stages and non-redundant sites) over the course of human embryogenesis (Figures 3 and S3 and Table S6). Second, human mtDNA displayed similar types of mt-ASFP sites as seen with mouse mtDNA (Figure 3C): a core set of mt-ASFP sites that was shared by all embryonic stages (52% of mt-ASFP), a group of mt-ASFP sites that appeared only after several cell divisions, yet persisted in all subsequent developmental stages (10% of mt-ASFP), and mt-ASFP sites that alternately appeared during the course of embryogenesis (38% of mt-ASFP). Third, as in mouse, we utilized a comprehensive list of human NUMTs (Li et al., 2012) and found that the proportion of NUMT reads was not statistically different between mt-ASFP sites and non-ASFP sites across the entire human mtDNA (ASFP sites = 0.16%, SD = 1.32%; non- ASFP sites = 0.195%, SD = 0.46%; Figure S1F). Thus, human NUMT reads had only little impact on our ATAC-seq analysis. Despite the similarities in the dynamics of mt-ASFP sites between mice and humans, it is noteworthy that whereas human mtDNA was

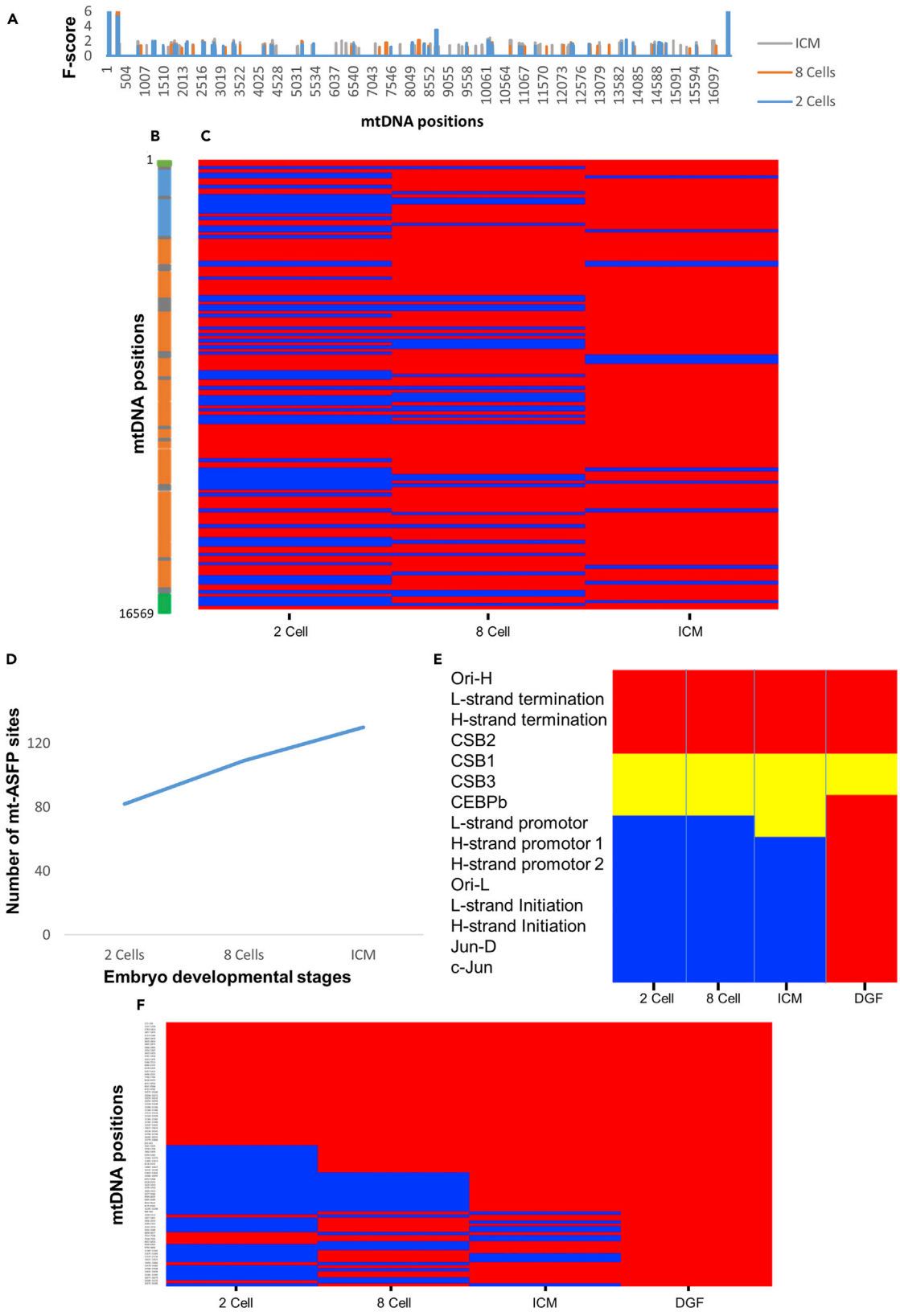

A
F-score
6
4
2
0
1
504
1007
1510
2013
2516
3019
3522
4025
4528
5031
5534
6037
6540
7043
7546
8049
8552
9055
9558
10061
10564
11067
11570
12073
12576
13079
13582
14085
14588
15091
15594
16097
mtDNA positions
ICM
8 Cells
2 Cells
B
C
1
16569
mtDNA positions
2 Cell
8 Cell
ICM
D
Number of mt-ASFP sites
120
80
40
0
2 Cells
8 Cells
ICM
Embryo developmental stages
E
Ori-H
L-strand termination
H-strand termination
CSB2
CSB1
CSB3
CEBPb
L-strand promotor
H-strand promotor 1
H-strand promotor 2
Ori-L
L-strand Initiation
H-strand Initiation
Jun-D
c-Jun
2 Cell
8 Cell
ICM
DGF
F
mtDNA positions
2 Cell
8 Cell
ICM
DGF

Figure 3. : mt-ASFP Site Dynamics during Human Embryogenesis

(A) A summary of overall mt-ASFP site distribution across the human mtDNA. Bars indicate mt-ASFP sites identified during each of the indicated developmental stages (see color code to the left of the panel). x axis, mtDNA positions.

(B) Linear map of the human mtDNA. Blue boxes represent rRNA genes. Orange boxes represent protein-coding genes. Gray boxes represent tRNA genes. Green panel represents the D loop.

(C) Gradual increase in mt-ASFP site distribution during the indicated human embryonic developmental stages. Red, occupied sites; blue, unoccupied sites relative to previously analyzed stages. Notice that sites that remained unoccupied during all developmental stages (see A) were excluded. y axis, relative mtDNA position; x axis, developmental stage.

(D) A graph demonstrating mt-ASFP site dynamics during human embryogenesis.

(E) Heatmap representing the association of mt-ASFP sites with known mtDNA regulatory elements during human embryogenesis. Red, positive association; blue, negative association; yellow, an mt-ASFP site was identified <40 bp from the indicated mtDNA regulatory site.

(F) Heatmap representing all mt-ASFP sites associated with mt-DGF sites in 70 adult human samples (>10% of the samples).

already highly occupied during the pre-implantation stages addressed, a similar degree of mt-ASFP site density was observed only during post-implantation stages in mouse (Figures 2D and 3D). Taken together, mouse and human mt-ASFP patterns showed very similar dynamics, thus suggesting that the establishment of higher-order mtDNA organization during embryogenesis is conserved in mammals.

Next, we asked whether some of the human mt-ASFP sites persisted from the embryonic stage to adulthood. Accordingly, publicly available DNase-seq experimental results from 70 different human cell samples (ENCODE consortium) were consulted (see Transparent Methods) (Blumberg et al., 2018). This analysis revealed a total of 144 mt-DGF sites, which were shared by at least 10% of the samples (Table S7) (Blumberg et al., 2018). While comparing the precise mtDNA positions of mt-DGF sites (in samples from adult tissues) with the overall set of mt-ASFP sites identified during embryogenesis, we found that 88 of 145 adult mt-DGF sites overlapped with "embryonic" mt-ASFP sites. Notably, 88% of these sites could be identified as early as during pre-implantation (Figure 3F and Table S6), suggesting that at least some of the mtDNA sites that were occupied during embryogenesis persisted until adulthood. Similar to what was seen in mouse, our analysis revealed that the human mt-ASFP sites co-localized with known human mtDNA regulatory elements shared by all pre-implantation stages tested (Figure 3E). Among these sites were those overlapping with the heavy-strand origin of replication, heavy- and light-strand transcription termination sites, the three CSBs, and the CEBPβ-binding site. Furthermore, as in mouse, all the human embryonic mt-ASFP sites that were associated with known regulatory elements were such also in the adult (Figure 3E). Hence, in both humans and mice, the mt-ASFP pattern associates with known mtDNA regulatory elements, further supporting the potential regulatory importance of mt-ASFPs.

DISCUSSION

Our analyses of ATAC-seq and DNase-seq experiments from early embryonic stages in both mouse and human samples indicate gradual increase in the number of mtDNA footprinting sites during the course of mammalian embryogenesis. The consistency of the footprinting pattern between two different methods (ATAC-seq and DNase-seq), and between different organisms, attests for the robustness of this observation. Hence, higher-order mtDNA organization is likely more regulated and more dynamic than previously thought, even during embryogenesis. As such organization is already apparent during pre-implantation stages, and as the footprinting sites co-localize with known regulatory elements, it is tempting to suggest that regulatory activation of the mtDNA transpires during early embryogenesis.

The co-localization of our mt-ASFP sites with regulatory elements, particularly with sites that associate with mtDNA transcription, tempts to assess the connection between our observed footprinting dynamics during development and mtDNA transcription. With this in mind the recent discovery of TEAD4, a critical factor in mouse preimplantation embryogenesis, which also regulates mtDNA transcription in the trophectoderm, is noteworthy (Kumar et al., 2018). Once quantitative techniques that measure nascent RNA formation, such as PRO-seq (Kwak et al., 2013), which was recently adapted to mtDNA analysis (Blumberg et al., 2017), are adapted to single-cell analysis, one will be able to experimentally assess such correlation.

The similarity in mt-ASFP dynamics during mouse and human development, and partial similarity in distribution of mt-ASFP sites, raises the possibility that the two species share the underlying mechanism

of mt-ASFP formation. This interpretation is further supported by the tendency of both mouse and human mt-ASFP sites to adopt secondary DNA structures (such as G-quadruplex) and co-localization with regulatory elements. It would be of great interest to assess whether certain regulatory factors tend to bind such sites *in vivo* in the mitochondria of both mouse and man. Nevertheless, as human mtDNA appears to be more occupied already during early embryogenesis than the mouse mtDNA, some mechanistic differences should be considered. Such interpretations should be studied in the future.

In summary, we have provided the first direct evidence for a dynamic chromatin-like organization of mtDNA during mammalian embryogenesis. Specifically, we found that both mouse and human mtDNA display ASFP sites that precisely overlap with known mtDNA regulatory elements. These footprinting sites emerge gradually during embryogenesis and exhibit similar dynamics in both species. Furthermore, in both mammals, we identified mt-ASFP sites that were occupied during all developmental stages, sites that emerged and are maintained in subsequent stages after their appearance, transient sites that emerged but are not retained, and finally, stage-specific sites. Such similarity in the dynamics of mt-ASFP sites, in conjunction with their co-localization with known regulatory sites, suggests the physiological importance of these mt-ASFP sites in mammalian development.

Limitations of the Study

- Although pre-implantation embryonic stages were analyzed in both human and mice, post-implantation stages were analyzed only in mice. These should be analyzed only upon availability of such data.
- The analysis of ATAC-seq data was limited to human and mice embryogenesis; it would be interesting to analyze ATAC-seq data from carefully sampled embryonic stages in non-mammalian vertebrates and invertebrates.
- Some of the analyzed ATAC-seq data originated from Wu et al., who used CARM to deplete mtDNA reads. Despite such approach, we found that multiple mtDNA reads remained and that the gradual increase in mtDNA footprinting sites was observed not only in the ATAC-seq data, but in DNase-seq experiments from the same stages.
- Although ATAC-seq footprinting sites in mouse and human mtDNA associated with known transcriptional and replication regulatory elements, the functional impact of the majority of the identified footprinting sites is yet to be investigated.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, three figures, and seven tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.12.032.

ACKNOWLEDGMENTS

The authors thank Dr. Irene Kaplow and Ms Tal Cohen for technical assistance in the initial steps of analysis. This work was funded by research grants from the Israel Science Foundation (372/17) and the US Army Life Sciences Division (LS67993) awarded to D.M., as well as a US-Israel Binational Science Foundation grant awarded to D.M. and A.K. (2013060). The authors would also like to acknowledge the Negev Scholarship for Excellent Graduate Students awarded to S.M., and the Harbor Foundation for a scholarship for excellent graduate students awarded to A.B.

AUTHOR CONTRIBUTIONS

S.M. performed most of the analysis in the manuscript and contributed to writing the manuscript; A.B. performed the mt-DGF analysis, A.K. contributed software and algorithm development; and D.M. conceived the study and wrote the manuscript.

DECLARATION OF INTEREST

The authors declare no conflict of interest.

Received: August 17, 2018
Revised: November 8, 2018
Accepted: December 27, 2018
Published: February 22, 2019

REFERENCES

Akiyama, T., and Okada, M. (1992). Spatial and developmental-changes in the respiratory activity of mitochondria in early drosophila embryos. Development *115*, 1175–1182.

Aloni, Y., and Attardi, G. (1971). Symmetrical in vivo transcription of mitochondrial DNA in HeLa cells. Proc. Natl. Acad. Sci. U S A *68*, 1757–1761.

Aquilano, K., Vigilanza, P., Baldelli, S., Pagliei, B., Rotilio, G., and Ciriolo, M.R. (2010). Peroxisome proliferator-activated receptor gamma co-activator 1alpha (PGC-1alpha) and sirtuin 1 (SIRT1) reside in mitochondria: possible direct function in mitochondrial biogenesis. J. Biol. Chem. *285*, 21590–21599.

Avital, G., Buchshtav, M., Zhidkov, I., Tuval Feder, J., Dadon, S., Rubin, E., Glass, D., Spector, T.D., and Mishmar, D. (2012). Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. Hum. Mol. Genet. *21*, 4214–4224.

Barshad, G., Marom, S., Cohen, T., and Mishmar, D. (2018). Mitochondrial DNA transcription and its regulation: an evolutionary perspective. Trends Genet. *34*, 682–692.

Battey, J., and Clayton, D.A. (1978). The transcription map of mouse mitochondrial DNA. Cell *14*, 143–156.

Blumberg, A., Danko, C.G., Kundaje, A., and Mishmar, D. (2018). A common pattern of DNase I footprinting throughout the human mtDNA unveils clues for a chromatin-like organization. Genome Res. *28*, 1158–1168.

Blumberg, A., Rice, E.J., Kundaje, A., Danko, C.G., and Mishmar, D. (2017). Initiation of mtDNA transcription is followed by pausing, and diverges across human cell types and during evolution. Genome Res. *27*, 362–373.

Blumberg, A., Sailaja, B.S., Kundaje, A., Levin, L., Dadon, S., Shmorak, S., Shaulian, E., Meshorer, E., and Mishmar, D. (2014). Transcription factors bind negatively selected sites within human mtDNA genes. Genome Biol. Evol. *6*, 2634–2646.

Calabrese, F.M., Simone, D., and Attimonelli, M. (2012). Primates and mouse NumtS in the UCSC genome browser. BMC Bioinformatics *13* (*Suppl 4*), S15.

Chang, D.D., and Clayton, D.A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. U S A *82*, 351–355.

Chatterjee, A., Seyfferth, J., Lucci, J., Gilsbach, R., Preissl, S., Bottinger, L., Martensson, C.U., Panhale, A., Stehle, T., Kretz, O., et al. (2016). MOF acetyl transferase regulates transcription and respiration in mitochondria. Cell *167*, 722–738.e3.

Chen, S.H., Suzuki, C.K., and Wu, S.H. (2008). Thermodynamic characterization of specific interactions between the human Lon protease and G-quartet DNA. Nucleic Acids Res. *36*, 1273–1287.

Daga, A., Micol, V., Hess, D., Aebersold, R., and Attardi, G. (1993). Molecular characterization of the transcription termination factor from human mitochondria. J. Biol. Chem. *268*, 8123–8130.

Dong, D.W., Pereira, F., Barrett, S.P., Kolesar, J.E., Cao, K., Damas, J., Yatsunyk, L.A., Johnson, F.B., and Kaufman, B.A. (2014). Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. BMC Genomics *15*, 677.

Ekstrand, M.I., Falkenberg, M., Rantanen, A., Park, C.B., Gaspari, M., Hultenby, K., Rustin, P., Gustafsson, C.M., and Larsson, N.G. (2004). Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Hum. Mol. Genet. *13*, 935–944.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.G., and Gustafsson, C.M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. Nat. Genet. *31*, 289–294.

Fisher, R.P., Topper, J.N., and Clayton, D.A. (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. Cell *50*, 247–258.

Folmes, C.D., Dzeja, P.P., Nelson, T.J., and Terzic, A. (2012). Metabolic plasticity in stem cell homeostasis and differentiation. Cell Stem Cell *11*, 596–606.

Folmes, C.D.L., Nelson, T.J., Martinez-Fernandez, A., Arrell, D.K., Lindor, J.Z., Dzeja, P.P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. Cell Metab. *14*, 264–271.

Gustafsson, C.M., Falkenberg, M., and Larsson, N.G. (2016). Maintenance and expression of mammalian mitochondrial DNA. Annu. Rev. Biochem. *85*, 133–160.

Hazkani-Covo, E., Sorek, R., and Graur, D. (2003). Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications. J. Mol. Evol. *56*, 169–174.

Hom, J.R., Quintanilla, R.A., Hoffman, D.L., de Mesy Bentley, K.L., Molkentin, J.D., Sheu, S.S., and Porter, G.A., Jr. (2011). The permeability transition pore controls cardiac mitochondrial maturation and myocyte differentiation. Dev. Cell *21*, 469–478.

Iborra, F.J., Kimura, H., and Cook, P.R. (2004). The functional organization of mitochondrial genomes in human cells. BMC Biol. *2*, 9.

Kukat, C., Davies, K.M., Wurm, C.A., Spahr, H., Bonekamp, N.A., Kuhl, I., Joos, F., Polosa, P.L., Park, C.B., Posse, V., et al. (2015). Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. Proc. Natl. Acad. Sci. U S A *112*, 11288–11293.

Kumar, R.P., Ray, S., Home, P., Saha, B., Bhattacharya, B., Wilkins, H.M., Chavan, H., Ganguly, A., Milano-Foster, J., Paul, A., et al. (2018). Regulation of energy metabolism during early mammalian development: TEAD4 controls mitochondrial transcription. Development *145*, https://doi.org/10.1242/dev.162644.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Legros, F., Malka, F., Frachon, P., Lombes, A., and Rojo, M. (2004). Organization and dynamics of human mitochondrial DNA. J. Cell Sci. *117*, 2653–2662.

Li, M., Schroeder, R., Ko, A., and Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. Nucleic Acids Res. *40*, e137.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing chromatin regulatory landscape during mouse preimplantation development. Cell *165*, 1375–1388.

Lyonnais, S., Tarres-Sole, A., Rubio-Cosials, A., Cuppari, A., Brito, R., Jaumot, J., Gargallo, R., Vilaseca, M., Silva, C., Granzhan, A., et al. (2017). The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. Sci. Rep. *7*, 43992.

Macias, E., Rao, D., Carbajal, S., Kiguchi, K., and DiGiovanni, J. (2014). Stat3 binds to mtDNA and regulates mitochondrial gene expression in keratinocytes. J. Invest. Dermatol. *134*, 1971–1980.

Martins, A.L., Walavalkar, N.M., Anderson, W.D., Zang, C., and Guertin, M.J. (2018). Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. Nucleic Acids Res. *46*, e9.

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.M., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., et al. (2011). The human mitochondrial transcriptome. Cell *146*, 645–658.

Mishmar, D., Ruiz-pesini, E., Brandon, M., and Wallace, D.C. (2004). Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into

our African origins and the mechanism of foreign DNA integration. Hum. Mutat. *23*, 125–133.

Morozov, Y.I., Agaronyan, K., Cheung, A.C., Anikin, M., Cramer, P., and Temiakov, D. (2014). A novel intermediate in transcription initiation by human mitochondrial RNA polymerase. Nucleic Acids Res. *42*, 3884–3893.

Neijts, R., Amin, S., van Rooijen, C., Tan, S., Creyghton, M.P., de Laat, W., and Deschamps, J. (2016). Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. Genes Dev. *30*, 1937–1942.

Parisi, M.A., and Clayton, D.A. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. Science *252*, 965–969.

Reid, B.D., and Parsons, P. (1971). Partial purification of mitochondrial RNA polymerase from rat liver. Proc. Natl. Acad. Sci. U S A *68*, 2830–2834.

Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. Nature *478*, 269–273.

She, H., Yang, Q.A., Shepherd, K., Smith, Y., Miller, G., Testa, C., and Mao, Z.X. (2011). Direct regulation of complex I by mitochondrial MEF2D is disrupted in a mouse model of Parkinson disease and in human patients. J. Clin. Invest. *121*, 930–940.

Takamatsu, C., Umeda, S., Ohsato, T., Ohno, T., Abe, Y., Fukuoh, A., Shinagawa, H., Hamasaki, N., and Kang, D. (2002). Regulation of mitochondrial D-loops by transcription factor A and single-stranded DNA-binding protein. EMBO Rep. *3*, 451–456.

Uchida, A., Murugesapillai, D., Kastner, M., Wang, Y., Lodeiro, M.F., Prabhakar, S., Oliver, G.V., Arnold, J.J., Maher, L.J., Williams, M.C., et al. (2017). Unexpected sequences and structures of mtDNA required for efficient transcription from the first heavy-strand promoter. Elife *6*, e27283.

Wang, Y.E., Marinov, G.K., Wold, B.J., and Chan, D.C. (2013). Genome-wide analysis reveals coating of the mitochondrial genome by TFAM. PLoS One *8*, e74513.

Wellen, K.E., and Thompson, C.B. (2012). A two-way street: reciprocal regulation of metabolism and signalling. Nat. Rev. Mol. Cell Biol. *13*, 270–276.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S., et al. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature *557*, 256–260.

ISCI, Volume *12*

Supplemental Information

mtDNA Chromatin-like Organization Is Gradually Established during Mammalian Embryogenesis

Shani Marom, Amit Blumberg, Anshul Kundaje, and Dan Mishmar

Marom et al 2018: Inventory of Supplemental Information

Transparent Methods: Experimental Procedures.

Supplemental Figures

Supplemental Figure S1, related to figures 2-3: Dynamics of mt-ASFP sites and their distribution during mouse embryogenesis. A. Pie chart summarizing the distribution of mt-ASFP site prevalence during mouse embryonic stages. Blue area - ASFP sites shared by all embryonic stages tested (3% of ASFP). Light blue area - ASFP sites that appeared only after several cell divisions, yet persisted in all subsequent developmental stages (64% of ASFP). Gray area - ASFP sites that appeared only during certain pre-implantation stages (20% of ASFP). Orange area - ASFP sites with alternating appearance during embryogenesis (13% of ASFP). B. Graph illustrating the genomic distribution of the ASFP sites across the different mtDNA regions. C. Bar graph demonstrating ASFP site gain/loss, as compared to ASFP site distribution in the preceding embryonic stage. D. A graph demonstrating mt-DGF site dynamics during mouse embryogenesis. E. A diagram comparing the prevalence of NUMT reads at ASFP versus non-ASFP mtDNA sites. F.A diagram comparing the prevalence of human NUMT reads at ASFP versus non-ASFP mtDNA sites.

Supplemental Figure S2, related to figures 2-3: Mouse ASFP site length, read coverage and ATAC-seq site read coverage and F-score data. A. Box plot representing mt-ASFP site length distribution. Average length is indicated. B. A table summarizing the average, median, maximum and minimum mt-ASFP site lengths at each embryonic stage tested. C. Graph representing the read coverage for a representative mouse ATAC-seq experiment. Notice that the reduced read coverage in the end of the linearized mtDNA is presented prior to re-mapping to correct for the circularity of the molecule (applies also to Figure S3). X axis - mtDNA nucleotide positions, Y axis - number of reads per site. D. Distribution of calculated F-scores in a representative mouse sample. X axis - the mtDNA nucleotide positions, Y axis - F-scores per position. E. Graph representing specific F-score data for the Ori-L site (underlined in red).

Supplemental Figure S3, related to figures 2-3: Human mt-ASFP site length, read coverage and ATAC-seq site read coverage and F-score data. A. Box plot representing mt-ASFP site length distribution. Average length is indicated. B. A table summarizing the average, median, maximum and minimum of ASFP site lengths at each embryonic stage tested. C. Graph representing the read coverage for a representative human ATAC-seq experiment. D. Distribution of calculated F-scores in a representative human sample. X axis - the mtDNA nucleotide positions, Y axis - F-scores per position. E. Graph representing specific F-score data for the Ori-L site (underlined in red – similar to Figure S2).

Supplemental Tables

Supplemental Table S1, related to Figure 1 and Figure 3: Calculated chi-square for K-mer analysis. A. Table of chi-square analysis for mouse embryogenic stages. B. Table of chi-square analysis for human embryonic stages.

Supplemental Table S2, related to Figure 2: Calculated F-scores for mouse mt-ASFP sites per mouse embryonic stage. Last column represents mt-DGFs that co-localized with mt-ASFP sites.

Supplemental Table S3, related to Figure 2: Calculated F-scores for mouse mt-DGF sites per embryonic stage.

Supplemental Table S4, related to Figure 2: Mouse mt-DGF sites analysis as a validation for the results of mouse ATAC-seq experiments. mt-DGF co-localize with mt-ASFP sites at mouse mtDNA regulatory elements during early mouse embryogenesis. Plus/minus signs indicate whether our identified mt-ASFP or mt-DGF sites co-localize with known mtDNA regulatory elements. Asterisk - an mt-ASFP site located no more than 40 bp from the indicated regulatory elements.

Supplemental Table S5, related to Figure 2: mt-DGF sites in mouse adult cell lines. mt-DGF sites that were present in more than 10% of the analyzed cell lines (N=43) (according to Blumberg et al. 2018, **Genome Research**).

Supplemental Table S6, related to Figure 3: Calculated F-scores for mt-ASFP sites per mouse embryonic stage. The last column represents adult mt-DGFs that co-localize with mt-ASFP sites.

Supplemental Table S7, related to Figure 3: mt-DGF sites in human adult cell lines. mt-DGF sites that were present in more than 10% of the analyzed cell lines (N=70) (according to Blumberg et al. 2018, **Genome Research**).

Experimental Procedures

Samples used for data analysis: ATAC-seq data for mouse embryogenesis were obtained from GEO Accession number GSE66582 (Neijts et al., 2016; Wu et al., 2016). DNase-seq data for mouse embryogenesis were obtained from GEO Accession number GSE76642 (Lu et al., 2016). While considering mouse preimplantation and post-implantation stages, we analyzed a total of 28 experiments, with each developmental stage represented by two biological replicates each performed in duplicates (i.e. 4 experiments per developmental stage). ATAC-seq data for human embryogenesis were obtained from GEO Accession number GSE101571 (Wu et al., 2018). While considering human samples, we analyzed a total of 12 experiments, with each developmental stage represented by two biological replicates (two independent samples), each performed in duplicates (i.e. 4 experiments per developmental stage). Notably, both ATAC-seq and DNase-seq data were obtained from whole cells, not isolated nuclei.

Sample-specific mtDNA sequence reconstruction and mapping, read coverage calculation and circular-like mapping of sample-specific mtDNA sequences: Analyses were performed as described previously (Blumberg et al., 2017). In brief, after mapping reads against the mtDNA reference genome (NCBI accession numbers NC_005089.1 (mouse) and NC_012920.1 (human)), sample-specific mtDNA reference genomes were re-constructed and ATAC-seq and DNase-seq reads were aligned while taking into account the circular organization of the mtDNA, as recently performed (Blumberg et al., 2017). Read coverage for each position was calculated using the 'genomecov' command in BEDtools (http://bedtools.readthedocs.org/en/latest/ version

2.25)(Quinlan and Hall, 2010) (for mouse, Figure S2; for human, Figure S3). Only positions covered by at least 300X read depth were taken into consideration. For ATAC-seq analysis, average, median, maximum and minimum coverage are presented (for mouse, Figure S2; for human, Figure S3).

ATAC-seq analysis: Publically available Sequence Reads Archive (SRA) files were converted into a FASTq format using sratoolkit (www.ncbi.nlm.nih.gov/Traces/sra/?view=toolkit_doc). The sequencing adaptors were trimmed by applying an ATAC-seq analysis pipeline, specifically using 'function detect adaptors' and 'function trim adaptors' (http://github.com/kundajelab/atac_dnase_pipelines). mt-ASFP sites were identified using a similar set of criteria as recently described for the identification of DNase footprinting sites (Blumberg et al., 2017). In brief, for each mtDNA nucleotide position, an F-score was calculated in sliding read windows of ~120 bp using the following equation: $F = (C + 1)/L + (C + 1)/R$, where C represents the average number of reads in the central fragment, L represents the average read count in the proximal fragment, and R represents the average read count in the distal fragment. The lowest F-scores were interpreted as reflecting ASFP sites. To identify overlapping mt-ASFP sites in experimental replicates, we employed the BEDtools command 'intersect' using the '-a' and '-b' options for each of the different replicates. Only mt-ASFP sites that were common to all replicates of a given experiment (i.e. overlapping by at least 1 nucleotide position) were used for subsequent analysis. In addition, in cases where triplicates were available, we first recorded overlapping sites in two of them using BEDtools, followed by assessment of overlapping sites with the third replicate. For analysis of mt-ASFP sites dynamics,

we considered mt-ASFP sites that overlap in all embryonic stages analyzed. To this end, mt-ASFP data from all developmental stages were combined in all analyzed samples. The length of each combined mt-ASFP site was measured between the 5' and 3' nucleotide positions of the most proximal and distal overlapping mt-ASFP sites, respectively.

Statistical comparison between BED files: Statistical comparison between sets of sites represented in BED format was performed using BEDtools (Quinlan and Hall 2010).

DNase-seq analysis: ENCODE DNase-seq fastq files were downloaded from the ENCODE consortium website: http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeUwDnase/). mt-DGF sites were identified by following a method similar to the approach described in the above sub-section 'ATAC-seq analysis'(Blumberg et al., 2018).

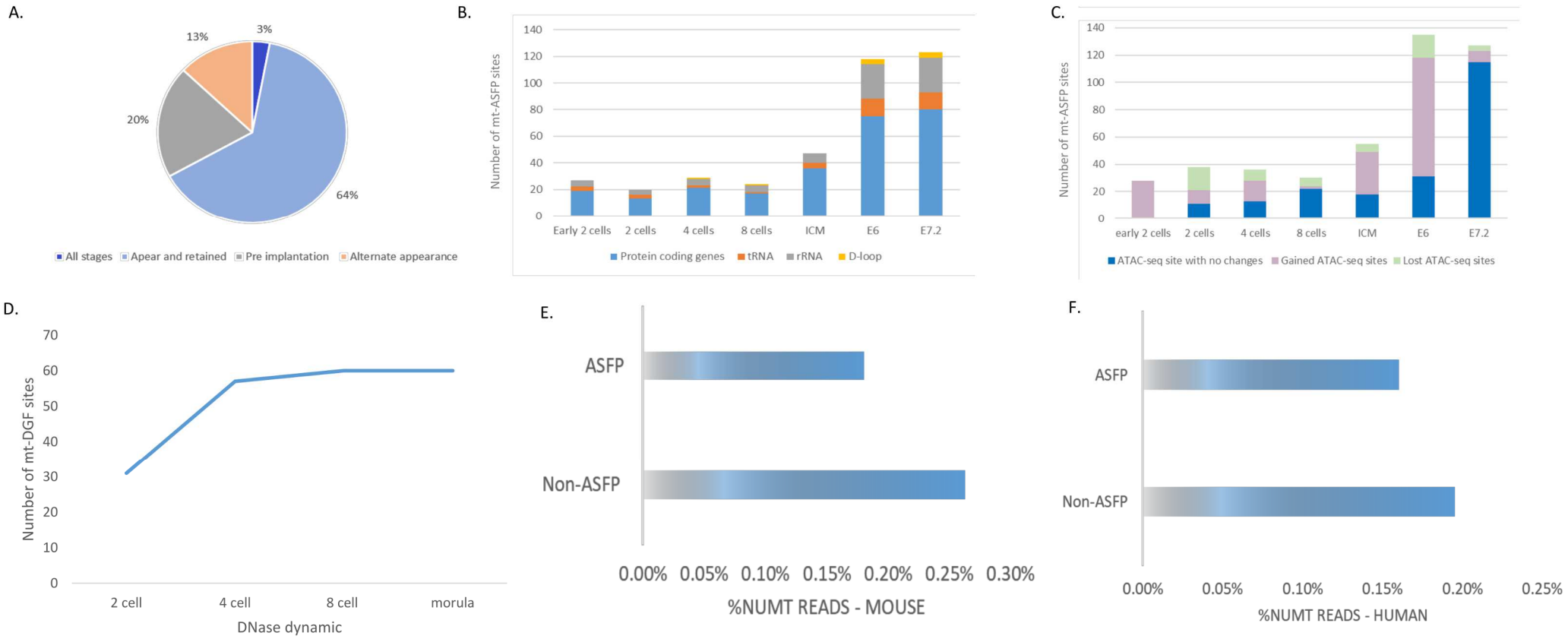
K-Mer analysis: To control for possible Tn5 digestion bias we screened for possible sequence bias in ATAC-seq and DNase-seq reads, in 6 nucleotide window size (K-mer). K-mer scores were calculated using the seqOutBias pipeline (https://github.com/guertinlab/seqOutBias) as previously reported (Martins et al., 2018). In brief, seqOutBias was applied to BAM files available from all tested human and mouse ATAC-seq data files, and k-mer values were calculated for each mtDNA position per sample. Next, we used these k-mer values (per mtDNA position) to calculated average and standard deviation. Then, we tested whether the distribution of k-mer values in mt-ASFP sites recorded for all embryonic stages deviated from

the mean values (+SD) calculated for the entire mtDNA per sample; by that we referred only to the mt-ASFP sites that obtained across all the tested embryonic stages). In order to do so, we asked whether the K-mer calculated values in the mtDNA ASFP sites significantly deviated by at least two standard deviations from the mean K-mer values calculated for randomly selected mtDNA sites (chi square test). Such comparison was performed for each of the tested samples.

NUMT analysis: The proportion of NUMT reads was estimated using bam-readcount (https://github.com/genome/bam-readcount) by counting mtDNA-mapped ATAC-seq reads (within BAM files) that contained NUMT variants. NUMT genetic variants was obtained from the published collection of 150 mouse NUMT variants (Calabrese et al., 2012). For each analyzed sample, a sample-specific NUMT collection was generated by screening the reconstructed sample-specific mtDNA sequences. The proportion of sequencing reads harboring a specific NUMT mutation was estimated per mtDNA position.

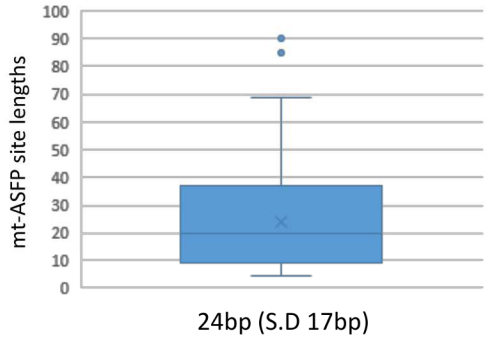
Prediction of the potential of DNA sequences to adopt a G-quadruplex structure: G-quadruplex DNA structures were predicted for both the light and heavy mtDNA strands using QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php)(Kikin et al., 2008). We used the previously published prediction parameters (Dong et al., 2014) (GQP max length = 33. Min; G-group size = 2; loop size = 0 to 36). mtDNA site coordinates were listed after merging the predicted site coordinates for both mtDNA strands. Statistical assessment of mt-ASFP site enrichment within sequences with the propensity to adopt secondary structure (GQP and mtDNA-encoded tRNAs, separately) was compared to their association with a set of random

non-ASFP mtDNA sites of comparable sequence length to the average mt-ASFP site (Chi square test for goodness of fit).



Supplemental Figure S1, related to Figures 2-3: Dynamics of mt-ASFP sites and their distribution during mouse embryogenesis. A. Pie chart summarizing the distribution of mt-ASFP site prevalence during mouse embryonic stages. Blue area - ASFP sites shared by all embryonic stages tested (3% of ASFP). Light blue area - ASFP sites that appeared only after several cell divisions, yet persisted in all subsequent developmental stages (64% of ASFP). Gray area - ASFP sites that appeared only during certain pre-implantation stages (20% of ASFP). Orange area - ASFP sites with alternating appearance during embryogenesis (13% of ASFP). B. Graph illustrating the genomic distribution of the ASFP sites across the different mtDNA regions. C. Bar graph demonstrating ASFP site gain/loss, as compared to ASFP site distribution in the preceding embryonic stage. D. A graph demonstrating mt-DGF site dynamics during mouse embryogenesis. E. A diagram comparing the prevalence of NUMT reads at ASFP versus non-ASFP mtDNA sites. F.A diagram comparing the prevalence of human NUMT reads at ASFP versus non-ASFP mtDNA sites..

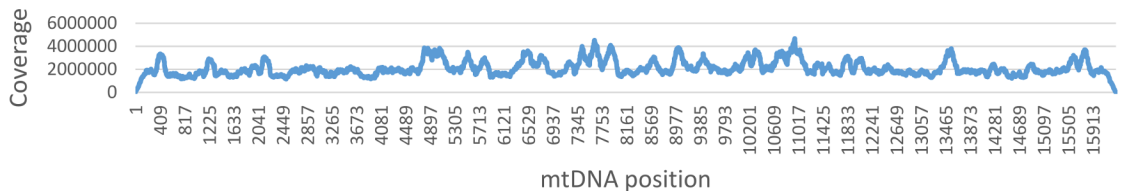
A.



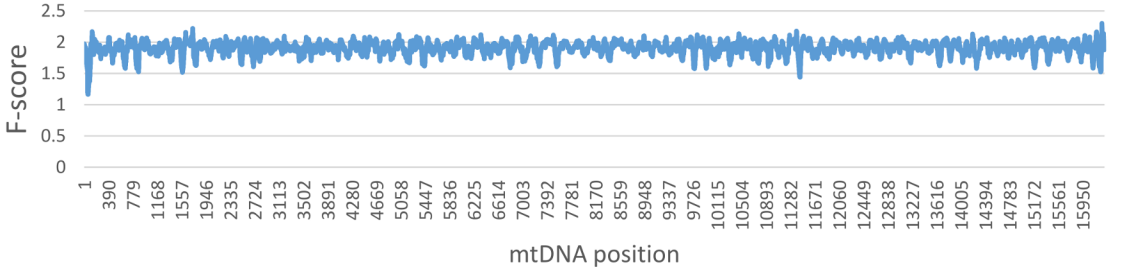
B.

	Early 2 cells	2 cells	4 cells	8 cells	ICM	E6	E7
Average coverage	1509348	725630	383893	78946	291629	62273	68869
Median	1237767	688781	386278	79224	293156	62508	69149
Max coverage	6156885	1588738	526590	110832	384044	87389	97552
Min coverage	8576	2923	336	471	1717	466	529

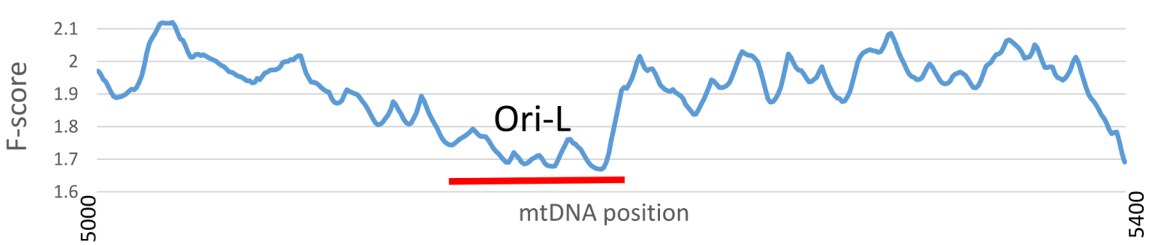
C.



D.

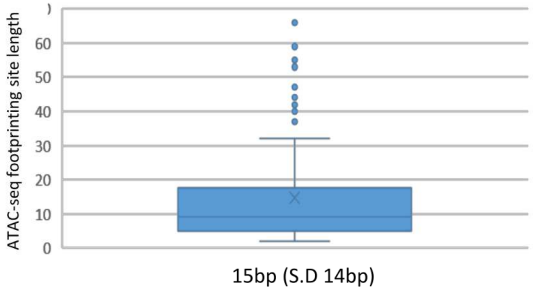


E.



Supplemental Figure S2, related to Figures 2-3: Mouse ASFP site length, read coverage and ATAC-seq site read coverage and F-score data. A. Box plot representing mt-ASFP site length distribution. Average length is indicated. B. A table summarizing the average, median, maximum and minimum mt-ASFP site lengths at each embryonic stage tested. C. Graph representing the read coverage for a representative mouse ATAC-seq experiment. Notice that the reduced read coverage in the end of the linearized mtDNA is presented prior to re-mapping to correct for the circularity of the molecule (applies also to Figure S3). X axis - mtDNA nucleotide positions, Y axis - number of reads per site. D. Distribution of calculated F-scores in a representative mouse sample. X axis - the mtDNA nucleotide positions, Y axis - F-scores per position. E. Graph representing specific F-score data for the Ori-L site (underlined in red).

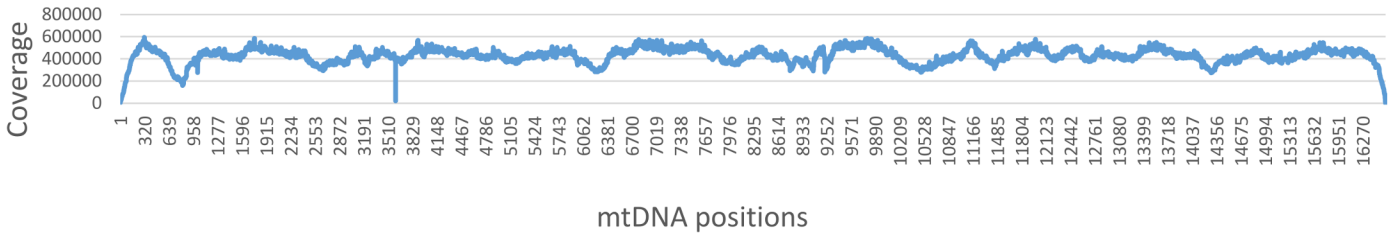
A



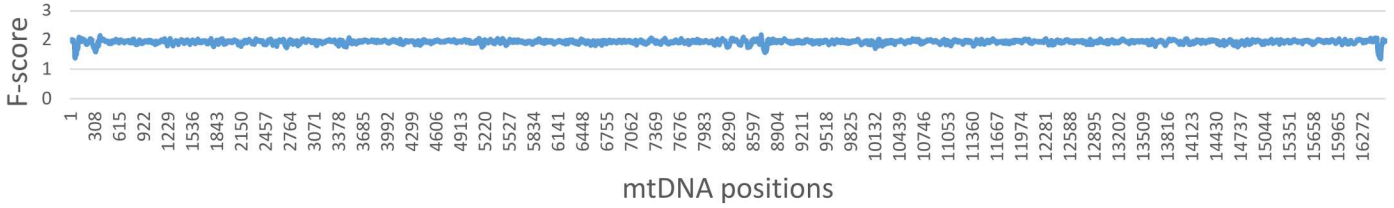
B.

	2 cells	8 cells	ICM
Average coverage	1393380	460614	332088
Median	1400438	466304	336495
Max coverage	1953866	659452	46683
Min coverage	1462	1104	811

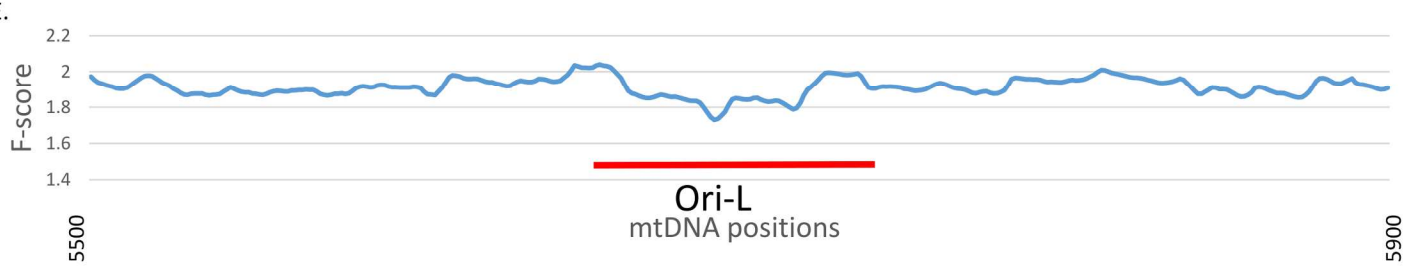
C.



D.



E.



Supplemental Figure S3, related to Figure 2-3: Human mt-ASFP site length, read coverage and ATAC-seq site read coverage and F-score data. A. Box plot representing mt-ASFP site length distribution. Average length is indicated. B. A table summarizing the average, median, maximum and minimum of ASFP site lengths at each embryonic stage tested. C. Graph representing the read coverage for a representative human ATAC-seq experiment. D. Distribution of calculated F-scores in a representative human sample. X axis - the mtDNA nucleotide positions, Y axis - F-scores per position. E. Graph representing specific F-score data for the Ori-L site (underlined in red – similar to Figure S2).

A.	Mouse embryonic stage	Kmer within the ASFP sites (Observed)	Kmer within random mtDNA sites (maintain site sample size)	Total Kmer calculated
	Early 2 2 cells rep1	142	199	828
	Early 2 2 cells rep2	123	205	798
	2 cells rep 1	141	193	750
	2 cells rep 2	136	211	787
	4 cells rep 1	121	188	805
	4 cells rep 2	139	197	792
	8 cells rep 1	144	183	840
	8 cells rep 2	145	201	801
	ICM rep 1	132	198	838
	ICM rep 2	146	185	852
	E6 rep 1	135	189	720
	E6 rep 2	127	202	766
	E7.2 rep 1	119	196	785
	E7.2 rep 2	117	193	793

B.	Human embryonic stage	Kmer within the ASFP sites (Observed)	Kmer within random mtDNA sites (maintain site sample size)	Total Kmer calculated
	2 cells rep 1	72	735	1590
	2 cells rep 2	69	712	1483
	8 cells rep 1	62	698	1464
	8 cells rep 2	75	501	1435
	ICM rep 1	58	771	1500
	ICM rep 2	78	706	1401

Supplemental Table S1, related to Figure 1 and Figure 3: Calculated chi-square for K-mer analysis. A. Table of chi-square analysis for mouse embryogenic stages. B. Table of chi-square analysis for human embryonic stages.

Regulatory Element - DNAse	Position	2 cell	4 cell	8 cell	Morula
mTERF1	2660-2681	+	+	+	+
L-strand origin of replication	5160-5191	*	*	*	+
CSB 1	16035-16058	+	+	+	+
CSB 2	16089-16104	+	+	+	+
CSB 3	16114-16131	+	+	+	+
L-strand promotor	16183	*	*	*	*
L-strand pausing site	16159	*	*	*	*
H-strand termination site	16247	+	+	+	+
H-strand promotor 2	65	+	+	+	+
H-strand promotor 1	16282-16290	*	*	*	*
L-strand termination site	16284	*	*	*	*
H-strand pausing site	74	+	+	+	+

Regulatory Element - ASFP	Position	2 cells	4 cells	8 cells	ICM
mTERF1	2660-2681	+	+	+	+
L-strand origin of replication	5160-5191	+	+	+	+
CSB 1	16035-16058	-	-	-	*
CSB 2	16089-16104	-	-	-	+
CSB 3	16114-16131	-	-	-	+
L-strand promotor	16183	-	*	*	*
L-strand pausing site	16159	-	-	-	-
H-strand termination site	16247	-	*	*	-
H-strand promotor 2	65	-	-	-	-
H-strand promotor 1	16282-16290	-	*	*	-
L-strand termination site	16284	-	*	*	-
H-strand pausing site	74	-	-	-	-

Supplemental Table S4, related to Figure 2: Mouse mt-DGF sites analysis as a validation for the results of mouse ATAC-seq experiments. mt-DGF co-localize with mt-ASFP sites at mouse mtDNA regulatory elements during early mouse embryogenesis. Plus/minus signs indicate whether our identified mt-ASFP or mt-DGF sites co-localize with known mtDNA regulatory elements. Asterisk - an mt-ASFP site located no more than 40 bp from the indicated regulatory elements.

ATAC-seq sites		F-Score				ATAC-seq sites		F-Score			
Start	End	2 Cells	8 Cells	ICM	DGF>90%	Start	End	2 Cells	8 Cells	ICM	DGF>90%
51	105	5.64	5.38	1.13		9350	9355	1.12	0.00	1.50	
203	218	1.23	0.00	0.00		9372	9381	0.00	1.23	1.47	
274	345	3.40	6.10	5.68		9436	9442	1.11	1.03	1.25	
353	364	1.86	1.08	1.63		9535	9545	0.00	0.00	1.40	
820	852	1.35	1.34	1.93		9552	9584	1.37	1.22	1.74	
889	894	0.00	1.25	1.23		9626	9630	1.19	0.00	1.18	
909	927	0.00	1.41	0.00		9797	9801	0.00	0.00	1.32	
1069	1076	1.33	1.16	1.47		9963	9969	1.48	1.17	1.50	
1221	1230	1.69	1.82	2.07		9976	9991	1.03	0.00	1.14	
1268	1279	0.00	1.25	1.41		10050	10054	1.25	0.00	0.00	+
1294	1301	1.82	1.73	1.94		10071	10082	1.43	0.00	0.00	+
1361	1368	1.25	0.00	0.00		10112	10155	1.84	1.93	2.26	+
1510	1514	1.31	1.05	1.25		10175	10181	2.14	1.86	2.46	
1522	1526	0.00	1.03	1.28		10200	10216	1.31	1.41	1.40	
1640	1644	0.00	0.00	1.20		10230	10243	1.77	1.81	1.83	
1705	1711	0.00	0.00	1.50		10254	10260	1.36	1.10	1.42	
1721	1733	1.04	1.52	1.83		10386	10391	1.36	0.00	1.25	
1740	1751	1.05	1.22	1.23		10709	10751	1.12	1.01	2.12	
1759	1766	0.00	1.43	1.77		10946	10978	1.33	1.35	1.74	
1794	1814	1.75	1.69	2.08	+	11029	11040	1.46	1.27	1.69	
1828	1842	1.14	1.11	1.65	+	11066	11080	1.42	1.23	1.43	+
1858	1871	1.26	1.44	1.65		11090	11107	1.72	1.25	1.30	+
1937	1941	1.28	0.00	1.45		11244	11255	1.06	0.00	1.09	+
1961	1971	1.11	1.10	1.27		11369	11387	1.96	1.26	1.79	
2007	2011	0.00	1.09	0.00		11397	11403	1.36	1.11	1.32	
2114	2161	1.94	1.65	1.97		11465	11469	1.20	1.20	1.47	
2290	2315	1.10	1.43	1.51		11510	11519	1.30	1.13	1.59	
2438	2474	1.33	0.00	0.00		11676	11700	0.00	1.29	0.00	
2547	2584	1.15	1.83	1.75		11806	11814	0.00	1.18	1.21	
2598	2613	1.67	1.80	1.86		11824	11832	1.05	0.00	1.42	
2703	2751	2.17	0.00	1.26	+	12086	12098	0.00	0.00	1.48	
2804	2825	1.03	1.39	1.04		12125	12135	0.00	1.44	0.00	
2836	2844	1.07	1.21	1.93		12241	12249	0.00	0.00	1.24	
2866	2875	1.87	1.78	2.28		12263	12274	0.00	1.02	1.52	
2887	2896	1.22	1.32	1.51		12321	12330	1.52	1.83	1.72	
3097	3115	1.81	1.28	1.28		12344	12349	0.00	1.09	1.56	
3144	3154	0.00	1.17	0.00		12367	12384	1.39	1.05	1.63	
3163	3167	0.00	1.16	0.00		12392	12400	1.33	1.26	1.37	
3351	3388	1.84	1.90	2.13		12797	12819	1.32	1.61	1.57	
3404	3424	1.26	1.17	1.51		12835	12846	1.37	1.51	1.58	
3445	3492	1.85	0.00	0.00		12875	12883	1.23	0.00	1.05	
3541	3551	1.02	1.45	1.84		13022	13026	0.00	1.31	0.00	
3557	3569	0.00	1.16	1.56		13057	13061	1.05	0.00	1.45	
4148	4152	1.02	1.24	1.54		13081	13095	1.05	1.05	1.35	+

4175	4184	1.18	1.20	1.21	+	13338	13343	1.14	1.43	1.52	
4192	4219	1.06	1.26	1.68	+	13582	13586	1.54	1.15	1.48	
4333	4377	1.31	1.50	1.63		13612	13617	1.26	1.36	1.76	
4401	4430	1.43	1.10	1.13		13643	13647	0.00	0.00	1.62	
4440	4444	0.00	0.00	1.29		13657	13688	2.02	1.69	2.05	
4510	4514	0.00	0.00	1.41		13778	13809	1.63	1.82	1.14	
4544	4548	1.04	1.10	1.10		13969	13979	1.33	1.67	1.86	
5056	5071	0.00	0.00	1.29	+	13988	14029	1.11	1.23	1.28	
5078	5087	0.00	0.00	1.31	+	14093	14111	1.32	0.00	0.00	
5168	5213	1.59	1.31	1.89		14121	14130	1.14	0.00	0.00	
5460	5466	0.00	1.29	1.46		14449	14455	1.03	0.00	1.62	
5484	5514	1.61	1.56	1.59	+	14479	14484	1.59	1.04	1.44	+
5822	5826	1.32	0.00	0.00		14493	14498	1.50	1.44	1.45	+
6087	6103	1.13	1.27	1.73		14535	14546	1.78	1.44	1.91	+
6240	6245	1.23	1.35	2.03	+	14561	14570	1.25	0.00	1.04	
6353	6369	0.00	0.00	1.73	+	14634	14682	1.61	1.71	1.86	
6404	6409	1.34	0.00	1.08	+	14701	14750	2.01	1.57	1.27	+
6497	6502	1.03	1.30	1.59		14787	14799	1.50	0.00	0.00	+
6539	6550	1.01	0.00	1.20		14832	14840	0.00	1.43	1.71	
6557	6563	1.73	1.55	1.81		14851	14863	1.24	1.05	0.00	
6575	6579	1.04	0.00	1.52		15012	15026	1.07	0.00	1.34	
6609	6635	1.19	0.00	1.26		15041	15046	1.28	1.41	2.04	
6679	6684	1.12	0.00	0.00		15231	15242	1.01	0.00	1.46	
6691	6703	1.61	0.00	1.07		15392	15396	0.00	1.07	1.23	
7132	7139	1.57	1.13	1.66		15463	15467	0.00	1.23	0.00	
7237	7263	1.06	1.10	1.14		15476	15480	0.00	1.02	0.00	
7408	7452	1.29	1.76	0.00	+	15647	15654	1.33	1.31	1.83	
7534	7539	1.22	1.18	0.00		15667	15683	1.02	1.16	1.04	+
7549	7556	1.11	1.15	0.00		16021	16025	1.47	0.00	0.00	+
7791	7795	1.33	1.11	1.55		16075	16080	0.00	0.00	2.06	+
8101	8126	1.49	1.54	1.68		16090	16127	1.18	0.00	2.06	+
8137	8143	0.00	1.40	1.68		16142	16147	0.00	1.01	1.48	+
8175	8183	1.48	0.00	1.42		16164	16186	1.03	1.35	0.00	+
8263	8322	2.32	2.14	1.63	+	16466	16528	6.23	5.85	3.77	
8413	8454	1.55	1.65	1.55							
8529	8581	2.33	1.11	1.11							
8590	8603	0.00	0.00	1.32							
8682	8687	1.29	0.00	1.30							
8723	8783	4.39	3.25	1.89							
8813	8819	1.00	0.00	1.26							
9108	9113	0.00	1.00	1.37	+						

Supplemental Table S6, related to Figure 3: Calculated F-scores for mt-ASFP sites per human embryonic stage. The last column represents adult mt-DGFs that co-localize with mt-ASFP sites.

Human mt-DGF sites >10%			Human mt-DGF sites >10%		
Start	**End**	**% of tested samples**	**Start**	**End**	**% of tested samples**
195	335	88.406%	8201	8347	94.203%
400	438	10.145%	8440	8529	89.855%
484	598	91.304%	8546	8607	11.594%
628	687	91.304%	8695	8767	88.406%
832	877	88.406%	8883	8901	36.232%
922	987	76.812%	8923	8976	52.174%
1133	1189	40.580%	9063	9119	98.551%
1229	1253	23.188%	9248	9293	20.290%
1307	1369	65.217%	9414	9509	89.855%
1408	1458	69.565%	9579	9631	50.725%
1475	1556	86.957%	9663	9734	100.000%
1624	1654	73.913%	9741	9764	28.986%
1744	1748	10.145%	9823	9861	17.391%
1785	1857	95.652%	10048	10121	95.652%
1945	2028	78.261%	10166	10252	98.551%
2059	2064	13.043%	10294	10341	98.551%
2088	2155	46.377%	10424	10508	94.203%
2190	2271	17.391%	10609	10694	11.594%
2295	2376	39.130%	10845	10905	88.406%
2404	2416	47.826%	10948	10968	86.957%
2445	2496	89.855%	11046	11121	97.101%
2600	2619	73.913%	11182	11217	40.580%
2660	2674	14.493%	11345	11393	92.754%
2690	2769	95.652%	11432	11465	63.768%
2805	2885	71.014%	11508	11538	40.580%
3089	3202	37.681%	11550	11571	15.942%
3254	3315	73.913%	11581	11603	26.087%
3338	3359	10.145%	11647	11726	72.464%
3416	3450	10.145%	11800	11847	73.913%
3489	3537	53.623%	11917	11966	36.232%
3603	3621	11.594%	11983	11987	24.638%
3694	3768	78.261%	12059	12105	79.710%
3817	3830	28.986%	12144	12196	73.913%
3880	3935	15.942%	12236	12296	60.870%
3980	4032	89.855%	12337	12387	85.507%
4088	4119	86.957%	12431	12449	18.841%
4165	4207	98.551%	12573	12607	89.855%
4259	4323	97.101%	12672	12718	100.000%
4427	4476	23.188%	12781	12795	52.174%
4555	4608	68.116%	12843	12882	71.014%
4633	4686	24.638%	12939	12956	47.826%
4731	4791	85.507%	13012	13038	50.725%
4885	4928	91.304%	13066	13115	94.203%
5049	5106	98.551%	13123	13136	20.290%
5180	5285	79.710%	13155	13199	79.710%

5375	5467	72.464%	13207	13216	18.841%
5480	5577	92.754%	13229	13274	88.406%
5633	5665	66.667%	13337	13346	20.290%
5707	5790	91.304%	13392	13450	98.551%
5832	5921	34.783%	13545	13583	60.870%
5938	5982	68.116%	13701	13734	14.493%
6001	6037	76.812%	13780	13793	75.362%
6099	6135	82.609%	13855	13938	78.261%
6182	6257	95.652%	13944	13966	68.116%
6355	6454	95.652%	14009	14028	62.319%
6480	6502	15.942%	14048	14114	81.159%
6568	6624	18.841%	14181	14213	34.783%
6717	6737	17.391%	14267	14331	82.609%
6782	6800	15.942%	14479	14548	98.551%
6904	6915	13.043%	14586	14618	65.217%
7029	7067	66.667%	14643	14683	33.333%
7148	7179	50.725%	14743	14815	97.101%
7270	7364	94.203%	14987	15001	14.493%
7365	7457	95.652%	15156	15189	47.826%
7498	7570	82.609%	15262	15300	24.638%
7610	7665	26.087%	15398	15467	59.420%
7738	7799	28.986%	15673	15726	94.203%
7827	7836	10.145%	15803	15849	47.826%
7888	7937	55.072%	15899	15917	85.507%
8031	8076	91.304%	15950	16027	94.203%
8109	8112	13.043%	16055	16234	92.754%
8125	8146	14.493%	16274	16340	62.319%

Supplemental Table S7, related to Figure 3: mt-DGF sites in human adult cell lines. mt-DGF sites that were present in more than 10% of the analyzed cell lines (N=70) (according to Blumberg et al. 2018, **Genome Research**).