Regulatory principles of human mitochondrial gene expression revealed by kinetic analysis of the RNA life cycle

Erik McShane[1], Mary Couvillion[1], Robert Ietswaart[1], Gyan Prakash[1], Brendan M. Smalec[1], Iliana Soto[1], Autum R. Baxter-Koenigs[1], Karine Choquet[1], L. Stirling Churchman[1,*]

[1]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA

*Corresponding author:

L. Stirling Churchman

Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115

E-mail: churchman@genetics.med.harvard.edu

Abstract

Mitochondria play a critical role in cellular metabolism primarily through hosting the oxidative phosphorylation (OXPHOS) machinery that is encoded by mitochondrial DNA (mtDNA) and nuclear DNA, with each genome separately regulated in their respective compartments. To unravel how the two gene expression systems collaborate to produce the OXPHOS complexes, the regulatory principles controlling the production of mtDNA-encoded proteins need to be elucidated. Here, we performed a quantitative analysis of the mitochondrial messenger RNA (mt-mRNA) life cycle to gain insight into which steps of gene expression experience the most regulatory control. Our analysis revealed a unique balance between the rapid turnover and high accumulation of mt-mRNA, leading to a 700-fold higher transcriptional output than nuclear-encoded OXPHOS genes. Additionally, we observed that mt-mRNA processing and its association with the mitochondrial ribosome occur rapidly and that these processes are linked mechanistically. These data resulted in a model of mtDNA expression that is predictive across human cell lines, revealing that differential turnover and translation efficiencies are the major contributors to mitochondrial-encoded protein synthesis. Applying this framework to a disease model of Leigh syndrome, French–Canadian type, we found that disrupting the responsible nuclear-encoded gene, *LRPPRC*, perturbs OXPHOS biogenesis predominantly through altering mt-mRNA stability. Our findings provide a comprehensive view of the intricate regulatory mechanisms governing mtDNA-encoded protein synthesis, highlighting the importance of quantitatively analyzing the mitochondrial RNA life cycle for decoding the regulatory principles of mtDNA expression.

Introduction

Determining how DNA and RNA connect to protein abundance is a major goal of genome science. The main principles underlying these connections have been quantitatively investigated for the bacterial and eukaryotic nuclear genomes[1,2]. In these systems, transcription initiation plays a major role with other layers of gene regulation serving to augment the large amount of control that occurs at gene promoters. However, the animal cell maintains and expresses two genomes, the nuclear chromosomes and the small, circular, and extremely polyploid mitochondrial DNA (mtDNA). Mitochondrial gene expression is spatially separated from nuclear gene expression processes and is governed by evolutionary distinct transcription, RNA processing, and translation machinery[3]. The expression of nuclear and mitochondrial DNA need to be coordinated as they encode the dual-origin oxidative phosphorylation (OXPHOS) complexes responsible for cellular respiration and ATP production[3–5]. In human cells, the relative abundance of mtDNA-encoded RNAs and proteins varies across cell types and conditions in a manner that is matched with their nuclear-encoded OXPHOS counterparts[3,5,6]. When protein synthesis is deregulated between the two compartments proteostatic stress follows[5,7] and mutations in genes involved in mitochondrial gene expression cause severe human disease syndromes[8].

Nuclear-encoded genes are transcribed into monocistronic transcripts that are spliced, polyadenylated, and exported for translation in the cytoplasm. In contrast, mitochondrial genes are transcribed into two long polycistronic primary transcripts encoded on each strand of the mtDNA, the heavy and light strands (Figure 1A). The primary transcripts are cleaved into individual mt-mRNA, transfer RNA (mt-tRNA), and ribosomal RNA (mt-rRNA). After processing, mt-mRNA is typically adorned with short poly(A) tails before being translated by the mitochondrial ribosomes (mitoribosomes) at the inner mitochondrial membrane[9]. Although the steps of mitochondrial gene expression are well defined, the quantitive and kinetic rules governing their expressions have not been fully resolved. For example, unlike for nuclear or bacterial chromosomes, regulation of

transcription initiation does not contribute much to the relative levels of mitochondrial-encoded proteins as mt-RNAs are produced from only three promoters[10]. Thus, the steady-state mitochondrial transcriptome must be established by post-transcriptional processes. However, the timing of these processes in the mitochondrial RNA life cycle, such as processing, ribosome association, and turnover, remain obscure largely due to a lack of quantitative analysis of mitochondrial gene expression and post-transcriptional processes. To understand how dual-origin OXPHOS complexes are co-regulated, we first need to determine the relative contributions of core regulatory modules to mitochondrial gene expression. The recent development of tools such as nucleotide-conversion-based strategies to measure RNA turnover (Fig 1B)[11,12] and direct RNA-sequencing to measure RNA abundance and processing (Fig 1G) enables the quantitative and kinetic analysis of the mt-RNA lifecycle with sufficient resolution to answer these questions for the first time[13].

In this study, we leveraged these tools to quantitatively measure the lives of mitochondrial RNAs by determining the rates of each major life-cycle step. We found that mt-mRNA has a 700-fold higher production rate than nuclear-encoded OXPHOS transcripts and that differential turnover explains the steady-state mt-mRNA abundance with little regulation by transcriptional mechanisms. Mt-mRNA processing, which includes cleavage out of the primary transcript and poly(A) tailing, generally occurs rapidly (<1 min). We found that mitoribosome association occurs on the same time scale as processing and most mt-mRNAs are mitoribosome-bound. Interestingly, our data indicate that in certain cases, mt-mRNA 5 prime-end processing is rate-limiting for ribosome association. Combining all our kinetic measurements into one quantitative gene expression model revealed that mitochondrial protein synthesis can be explained by mt-mRNA turnover rates (~60% of variability), translation efficiency (~40%), and ribosome-association rates (<1%). Intriguingly, we found that mitochondrial translational efficiencies are largely constant across the cell lines we analyzed. Finally, we used our model to disentangle the role of the disease-associated Leucine Rich Pentatricopeptide Repeat Containing (LRPPRC) protein in gene expression control and we found that LRPPRC's key

impact is in the regulation of mt-RNA stability. Our model of mitochondrial gene expression control is the first systematic, quantitative, and comparative kinetic analysis of the mitochondrial transcriptome, shedding light on how its expression can be controlled by and coordinated with nuclear-encoded genes.

Results

Differential mt-mRNA degradation rates determine the heavy-strand transcriptome

The mtDNA heavy strand encodes for two mitochondrial rRNAs, *RNR1* and *RNR2*, and 11 of the 12 mt-mRNAs (Fig 1A). These are produced in a single primary transcript and processed into individual RNAs. Although made at equimolar proportions, heavy-strand RNAs accumulate at levels that vary on the order of 100-fold[6,14]. RNA stability has been proposed to be a leading contributor to the differential steady-state levels of the individual transcripts[15,16]. However, testing this hypothesis has been challenging due to a lack of methods to quantify RNA turnover that does not rely on strong perturbations, such as transcription inhibition that can lead to compensatory changes to RNA turnover rates[14,15]. Recent advances in the detection of modified bases by next-generation sequencing have led to the development of nucleotide-conversion-based strategies to measure RNA turnover (Fig 1B)[11,12]. We sought to leverage this strategy to produce quantitative measurements of mitochondrial RNA and adapted TimeLapse-seq (TL-seq)[12] to be suitable for the measurement of mitochondrial RNA (mito-TL-seq).

In TL-seq, the nucleotide analog 4-Thiouridine (4sU) is introduced to cells for various amounts of time. After RNA extraction, incorporated 4sU is chemically modified, causing it to be misread during reverse transcription and resulting in T to C mismatches during sequencing (Fig 1B). In developing mito-TL-seq, we found that 4sU was incorporated at a lower frequency in mitochondrial-encoded RNAs as compared to nuclear-encoded transcripts (Fig S1A). As high levels of 4sU lead to numerous secondary effects[17], we re-established computational approaches used in the analysis of these data[13,18] to properly quantify T-to-C conversions in mitochondrial RNA sequencing reads, given the low 4sU incorporation. We developed strategies to reduce background; such as creating single nucleotide polymorphism-cleaned mitochondrial genomes for each cell line and determined the T-to-C conversion rates using a custom binomial mixture model[13]. Next, we determined the fraction of new transcripts for each gene by the analysis of mismatches that occur after sequence alignment

and used a simple exponential decay model to estimate the half-lives of all transcripts. Using this approach, we analyzed HeLa cells labeled with as little as 50 uM 4sU, a concentration that is known to cause minimal to no secondary effects[17]. Our turnover rates (Fig 1C, Table S1) were reproducible between replicates and independent of 4sU concentration (Fig S1D). To test whether 4sU-incorporation impacted the stability of labeled RNA or if sequencing affected our estimates of RNA stability, we applied an alternative method where we analyze native RNA depleted from nascent 4sU-labeled RNA using a probe-based approach, MitoStrings (Fig S1B, C, and D, Table S1)[13,19]. We found a strong correlation between the MitoStrings and the mito-TL-seq half-lives (Figure S1D, $R^2 > 0.9$).

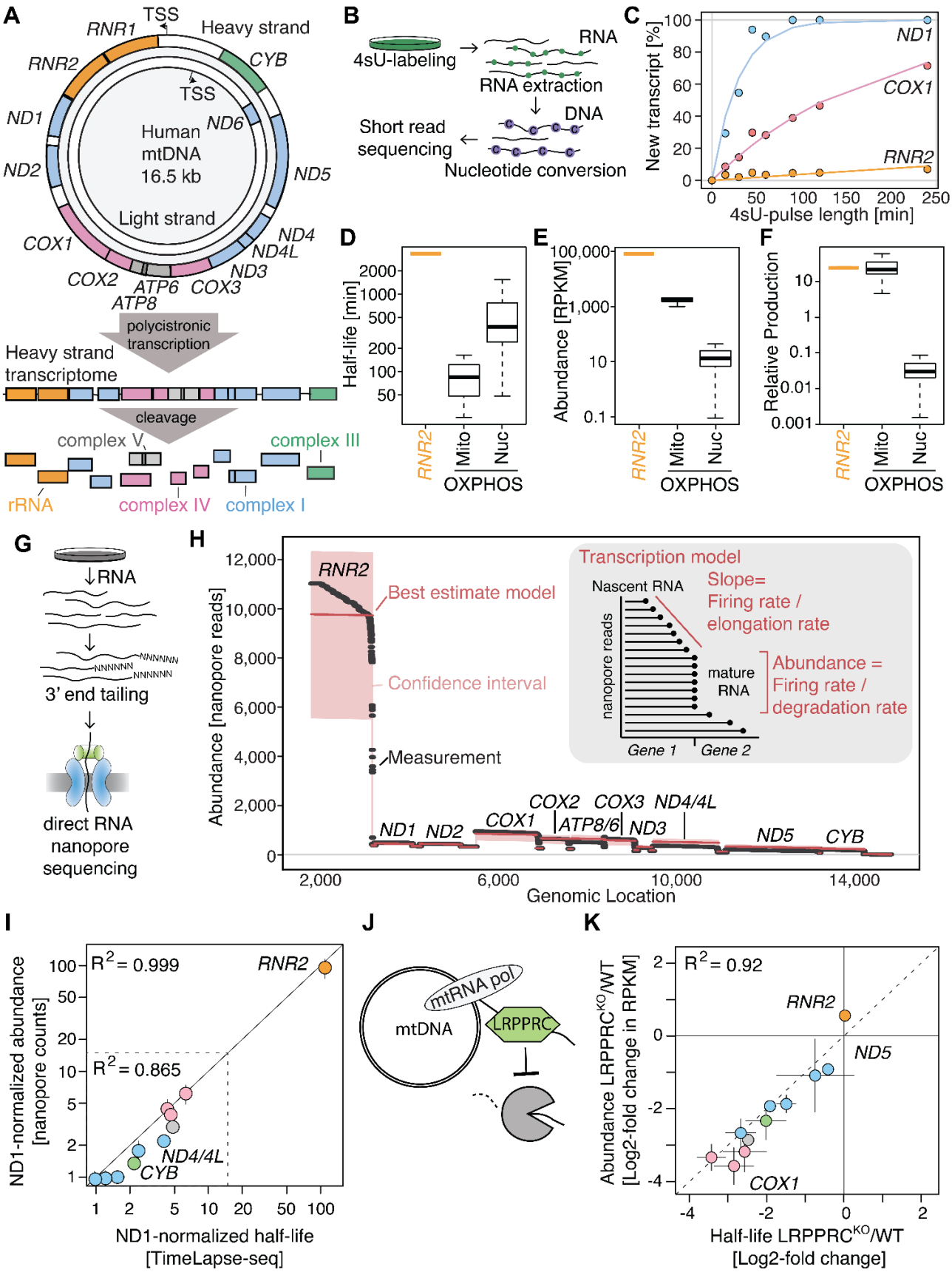

A
TSS
RNR1
Heavy strand
RNR2
CYB
TSS
ND1
ND6
Human
mtDNA
16.5 kb
ND2
ND5
Light strand
COX1
ND4
ND4L
ND3
COX2
ATP6
COX3
ATP8
polycistronic transcription
Heavy strand transcriptome
cleavage
complex V
complex III
rRNA
complex IV
complex I
B
4sU-labeling
RNA
RNA extraction
DNA
Short read sequencing
Nucleotide conversion
C
New transcript [%]
ND1
COX1
RNR2
4sU-pulse length [min]
D
Half-life [min]
RNR2
Mito
Nuc
OXPHOS
E
Abundance [RPKM]
RNR2
Mito
Nuc
OXPHOS
F
Relative Production
RNR2
Mito
Nuc
OXPHOS
G
RNA
3' end tailing
direct RNA nanopore sequencing
H
Abundance [nanopore reads]
RNR2
Best estimate model
Confidence interval
Measurement
Transcription model
Nascent RNA
Slope= Firing rate / elongation rate
nanopore reads
mature RNA
Abundance = Firing rate / degradation rate
Gene 1
Gene 2
ND1
ND2
COX1
COX2
ATP8/6
COX3
ND3
ND4/4L
ND5
CYB
Genomic Location
I
ND1-normalized abundance [nanopore counts]
R2 = 0.999
RNR2
R2 = 0.865
ND4/4L
CYB
ND1-normalized half-life [TimeLapse-seq]
J
mtRNA pol
mtDNA
LRPPRC
K
Abundance LRPPRCKO/WT [Log2-fold change in RPKM]
R2 = 0.92
RNR2
ND5
COX1
Half-life LRPPRCKO/WT [Log2-fold change]

Figure 1. RNA turnover is the main determinant of the relative abundance of the heavy-strand encoded transcriptome. A) The heavy strand of the mitochondrial genome encodes two ribosomal RNA, 10 mt-mRNA (two bicistronic), and several tRNAs while the light strand encodes one mRNA as well as tRNA. Mitochondrial transcription is polycistronic, as all transcription start sites are upstream of the coding regions[10], resulting in long primary transcripts which are cleaved into individual rRNA, tRNA, and mRNA (color-coded by OXPHOS complex) by processing enzymes. B) Overview of TimeLapse-seq which produces measurements of fraction new RNA. Green dots indicate metabolically incorporated 4sU which are modified during TL chemistry and then misread into Cs (purple dots) during reverse transcription. C) Turnover profiles for two mt-mRNAs, *COX1* and *ND1*, and one mt-rRNA, *RNR2*. Dots show the measured data and lines indicate the best fit for an exponential model. D) Distribution of half-lives derived from the exponential model for the mt-rRNA *RNR2*, all heavy strand encoded mt-mRNAs ("Mito", n=9), and all nuclear mRNA encoding OXPHOS subunits ("Nuc", n= 76). E) Same as D), but for transcript abundance estimated by taking the RPKMs from the same experiment. Zero values were excluded. F) Production rate estimated from the turnover rate in D) and the abundance in E) by dividing the RPKM by the half-life. G) Schematic of nanopore direct RNA sequencing to acquire minimally biased RNA abundances. 3 prime-end tailing was accomplished through *in vitro* addition of a poly(A) tail or ligation of a short DNA linker. H) Reads from one direct RNA-seq experiment mapping to the mitochondrial heavy strand (in black), with the results from the transcription model (shown in the right red box) overlaid in red. The model takes into account turnover and transcription parameters to estimate RNA abundance. Firing rate = slope of 3 prime ends of mapped reads from direct RNA-seq/elongation rate[20,21]. I) *ND1*-normalized abundance from nanopore direct RNA-seq is plotted against ND1-normalized turnover rates from mito-TL-seq. Error bars are standard deviation and the Pearson R^2, calculated in non-log space, is displayed for all transcripts (top left corner) and all mt-mRNAs (excluding *RNR2*, top left corner in dashed line box). Two distal genes are highlighted. J) LRPPRC is a multifunctional protein that stabilizes most mitochondrial mRNAs. K) Scatter plot showing a strong correlation between the change in RNA abundance and the change in turnover in HEK293T cells depleted of LRPPRC ($LRPPRC^{KO}$) compared to wild-type (WT) cells. Data from two replicates are shown. Error bars show the standard deviation and the Pearson R^2. All data in the figure are from HeLa cells except K) which uses HEK293T WT and $LRPPRC^{KO}$ cells.

We compared the half-lives of nuclear- and mitochondrial heavy strand-encoded OXPHOS mRNA, as well as *RNR2* (*RNR1* is removed during an rRNA depletion step). Half-lives ranged broadly with mitochondrial-encoded mRNA being rapidly turned over about 5-fold faster than nuclear-encoded OXPHOS mRNAs (Fig 1D). By contrast, *RNR2* was turned over >20-fold slower than any OXPHOS mRNA. Relative abundances measured by mito-TL-seq showed that mt-mRNA are >100-fold more abundant than the nuclear-encoded OXPHOS mRNAs (Fig 1E, Table S2). Moreover, *RNR2* is 100-fold more abundant than the mt-mRNA (Fig 1E). To estimate production levels for each RNA, we divided the abundance estimates by the half-lives for each gene (Fig 1F), revealing that mitochondrial RNA production is more than 700-fold higher than nuclear-encoded OXPHOS mRNAs. Importantly, we found comparable production rates for mitochondrial rRNA and mRNAs (Figure 1F), consistent with the polycistronic nature of mitochondrial transcription.

To obtain higher accuracy measurements of RNA abundance with minimal biases from library-generation steps, such as PCR amplification, we used nanopore direct RNA-sequencing to quantify the steady-state heavy-strand transcriptome (Fig 1 G, H, S1E, Table S2). We asked to what extent heavy-strand encoded mRNA abundance can be explained by differential RNA stability by comparing the half-lives from mito-TL-seq with the abundance from direct RNA-seq. The coefficient of determination (R^2) is 0.858 indicating that more than 85% of the variability in mRNA abundance can be explained by degradation alone (Fig 1I). Importantly, we find that the dynamic ranges of each measurement are matched (with ND1-relative abundance and half-lives both ranging between 0 to 100), which is essential if turnover is truly a driver of the steady-state abundance of mitochondrial RNAs.

We sought to test our approach by using it in a system with different turnover rates, so we analyzed *LRPPRC* knockout and matched control cells[5]. Mutations in the *LRPPRC* gene cause the French-Canadian type of Leigh syndrome and the encoded protein is involved in mt-mRNA stabilization, poly(A) tailing, and translation (Fig 1J)[5,22,23]. Depleting cells of LRPPRC had profound

effects on the steady-state levels of most heavy-strand encoded RNAs, as seen previously (Fig S1O). Moreover, we found a linear relationship between wild-type mt-mRNA stability and their degree of destabilization in the absence of LRPPRC (R^2 = 0.59), indicating that LRPPRC plays a major role in establishing the relative abundance of the mitochondrial transcriptome (Fig S1P). Importantly, changes in RNA half-lives were found to correlate strongly with changes in the steady-state levels of RNA in these cells (R^2 = 0.92) (Fig 1K). Thus, manipulating transcript stability leads to predictable alterations in RNA abundance, emphasizing the large role of transcript stability in establishing the mitochondrial transcriptome.

Slow mt-RNA polymerase elongation contributes a detectable RNA fraction to the mitochondrial transcriptome

Interestingly, most of the transcripts that deviate from the comparison between turnover and abundance are encoded on the promoter-distal end (E.g. *CYB* and *ND4/4L*) of the heavy-strand (Fig 1I). Intuitively, when transcription is polycistronic, more actively transcribing RNA polymerases will have traversed the promoter proximal genes before the promoter-distal genes at any one point in time (Fig S1Q, 1H). In a scenario where transcription is slow but RNA turnover is fast, transcription elongation will thus lead to more promoter-proximal than distal nascent RNA. We utilized a mathematical modeling approach[20] to formally predict total RNA abundance from the mature RNA turnover estimates and the nascent RNA due to transcription elongation (Fig 1H). We found that the transcript abundance differences of all heavy-strand genes could be well explained by the minimal model assumptions of a single transcription initiation site and rate, gene-specific RNA turnover rates, and an mt-RNA polymerase elongation rate estimated to be in the range of 0.06 - 0.12 kb/min (Fig 1H, S1R). These differ slightly from *in vitro* mt-RNA polymerase transcription elongation rate estimates, which range from 0.22 - 0.72 kb/min[24–26], which could indicate a transcriptional slow-down due to nucleoid-associated DNA binding factors. In summary, our kinetic measurements suggest that the steady-state mitochondrial transcriptome is shaped largely by differential degradation with transcription elongation providing an expression contribution for genes encoded proximal to the

heavy strand transcription start site, leaving little room for other mechanisms such as early transcription termination.

The light strand-encoded transcriptome is also determined by degradation kinetics

The light strand-encoded transcriptome consists of one mRNA, *ND6*, the 7S regulatory RNA, eight tRNAs, and numerous antisense (or 'mirror') RNAs with unknown functions. Quantifying the abundance of light-strand RNAs is uniquely challenging due to the strong expression of the opposite strand and slow or incomplete RNA processing. Strand-switching or index hopping in RNA-seq library generation[27] converted ~3% of the heavy-strand RNAs into contaminating reads in the light-strand mapping reads (Fig S1 F-I). Moreover, the accumulation of multiple light-strand RNA processing intermediates complicates their robust quantification by nanopore direct RNA-seq due to the limited read lengths of ~1 kb. Nevertheless, we reasoned that we could use our 4sU-induced sequencing mismatches to detect strand-switched reads (See MaM and Figure S1J and K). After correcting both the turnover (Figure S1L) and abundance (Figure S1K) measurements for contaminating reads, we found a strong correlation (R^2>0.9) between abundance and turnover (Fig S1 M and N), suggesting that the light strand-encoded transcriptome is also predominantly regulated by degradation kinetics.

Mitochondrial RNA processing is rapid and occurs predominantly co-transcriptionally

Mitochondrial precursor RNAs are processed into individual RNA molecules mainly through the excision of tRNAs that are encoded at the 5 prime and 3 prime ends of mRNAs[28]. We next investigated how RNA processing kinetics relate to transcription and turnover rates. We used two complementary approaches to estimate processing rates in HeLa cells. Firstly, nanopore direct RNA sequencing reads quantitatively capture both processed and unprocessed transcripts (Fig. 2A and B, Fig S2A). These data enabled measurements of the fraction of unprocessed transcripts at steady state, which we used to estimate processing rates (Fig 2C, Table S3). Secondly, we directly measured 5 prime-end processing for *COX2*, *COX3,* and *ATP8/6* mRNAs by performing a

time-series 4sU-labeling experiment capturing time points from 3 to 240 minutes. We then biotinylated and enriched the newly synthesized RNA and estimated the fraction of processed transcripts using MitoStrings probes that spanned processing junctions (Fig S2D, E). We found a good agreement between the indirect estimates with nanopore direct RNA sequencing and the direct measurements with MitoStrings (Fig S2F).

These analyses revealed that processing occurs rapidly with a median half-life of unprocessed transcripts under one minute (Fig 2D and Fig S2B). However, we found that some processing steps were notable exceptions, taking considerably longer to be processed (9-39 min). For example, the processing of *ATP8/6* and *COX3*, which had previously been observed to accumulate as a translation-competent tricistronic transcript[5,19,29], takes over 20 minutes (Fig 2A, D). We also observed the accumulation of the processing intermediate *RNR2-tRNA*^{Leu}*-ND1,* referred to as RNA19[29], (Fig 2B), and a form of this containing only *tRNA*^{Leu}*-ND1*, as well as significant amounts of unprocessed 5 prime ends of *COX1* (Fig S2A). We also observed the accumulation of these processing intermediates in direct RNA-seq libraries from HeLa cells, K562 cells, human myocytes, and myoblasts (Fig S2C). Finally, given that the mitochondrial transcription elongation rate is slower than 1kb/min (Figure 1H), we conclude that processing must occur predominantly co-transcriptionally.

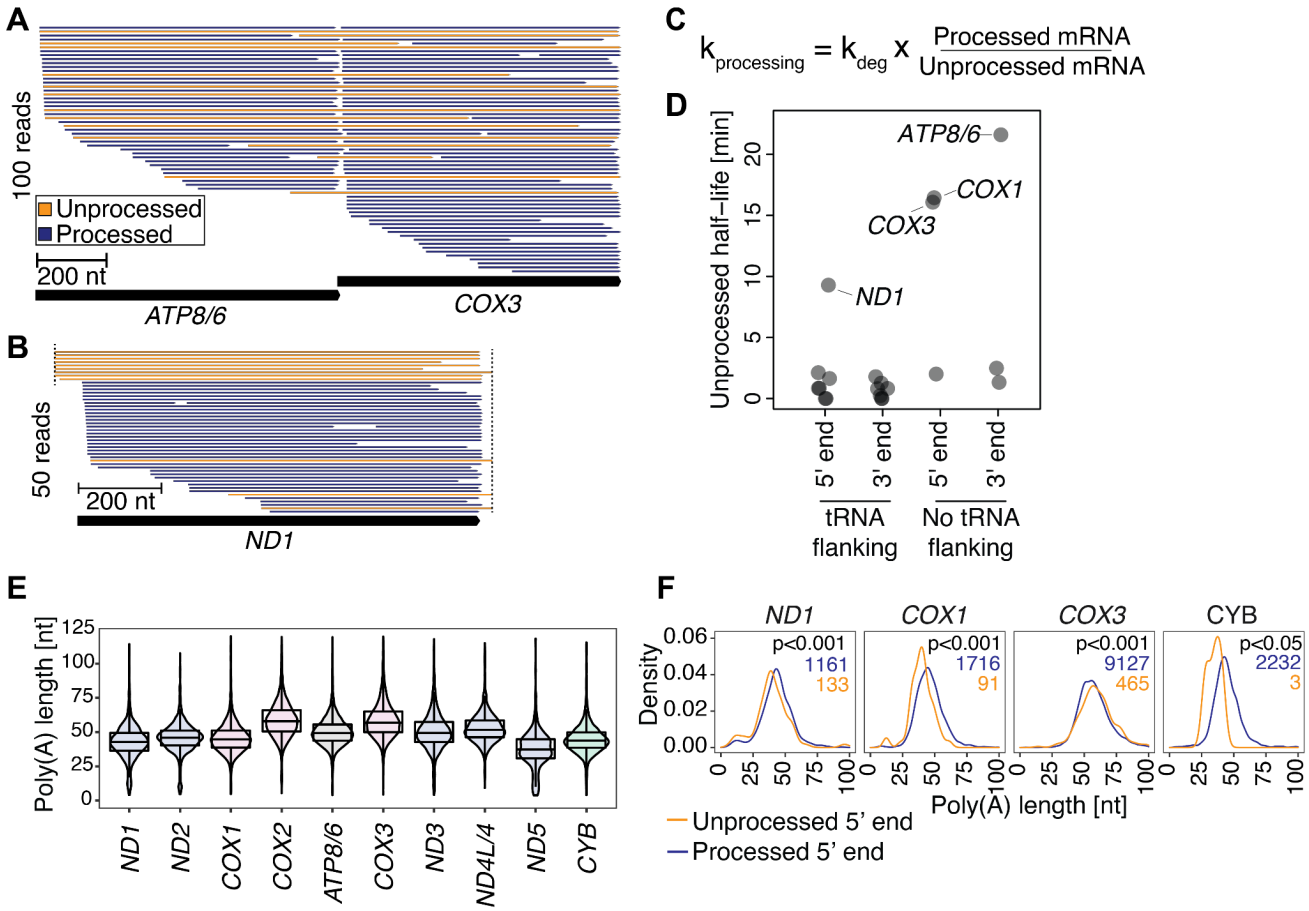


Figure 2. RNA processing is generally rapid with notable exceptions. A) 100 randomly sampled reads from nanopore direct RNA-seq that align to *ATP8/6* and/or *COX3* show many transcripts (orange reads) that are unprocessed at the 5 prime and/or 3 prime ends. B) Same as A) but for 50 reads mapping to *ND1*. C) Equation used to estimate processing rates from nanopore direct RNA seq and mito-TL-seq turnover measurements as shown in D). D) The amount of time required until half of the newly synthesized transcripts have been processed at a specific site. Processing sites are subdivided into 3 prime or 5 prime ends of the mRNA as well as into groups of canonical (tRNA-flanking) or non-canonical junctions. E) Distribution of poly(A) tail lengths measured by direct RNA-seq of poly(A)+ RNA. All pairwise comparisons between transcripts are significant (p-value < 0.05) by the Wilcoxon rank-sum test except *ND3* vs. *ATP8*/6. F) Density plot showing the distribution of poly(A) tail lengths for four genes. Orange lines show mRNAs with unprocessed 5 prime ends (more nascent on average) and blue lines show mRNAs with processed 5 prime ends. The numbers of reads underlying each distribution (in blue or yellow, respectively) and the p-value from the Wilcoxon rank-sum test comparing processed and unprocessed reads are displayed. All data in the figure are from HeLa cells.

Poly(A) tailing occurs rapidly after 3 prime end processing and can precede 5 prime end processing

Mitochondrial mRNAs, except ND6, have short poly(A) tails (median 48 nts)[30]. For most transcripts, the poly(A) tail is required to complete the stop codon[30]. To estimate poly(A) tailing kinetics, we adapted nanopore direct RNA sequencing to sequence RNAs with and without a poly(A) tail by ligating a short DNA linker to the 3 prime ends of all RNAs. With these data, we compared the fraction of mRNAs that have been cleaved at the 3 prime-end but not yet polyadenylated to those that are both cleaved and polyadenylated. Comparable to previous measurements[30], we found that most mRNAs have poly(A) tail lengths between 40 and 60 nucleotides, with *ND5* having shorter poly(A) tails than other transcripts (median: 28-37 nts, depending on the library preparation method, Fig 2E, Fig S2G, Table S4). Only a small fraction of cleaved transcripts (< 4%) still lacked poly(A) tails (Fig. S2H). Due to the low level of non-polyadenylated reads, there was not sufficient coverage to calculate polyadenylation rates. Nevertheless, the results indicate that polyadenylation occurs very rapidly following 3 prime-end cleavage**.**

To analyze the timing of 3 prime end polyadenylation relative to 5 prime end processing, we compared the lengths of poly(A) tails as a function of the 5 prime processing status and saw a significant difference for several mRNAs (Fig. 2F). *ND1*, *COX1,* and *CYB* mRNAs with unprocessed 5 prime-ends displayed shorter tails than those mRNAs with processed 5 prime-ends, an effect that was observed across different cell types (Fig. S2I). Nevertheless, the tail lengths for the unprocessed transcripts were still sizable, with a median of only 4-9 nucleotides shorter than the processed transcripts. Thus, although we observe that the initiation of polyadenylation is rapid, given that mRNAs with unprocessed 5 prime-ends are younger on average, these data suggest that the lengthening of the poly(A) tail is a slower process[31]. Interestingly, *COX3* mRNA showed the opposite trend, with unprocessed transcripts having slightly longer tails (median 2.3 nt longer) than the processed transcripts. As poly(A) tail length reflects the age of the transcripts, this observation is consistent with a higher stability of the tricistronic ATP8/6-COX3 mRNA than the processed COX3

mRNA. In sum, poly(A) tailing is initiated rapidly on the same time scales as processing and nascent mRNA (with unprocessed 5 prime ends) bears almost full-length tails consistent with poly(A) tailing being a one-step process[31].

mt-mRNAs arrive rapidly to mitoribosomes with rates that vary by 4-fold

mt-mRNAs are recruited directly to the monosome with the start codon acting as the only known ribosomal recruitment signal[32]. However, how long mt-mRNAs take before they associate with the mitoribosome is unknown and we performed mito-TL-seq on RNA associated with ribosomes, to answer this question. To enrich for mitoribosomes we performed sucrose gradient fractionations (Fig S3A). Because mtDNA and associated nascent RNA co-purify with the mitoribosome in sucrose gradient fractions[33], we performed DNAse treatment to remove any contaminating mtDNA and immunoprecipitated the mitoribosome from the fractions containing mitochondrial monosomes and polysomes, an approach we have shown faithfully reflects the full mitochondrial translatome[5] (Fig 3A, Fig S3B). These additional purification steps resulted in a strong enrichment (>20-fold) of mt-mRNA and mt-rRNA (Fig 3B) over extracting RNA directly from polysomes (compare Fig 3B with S3C and D). Due to the analysis challenges for light-strand RNAs described above, *ND6* was not included in these analyses.

To determine ribosome association rates, we developed a mathematical model (Fig 3C) that assumed equal mRNA degradation rates both for the unbound “free” and ribosome-bound transcripts. An Akaike information criterion (AIC) test consistently preferred this simpler model compared to a slightly more advanced model with independent degradation rates in the “free” and ribosome-bound state. Regardless of which model we applied, the ribosome-association rate estimates were very similar (Fig S3E). We first determined that the time it takes for *RNR2* to reach the monosome is >200 minutes, which is consistent with estimates of mitoribosome assembly of 2-3 hours (Fig 3D, E, Table S5)[34]. For mt-mRNA, the association times varied with half-lives ranging from 13 minutes for *COX1* to 3 minutes for *COX3*. The variability in mitoribosome association rates for

mt-mRNA was surprising because a physical barrier such as the nuclear envelope does not exist between nascent RNA and ribosomes in mitochondria and mt-mRNAs have minimal or absent 5 prime UTRs. However, these measurements were repeated three times in HeLa cells, twice using 100uM 4sU and once with 500uM 4sU to achieve higher signal-to-noise in the detection of T to C conversions (Fig 3E), with consistent results (Fig S3F). Comparison to association rates in HEK293T cells produced similar results as those in HeLa cells (Fig S3G). These trends are only weakly related to the size of the 5 prime UTR, most of which are 0 or 1 nts, or the start codon (Fig S3H). In conclusion, our ribosome association kinetics measurements show a surprising ~4-fold variability across mt-mRNA.

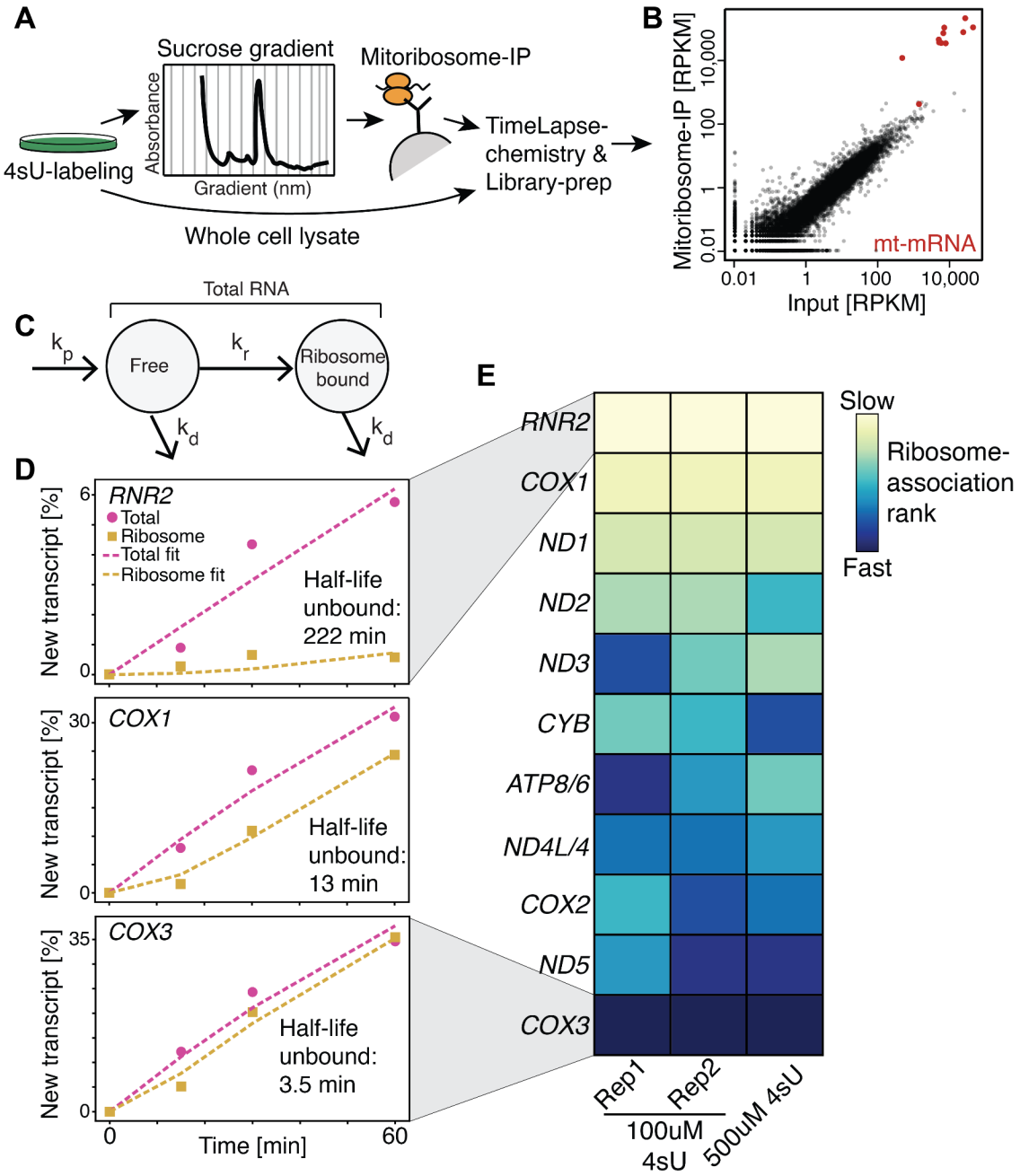


Figure 3. Mitoribosome-association kinetics are variable between transcripts. A) Cell lysates from 4sU-labeled cells are fractionated using sucrose gradients and the mitoribosomes are immunoprecipitated (IP) out of the gradient. RNA is extracted from the IPs and input (whole cell lysate) and sequenced by mito-TL-seq. B) Scatter plot showing the reads per kilobase per million reads (RPKM) in libraries from the mitoribosome-IP experiment and the input samples for the same experiment. Mt-ribosomal mRNA and mt-rRNA are enriched in the IP and shown in red. C) Model used to calculate ribosome-association rates. k_{deg} = degradation rate and k_r = transfer rate from the unbound to the ribosome-bound state. RNA can be either unbound "free" or bound by ribosomes. D) Three example profiles from one replicate experiment performed using 100uM 4sU. Black squares show the fraction new in the mitoribosome-IP and black circles show the fraction new in the input (Total). Dotted purple and yellow lines show the best fit for the total and mitoribosome-IP, respectively, deploying the model in C). Ribosome-association half-lives (the time it takes for half of the transcripts to bind to the ribosome), in minutes, estimated by the model for each gene, are shown (half-life unbound). E) Heatmap showing the rank of mt-RNAs in each replicate based on their ribosome-association half-lives. All the data in the figure are from HeLa cells.

mt-mRNA 5 prime but not 3 prime processing is a prerequisite for mitoribosome-association

Without physical barriers between transcription and translation, we investigated what instead could determine the range of ribosome-association rates. As the directly preceding step, we reasoned that RNA processing might be a rate-limiting step for ribosome association. Indeed, we found that generally ribosome association rates (median time until half of mRNA associated with ribosome: 5 min) were generally slower than processing rates (median 5 prime and 3 prime processing half-life: 100 s), albeit with three exceptions (Figure 4A and B). *ND1* and *COX1* 5 prime-end processing took the same amount of time as their ribosome association and the tricistronic *ATP8/6-COX3* transcript was processed slower than its ribosome association time (Fig 4A and B). These results suggest that processing for most transcripts occurs before mitoribosome association.

To validate this, we analyzed mitoribosome profiling data to determine whether ribosome-protected fragments (RPF) contain uncleaved mt-mRNA (Fig 4C-F)[5]. We found several transcripts with unprocessed 3 prime-ends in the ribosomes (Fig 4D, F and Fig S4C, Table S3), which is consistent with the fact that mRNAs lacking stop codons are nevertheless commonly translated by mitoribosomes[35]. However, only one transcript, *COX3*, had unprocessed 5 prime ends in the ribosome (Fig 4C, E, and S4B), in stark contrast to the nanopore direct RNA-seq where we found a large fraction of unprocessed transcripts (Fig 2). Accordingly, the uncleaved 3 prime-end of *ATP8/6* was by far the most common form of unprocessed 3 prime-end in the RPF data, consistent with the preference for *COX3* translation initiation to occur on the tricistronic transcript[5]. We interrogated ribosome profiling data and RNA-seq data from primary human fibroblasts and found consistent results (Fig S4D-E). The lack of unprocessed mRNA 5 prime ends in the ribosomes supports the idea that 5 prime-end processing is the rate-limiting step for translation for *COX1* and *ND1*. Altogether, mt-mRNA 5 prime-end processing is generally fast; a few sites are much slower to be cleaved and among transcripts with unprocessed 5 prime ends, only COX3 is found in translating

mitoribosomes (Fig 4G). Thus, 5 prime processing of *ND1* and *COX1* is a likely prerequisite for translation (Fig 4H). These results indicate that translation in mitochondria, in contrast to many bacteria[36], is unlikely to be co-transcriptional, despite the lack of a physical barrier between the two processes.

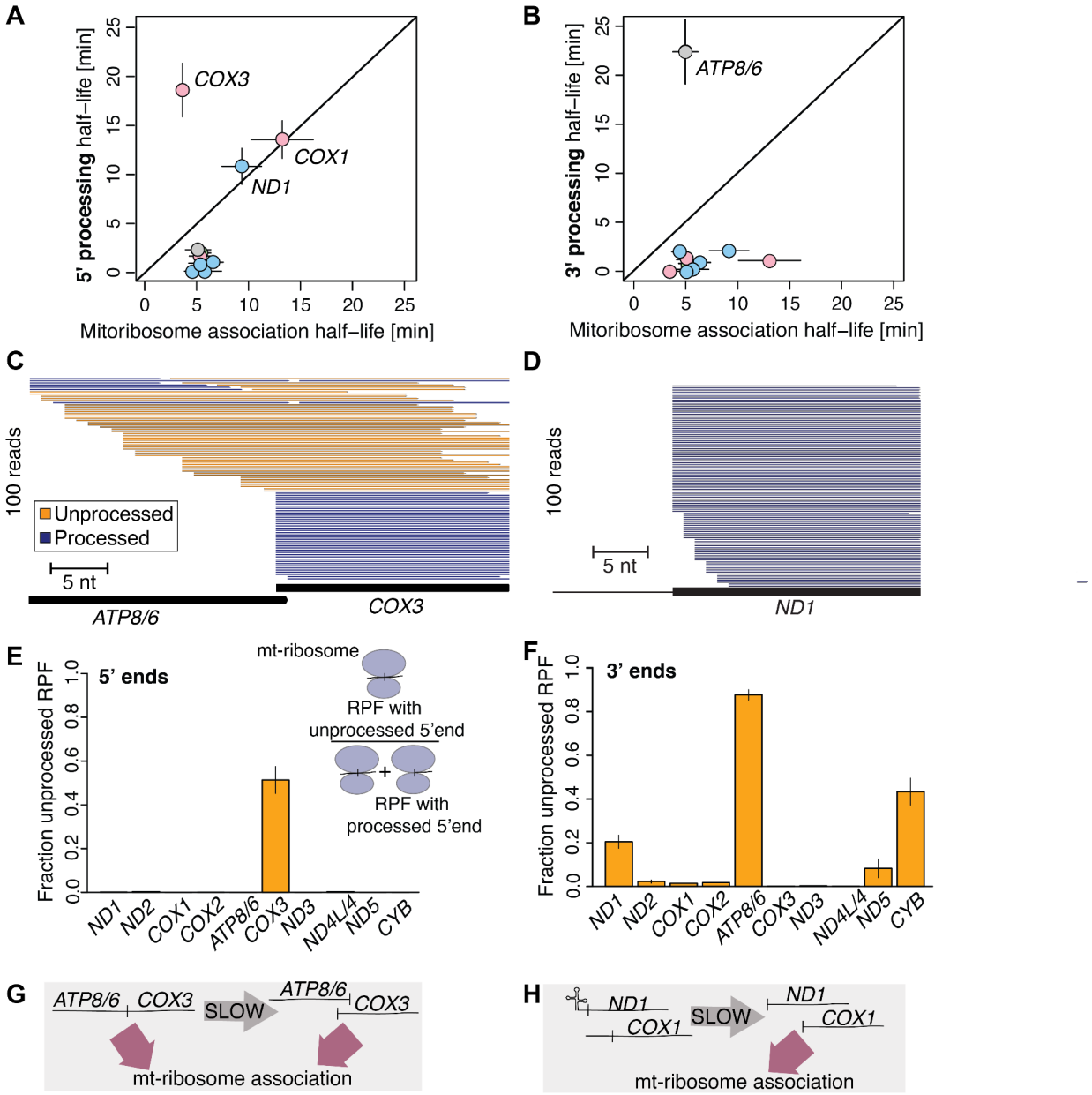


Figure 4. 5 prime RNA-processing might be a prerequisite for mitoribosome association. A) Scatter plot correlating 5 prime-end processing and mitoribosome-association half-lives. Error bars show standard deviations for three replicates. B) Same as A) but showing the 3 prime-end processing half-lives. C) Subsampling of 100 ribosome-protected fragments (RPF) with 5' ends within -30 and +3 nts from the 5' end of *COX3*. The scale bar shows the distance in nucleotides and the color code refers to the processing status of each read. D) Same as C) but for *ND1*. E) Quantification of all ribosome-protected fragments that align to the 5 prime ends of mitochondrial genes. The plot shows the fraction of the unprocessed reads (i.e. orange in D) and C)) over all reads aligning to the 5 prime end of the gene (i.e. blue and orange reads in D) and C)). Error bars show ranges of two replicates. F) Same as in E) but for 3 prime ends. G) Cartoon showing that *COX3* is processed slowly but the processing status does not affect ribosome association. H) Cartoon illustrating that *ND1* and *COX1* mRNAs are both processed slowly at the 5 prime-end and that only processed transcripts are associated with the ribosome. All data in the figure are from HeLa cells.

Ribosome association is independent of mRNA turnover and fine-tunes mitochondrial translational output

Nuclear-encoded OXPHOS transcripts spend a long time unbound by ribosomes due to their compartment-specific transcription, nuclear export, and translation kinetics[13]. Comparing mitochondrial and cytosolic ribosome association rates, we found that nuclear-encoded transcripts spent >10x more time in a non-ribosome-bound state compared to mitochondrial-encoded transcripts (median time nuclear-encoded: 113 min, median time mitochondrial-encoded: 5 min; Fig 5A). Furthermore, when compared to the time nuclear-encoded RNA spends between nuclear export and ribosome association (i.e., median residence time in the cytoplasm when not bound to ribosomes: 13 min, Fig 5A), mitochondrial RNA still spent only one-third of that time unbound by ribosomes, illustrating how very rapid the process is in mitochondria[13]. Using the ribosome association rates, we calculated the fraction that is expected to be associated with the ribosome for each mRNA. We found that 90% of mt-mRNA is found ribosome-bound on average (Fig 5B), consistent with biochemical measurements[37], and far higher than for nuclear-encoded transcripts (Fig 5B, ~50%).

Nevertheless, we wondered if mitoribosome association might act as an independent regulatory point for mitochondrial OXPHOS subunit production. First, we found that the ribosome association rates were not correlated with mRNA turnover (Fig S5A). Considering the high level of ribosome-bound mt-mRNA, we next asked whether the steady-state abundance of mRNA and the fraction associated per transcript (ranging between 73% and 98%) could predict levels of protein synthesis as measured by ribosome profiling (Fig 5C). We found that using the fraction mRNA that was bound by ribosomes modestly improved the correlation with the RPF data compared to using total RNA levels (from R^2 = 0.63 to R^2=0.65), suggesting that ribosome association rates fine-tune gene expression in mitochondria.

To test the overall importance of ribosome-association kinetics in determining protein synthesis, we developed a quantitative model that incorporated our measured RNA turnover rates

(Fig 1D), transcription initiation and elongation rates (Fig 1H) and ribosome-association rates to predict protein synthesis levels calculated from ribosome profiling data (Fig 5D). The remaining variability we attributed to translation efficiency (TE = Ribosome profiling/RNA-abundance). Applying this model, we found that RNA turnover had the largest impact, predicting 59.4% of the variability in protein synthesis levels, followed by translational efficiency which explained 40.1% (Fig 5E). Transcription initiation and elongation did not contribute to the variation, and ribosome association increased the predictive power modestly by 0.5%.

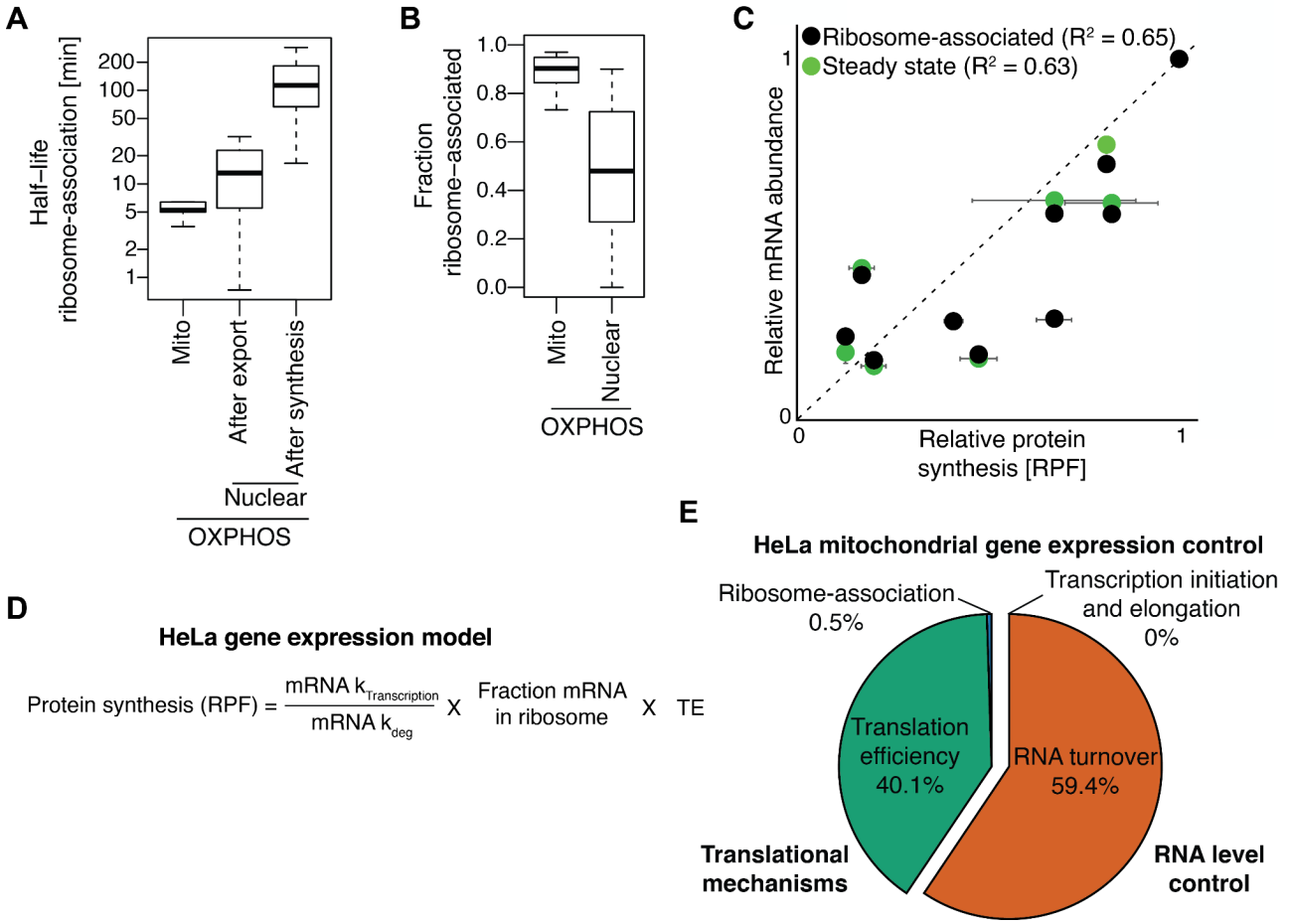


Figure 5. Ribosome association fine-tunes mitochondrial protein synthesis but has little predictive power relative to RNA turnover and translation efficiency. A) Distribution of ribosome-association half-lives for mitochondrial-encoded mRNA or nuclear-encoded OXPHOS subunits, measured either from nuclear export (shorter half-lives) or from synthesis (longer). Mito data is from this study and HeLa cells while nuclear-encoded data is from [13] and K562 cells. B) Distribution of the fraction of each mRNA that is bound by ribosomes. Mitochondrial mRNAs are on average more frequently associated with the mitoribosome than nuclear-encoded transcripts with the cytosolic ribosomes. C) Plot showing the correlation between either relative mRNA abundance at steady state (black dots) or mRNA abundance in the ribosome (green dots) and the relative protein synthesis as measured by ribosome profiling (ribosome protected fragments, RPF)[5]. In most cases, the green dots are closer to the midline than the black dots showing that ribosome association brings the total mRNA levels closer to relative protein synthesis levels. D) Quantitative gene expression model using the measured kinetic rates from this study. RPF: ribosome profiling, TE: translation efficiency, $k_{transcription}$: transcription rate, k_{deg} : turnover rate E) Proportion of protein synthesis explained by different steps of gene expression in HeLa cells. RNA turnover and translation efficiency explain the vast majority of protein synthesis as determined by ribosome profiling, while ribosome association and transcription play minor roles. All data in the figure are from HeLa cells.

A minimal model of gene expression predicts nuclear- and mitochondrial-encoded OXPHOS expression variation across cell types

Having identified multiple key regulatory points of mitochondrial gene regulation, we simplified the quantitative model we developed and compared the impact of different gene expression levels for both mitochondrial- and nuclear-encoded OXPHOS genes (Fig 6A). The model parameters were estimated, independently for the mitochondrial and nuclear genes, using only mRNA turnover, transcription, and translation efficiencies from HeLa cells (Fig 6A). As the model was fitted to explain 100% of the variability in the ribosome profiling data for a single biological replicate, we applied the model with the estimated parameters to predict RPF in a replicate dataset (Fig 6B) and found that the coefficient of determination was >0.9. Importantly, for nuclear genes variability in the transcription rate, and not mRNA turnover, largely predict the expression variation (Figure 6B), while for mitochondrial genes we found the reverse to be true: mRNA turnover but not transcription is the predominant determinant (Fig. 6B), consistent with our findings (Figs 1, 5). Interestingly, the translation efficiencies were less predictive for nuclear-encoded OXPHOS genes than for mitochondrial-encoded genes, explaining ~25% and ~47%, respectively.

To better understand the role of translation regulation, we compared published translation efficiencies (TEs) across five cell types, including human primary fibroblasts and myoblasts, and found that they correlate well (Fig S6A), suggesting that translation regulation could be fairly fixed in human mitochondria. To test this idea further, we used our minimal model of mitochondrial protein synthesis with all parameters, i.e. transcription, mRNA degradation, and TE, estimated from our HeLa cell fits. We found that HeLa transcription rates alone could predict ~46% of the variability in mitochondrial RPF in K562 and HEK293T cells (Fig 6B). However, when we applied the full model, including the HeLa TE data, the predictive power increased to 74% of protein synthesis in K562 and 85% in HEK293T cells (Fig 6B), consistent with translation efficiencies being comparable between these cell lines. The mitochondrial model slightly outperformed the nuclear model which explained 68% and 78% of RPF variability for K562 and HEK293T cells, respectively. The underperformance of

the nuclear gene expression model was due to the lower predictive power of the estimated TE parameter (~20% of RPF variability) compared to the mitochondrial model (~30%). Together, these models reveal that translation regulation is surprisingly robust and hardwired in mitochondria.

As a last test of our model, we analyzed how it performed when predicting mitochondrial protein synthesis in cells lacking LRPPRC, which associates with the ribosome and whose loss leads to disrupted polyadenylation and destabilized mt-mRNA[15,22,23]. We analyzed HEK293T cells lacking LRPPRC ($LRPPRC^{KO}$) and found that our quantitative HeLa-based model that performed well for other cell lines displayed little predictive power (Fig 6C). However, if we substituted the HeLa mt-mRNA degradation rates for $LRPPRC^{KO}$ degradation rates, the model was again able to explain half of the variability in protein synthesis alone (Fig 6C), similar to what we observed in other cell lines (Fig 6B). Interestingly, a model using the HeLa TE parameters and the $LRPPRC^{KO}$ mRNA degradation rates increased the predictive power to 73%. These results indicate that translation regulation is fixed even in cells with vastly different RNA abundance and that LRPPRC predominantly acts on RNA stability. While we found differences in $LRPPRC^{KO}$ TEs relative to wild-type cells (Fig 6C), they had a relatively small impact on the predictive power of our model.

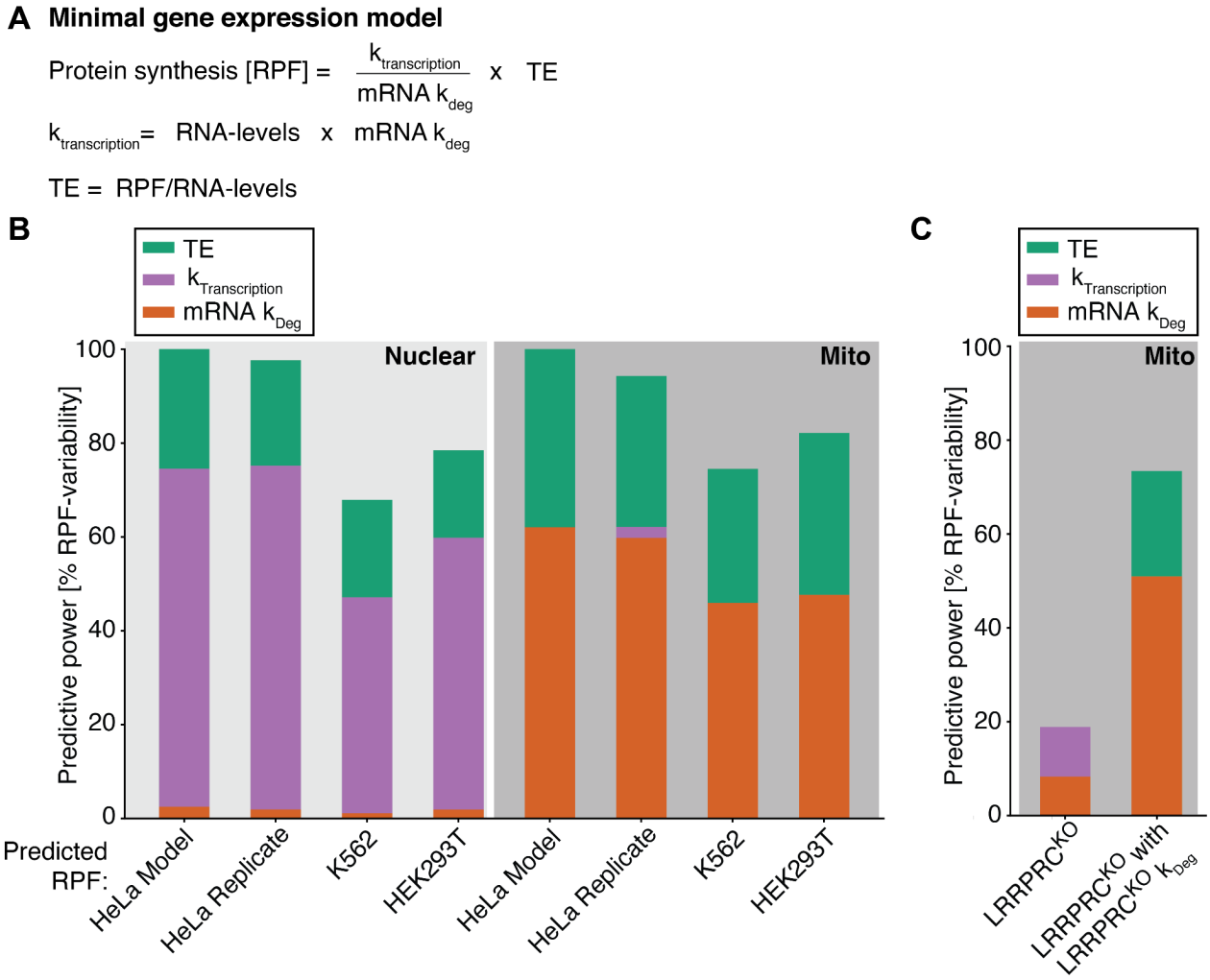


Figure 6. Mitochondrial gene expression is largely explained by RNA turnover and translation efficiencies even in situations where RNA levels are perturbed. A) Minimal model of mitochondrial gene expression used in B and C. B) The ability of the nuclear and mitochondrial models to predict variability in ribosome profiling data for nuclear or mitochondrial-encoded OXPHOS genes, respectively. The “HeLa Model” by definition explains 100% of the variability while the “HeLa Replicate" deploys the same estimated parameters to predict a replicate HeLa RPF dataset. The same model parameters are also used to predict RPF in K562 and HEK293T cells. Variability in nuclear gene expression is mostly explained by transcription levels while variability in mitochondrial gene expression is mostly explained by RNA degradation. TE is more predictive for mitochondrial genes than for nuclear-encoded genes. C) Same estimated HeLa parameters as in B) but predicting RPF in cells depleted of LRPPRC ($LRPPRC^{KO}$). Estimating the RNA degradation parameter (k_{deg}) using LRPPRC mRNA turnover data, instead of the HeLa k_{deg}, increases the predictive power.

Discussion

Here, we have developed a quantitative kinetic framework to analyze the human mitochondrial transcriptome, revealing many insights into mitochondrial gene regulation. We measured RNA processing, ribosome association, and turnover rates that differ starkly from nuclear-encoded OXPHOS genes, highlighting the unique features of mitochondrial gene expression. We found that the production rate of mt-mRNA is the same as for mt-rRNA, likely the most abundant transcript in mitochondria. These observations are naturally explained by our heavy-strand polycistronic transcription model that includes a single transcription start site[10] and rate. Our data reveal that RNA processing is very rapid relative to transcription elongation rates in human mitochondria, indicating that it is a co-transcriptional process. This is consistent with the fact that TEFM, the mitochondrial transcription elongation factor, is associated with RNA cleavage enzymes and that RNA processing is impaired in the absence of TEFM[38].

In bacteria, transcription factors directly interact with the ribosome creating an RNAP-ribosome expressome that co-transcriptionally translates the nascent mRNA[39] while in eukaryotes the physical separation between nuclear transcription and cytosolic translation prevents transcription-translation coupling. In mitochondria, we found that ribosome association is slower than processing and most likely initiated post-transcriptionally. Although an earlier model proposed that mt-mRNA first interacted with the small mitoribosome subunit and binding of the initiator tRNA stabilizes the mRNA-ribosome complex before monosome formation[40], it was recently shown that mt-mRNAs interact directly with the monosome[32]. Interestingly, our ribosome-association analysis indicates that 5 prime but not 3 prime processing is a prerequisite for monosome association, linking 5 prime processing status to translation. This is consistent with the finding that when *COX1* is poorly processed at the 5 prime end *COX1* is also poorly translated[41]. Mechanistically we propose that the uncleaved tRNA (or antisense tRNA) upstream of the mt-mRNA blocks the ribosome from binding prematurely to an mt-mRNA by blocking the start codon from entering the ribosome. In addition, we

noted a trend where mt-mRNAs that do not directly begin with a start codon (i.e. retain a short leader sequence) or use alternate start codons were slower to associate with the mitoribosome.

We measured a ~140-fold difference in transcript abundance between the mitochondrial- and nuclear-encoded OXPHOS mRNA (Fig 1) and found that mitochondrial transcripts are predominantly associated with ribosomes (Fig 5). If all mitochondrial ribosomes that associate with mRNA are translationally-active, mitochondrial ribosomes would have a translation rate constant (number of proteins synthesized per mRNA per time unit) of 1% relative to cytosolic ribosomes, assuming stoichiometric synthesis. Alternatively, mitochondrial-encoded proteins could be synthesized at the same rate and then turned over much faster than their nuclear-encoded counterparts. However, mitochondrial-encoded proteins are stable with half-lives much longer than the cell doubling time[42], leaving a low translation rate as the most likely explanation. What does that mean in absolute numbers? In HeLa cells, the translational output of the cytosolic ribosome has been estimated to be 1400 proteins/mRNA per hour, so the equivalent number for mitochondria would be 14 proteins/mRNA per hour. Considering some mt-mRNAs live less than an hour, many mRNAs would be translated only a few times in mitochondria.

We estimated parameters for a minimal gene expression model for mitochondria using our kinetic measurements over the mt-mRNA lifetime. Our findings indicate that RNA turnover is the prominent gene regulatory module, which alone explains up to two-thirds of the variability in mitochondrial protein synthesis. This is in stark contrast with nuclear-encoded OXPHOS subunits, which show variation due to their RNA production rates instead of mRNA turnover. Mitochondrial gene expression is then fine-tuned by ribosome association times and translation regulation. However, questions remain regarding the mechanisms by which mt-RNA half-lives are set and the extent to which RNA turnover is regulated between cell types and physiological situations in organelles. mt-mRNA levels vary broadly after the depletion of hundreds of factors and across cell lines, indicating that the mechanisms are likely to be complex, multifaceted, and important for cell

type-specific metabolism, although our analysis indicates that LRPPRC plays a predominant role (Figures 1K, S1O & S1P)[6,19,43].

Model parameters estimated in HeLa cells translated well to HEK293T and K562 cells, which can be explained by the consistency in both transcriptional and mRNA turnover in these cell types. This suggests that mt-mRNA may be largely hard-coded for translation efficiencies, resolving the apparent contradiction between the outsized importance of translation regulation in mitochondria and the lack of traditional regulatory features of nuclear-encoded or bacterial mRNA. In LRPPRC KO cells, RNA turnover remained a strong predictor of translational output and the HeLa TE parameter robustly increased the predictive power of the model. Thus, our modeling indicates that LRPPRC acts predominantly via its role in regulating RNA stability, illustrating how our approach can aid in identifying the core impact of regulatory factors. Applications of our models to other key regulatory factors, many of which play a role in disease[8], would illuminate the underlying mechanistic regulation of distinct mitochondrial gene expression modules. More broadly, our quantitative and kinetic framework for studying mitochondrial gene expression provides a critical foundation for gaining a deeper understanding of the mitochondrial and nuclear co-regulation driving OXPHOS biogenesis.

Acknowledgements

We thank Anna Wredenberg (Karolinska Institute), Christoph Freyer (Karolinska Institute), Jake Bridgers, and Chantal Guegler for their careful review of the manuscript. We thank members of the Churchman lab, Matthias Selbach (Max Delbrück Center for Molecular Medicine), and Karl McShane (Malmö Stad) for fruitful discussions. We thank Ari Sarfatis for help developing the data analysis pipeline; the Biopolymers Core Facility at Harvard Medical School for sequencing services and the Boston Children's Hospital Molecular Genetics Core for NanoStrings services.

Funding

This work was supported by National Institutes of Health grant R01-GM123002 (L.S.C.). E.M. is supported by European Molecular Biology Organization Long-Term Fellowship (ALTF 143-2018). K.C. is supported by post-doctoral fellowships from the Fonds de Recherche du Québec - Santé and the Canadian Institutes of Health Research. RI is supported by NIH/NIGMS T32 postdoctoral training grant GM007748-44.

Declaration of interests

R.I. is a founder, board member, and shareholder of Cellforma, unrelated to the present work.

Availability of data and materials

Raw and processed data will be available from GEO. Code for analysis of all data will be made publicly available at http://github.com/churchmanlab/mitoRNA_kinetics.

Methods

Cell culture

HeLa cells were grown in DMEM containing glucose and pyruvate (ThermoFisher, 11995073), supplemented with 10% FBS (Thermo Fisher, CAT A3160402.)

LRPPRC KO and matched WT HEK293T cells from[5] were cultured in DMEM supplemented with 20 µg/mL of uridine (Sigma, CAT U3750), 3 mM sodium formate (Sigma CAT 247596), 10% FBS, 1 mM GlutaMAX (Thermo Fisher Scientific, CAT 35050061) and 1 mM sodium pyruvate (Thermo Fisher Scientific, CAT 11360070).

Human myoblasts from anonymous healthy donors were kindly provided by Dr. Brendan Battersby (Institute of Biotechnology, University of Helsinki). Myoblasts were grown in Human Skeletal Muscle Cell Media with the provided growth supplement (HSkMC Growth Medium Kit, Cell Applications, 151K-500). For differentiation into myotubes, the media was replaced with DMEM (ThermoFisher, 11995073) with 2% heat-inactivated horse serum and 0.4 ug/mL dexamethasone. The differentiation media was replaced every two days.

4sU labeling

HeLa or HEK293T Cells were grown to 70-80% confluency and then pulsed by either 50, 100 or 500uM final concentration of 4-Thiouridine (4sU, Sigma T4509) in pre-conditioned cell culture media under standard cell culture conditions.

For mitoribosome-IP experiments, one 15 cm plate of cells was used per time point while for all other experiments, one 10 cm plate was used per time point.

Cell lysis and RNA extraction for standard time course experiments

Labeled cells were washed in ice-cold PBS (ThermoFisher) before being directly scraped in Trizol LS (ThermoFisher, 10296010). RNA extraction was performed using Trizol LS according to the

manufacturer's protocol, except with the addition of DTT (Sigma-Aldrich) at a final concentration of 0.2 mM DTT in the isopropanol.

Undifferentiated myoblasts (day 0) and differentiated myotubes (day 7) were collected and lysed in 1 mL of Qiazol (Qiagen). RNA from undifferentiated and differentiated myoblasts was extracted using the Qiagen miRNeasy kit (Qiagen) according to the manufacturer's protocol.

mtPolysome fractionation, mitoribosome-IP, and RNA extraction from mitoribosome-IP samples

After labeling cells were quickly washed in ice-cold PBS before being scraped in mt-polysome lysis buffer (0.25% Lauryl Maltoside, 10mM Tris pH 7.5, 0.5mM DTT, 20 mM magnesium chloride, 50 mM ammonium chloride 1× EDTA-free protease inhibitor cocktail (Roche)). Cell lysates were dounced 7 times in a 1mL dounce with the tight piston and then flash frozen. 1mL of thawed lysate was clarified by spinning twice at 10,000 rcf. To avoid contamination of nascent RNA still attached to mtDNA into the higher molecular weight sucrose gradient fractions we digested all DNA by 150 units of DNAse I (NEB) in the presence of superaseIn (ThermoFisher) and 0.5mM calcium chloride at room temperature for 1 h. To isolate mitoribosomes, lysates were loaded on 10–50% linear sucrose gradients and centrifuged in a Beckman ultra-centrifuge at 40,000 RPM for 3 h at 4°C using a SW41Ti rotor. Some of the input for the sucrose gradient fractionation was kept and treated as a standard sample described above. Gradients were mixed and fractionated using a BioComp instrument. To identify the mitoribosome containing fractions we western blotted for Mrpl12 (Proteintech 14795-1-AP) and Mrps18B (Proteintech 16139-1-AP) as described below. Monosomes and polysomes containing fractions were pooled and the mitoribosomes were immunoprecipitated out of the pooled fractions. For the immunoprecipitation, MRPL12 antibodies were conjugated to Protein A dynabeads (ThermoFisher) for 1 h in mt-polysome lysis buffer. After washing the beads, lysates were added and incubated for 3 h at 4C. After 3 h the supernatant was removed and the beads were washed three times in mt-polysome lysis buffer before the mitoribosomes were eluted in 0.2% SDS, 100 mM NaCl, 10 mM Tris pH 7.5, 1X EDTA-free protease inhibitor cocktail and

SuperaseIn. IP efficiency was confirmed by western blotting as described below. RNA was extracted from the eluates using Trizol LS as described above and then further cleaned by using 1.8x volume of RNAClean XP beads (Beckman Coulter A63987) and washed in 80% ethanol.

Western blotting

Samples were mixed with 4 x LDS sample buffer (ThermoFisher NP0007) and 0.1M DTT. Samples were loaded onto a 4-12% Bis-Tris gel (Invitrogen NP0321BOX) in 1x MOPs buffer and run at 160V for 1 hour. The gel was transferred to a nitrocellulose membrane using the wet transfer method in 1x transfer buffer (25mM Tris base, 192mM glycine, 20% methanol) at 400mA for 75 minutes at 4°C. The membrane was blocked in a blocking buffer (5% non-fat milk powder in 1x Tris buffer saline with 0.1% Tween 20) for at least 60 minutes. Primary antibodies were diluted in blocking buffer and incubated with membranes overnight at 4°C. Membranes were washed 3 x 20 minutes with 1 x TBST, incubated for 1 hour at 25°C with secondary HRP-conjugated antibodies against rabbit IgG (Cell Signaling Technology, 7074S), washed again 3x 20 minutes with 1x TBST, and developed using ECL Western Blotting Detection Reagents (Cytiva) and Amersham Hyperfilm ECL (Cytiva) using an automated M35 X-OMAT Processor developer (Kodak).

TimeLapse-sequencing

Extracted 4sU-labeled RNA from HeLa or HEK 293T cells were treated by TimeLapse-chemistry as described in[12,13]. In short, 2ug of RNA was treated with 0.1M sodium acetate pH 5.2, 4mM EDTA, 5.2% 2,2,2-trifluoroethylamine, and 10mM sodium periodate at 45°C for 1 hour. RNA was then cleaned using equal volume or RNAClean XP beads (Beckman Coulter A63987) and washed in 80% ethanol. The cleaned RNA was reduced in 0.1M DTT, in 0.1M Tris pH 7.5, 1M NaCl, and 0.01M EDTA for 30 min at 37 C before once again being cleaned by RNAClean XP beads. Sequencing libraries from treated RNA were created using the SMARTer Stranded Total RNA HI Mammalian kit (Takara 634873) following the manufacturer's instructions. Libraries were

sequenced on the NextSeq (Illumina, San Diego, CA) by the Biopolymers Facility at Harvard Medical School.

For the mitoribosome-IP the rRNA depletion step was left out for the IP:ed RNA as their cytosolic rRNA was already depleted by the IP step. In addition, the SMARTer RNA unique dual index kits (Takara 634452) were used instead of the primers that came with the SMARTer Stranded Total RNA HI Mammalian kit to allow for sequencing on the NovaSeq (Illumina, San Diego, CA) by the Biopolymers Facility at Harvard Medical School.

Creation of SNP-masked cell-line-specific reference genomes

To create cell-line-specific genomic single nucleotide polymorphism-corrected reference sequences, we used reads from samples without 4sU treatment. Reads were first aligned to the reference hg38 genome with STAR using parameters --outFilterMultimapNmax 100 --outFilterMismatchNoverLmax 0.09 --outFilterMismatchNmax 15 --outFilterMatchNminOverLread 0.66 --outFilterScoreMinOverLread 0.66 --outFilterMultimapScoreRange 0 --outFilterMismatchNoverReadLmax 1. Variants were then called with BCFtools (Li 2011) ‘mpileup’ and ‘call’ using two bam files as input. The resulting variant call file (VCF) was then split into a file with INDEL records only and a file without INDEL records (substitutions only). The "no INDEL" VCF was further split by frequency of substitution: loci covered by >= 5 reads and with a variant frequency >75% to a single alternate base were assigned the alternate base; loci with variants with an ambiguous alternate base were masked by "N" assignment. The reference FASTA was modified for these non-INDEL substitutions using GATK (McKenna et al 2010) FastaAlternateReferenceMaker. Finally, rf2m (https://github.com/LaboratorioBioinformatica/rf2m) was used with the INDEL-only VCF file to further modify the FASTA genome reference as well as the corresponding GTF annotation file.

Read Alignments

First, fastq files shared by the sequencing facility were concatenated to combine data across the sequencer lanes. The adapter sequences were trimmed from both of the paired-end reads using

Cutadapt v2.5 (Martin. EMBnet, 2011). Three more nucleotides were trimmed from the 5 prime end of the fragment to remove template switch bases, and five more from the 3 prime end to remove mismatches associated with random hexamer priming that result from the SMARTer Stranded library preparation. Read alignment was done using STAR 2.7.3 to cell-line-specific genomic reference sequences (see above) with different parameters for experiments with high and low 4sU treatment. For 500 uM 4sU experiments, parameters were --outFilterMismatchNmax 90 --outFilterMismatchNoverLmax 0.3 --outFilterMatchNminOverLread 0.2 --outFilterScoreMinOverLread 0.2. For 50 uM 4sU experiment, parameters were --outFilterMismatchNmax 10 --outFilterMismatchNoverLmax 0.05 --outFilterMatchNminOverLread 0.66 --outFilterScoreMinOverLread 0.66 --outFilterMultimapScoreRange 0.

T to C mismatch counting

For each sample, a subset of mitochondrial-aligned reads was used to determine the rate of T to C mismatches. Unless otherwise specified, this subset included all mitochondrial reads except for those aligning to MT-rRNA. For this analysis, reads were first filtered to remove non-primary alignments, reads with unmapped mates, and reads that map to more than 4 positions, and soft-clipped bases were removed. We then determined the number of T to C mismatches and the total number of T nts across each fragment (considering both read1 and read2 in each pair) using custom scripts from[13]. Two-dimensional distributions were generated containing the number of fragments with *n* Ts and *k* T to C mismatches.

Binomial mixture model to calculate T to C conversion rates

To estimate the 4sU-induced as well as background T to C conversion rates, we used a binomial mixture model we previously developed [13] to handle samples with low 4sU concentrations. In short, the binomial mixture model first estimates two background T to C conversion rates (p_{E1} and p_{E2}) and a global fraction (π_E) parameter, with $Binom$ a binomial distribution, given the measured *n* Ts and *k*

T to C mismatches in the unlabeled control sample. Giving the background T to C conversion rates as follows: $P_{BG}(k,n) = \pi_E Binom(k,n,p_{E1}) + (1-\pi_E)Binom(k,n,p_{E2})$. The three parameters in this model were fitted to the above T to C distributions using linear regression (Python v3.7.4, package lmfit v1.0.2, function minimize). We tested if our binomial mixture error model better fitted the T to C conversions of the untreated samples using the Akaike Information Criterion and adjusted our background model to include one or two background rates depending on the test outcome[44]. The 4sU sample T to C distributions were then modeled as the 4sU-induced T to C conversions (p_C) plus the background population $P_{BG}(k,n)$ once again with a global fraction (π_C) parameter: $P(k,n) = \pi_C Binom(k,n,p_C) + (1-\pi_C)P_{BG}(k,n)$. We note that it is necessary to calculate the T to C conversion rate from nuclear and mitochondrial encoded transcripts separately as the mitochondrial 4sU-incorporation rate is significantly lower than for nuclear-encoded genes. In experiments deploying high (500uM) 4sU concentration with no Uridine supplemented to the media, we had robust T to C conversion rates >1.5% and directly used GrandSlam to estimate conversion rates. The T to C conversion rates in the mitoribosome IP and input samples as well as samples from HEK293T cells were calculated from all reads aligning to the mitochondrial genome except reads aligning to COX1 and RNR2.

T to C conversion rate estimates for cases with very low 4sU incorporation

In two cases we ended up with T to C conversion rates lower than 0.6%: the earliest (15 min) time point for the HeLa mitoribosome IP experiments deploying 100uM 4sU (Fig 3D, E) and for all the time points in the HEK293T mitoribosome IP experiments (Fig S3 F). In these two cases, our binomial method failed and we instead calculated the T to C conversion rate based on the assumption that the ribosomal RNA would be turned over consistently. Practically, we set the T to C conversion rate such that it resulted in the fraction of new RNR2 being the same as at the equivalent time point in the high confidence time course experiments in Fig1. In the IP we set the T to C conversion rate so that the fraction of new RNR2 was lower than 0.5%, to allow for some noise. This

method gave similar results to the mitoribosome IP experiment with higher 4sU levels where we did not use any corrections (Fig 3E)

GRAND-SLAM parameters

Alignment files (.bam) containing all mitochondrial-aligned reads were converted into a .cit file using GRAND-SLAM v2.0.5f[18]. The no-4sU (0 min) alignment file was copied and both copies ("no4sU" and "0m") were included in the GRAND-SLAM analysis so that the "no4sU" sample would be used for the background T to C mismatch rate (p_E) and the sample details for the "0m" sample would also be output. GRAND-SLAM was run twice: first for .cit file creation and background rate determination, and a second time after modifying the T to C mismatch rates (p_C) according to the output of the binomial mixture model. Before the second run, the *.tsv and *ext.tsv files were removed from the output directory and the *rates.tsv file modified to add in new rates for single_new and double_new. For both runs the parameters used included -full -mode All -strandness Sense -snpConv 0.4 -trim5p 10 -trim3p 5.

Cell doubling correction of fraction new values

The fraction new RNA values for HeLa cells was corrected for dilution by cell doubling. The cell doubling time was estimated by growing HeLa cells under standard cell culture conditions to 70% confluency (0h). At 0h, 12h, and 24h, five replicates of 10cm plates were washed in PBS, and cells were collected by trypsinization. Cells were mixed with trypan blue and counted using an automated cell counter (Countess, Invitrogen). Cell doubling time was estimated to be 26.5 h by linear regression of the log2-fold change in cell number over time. The fraction new values were corrected as follows:

$$Corrected\ fraction\ new\ = Measured\ fraction\ new\ - \frac{\left(1-\left(\frac{1}{e^{\log(2)/\text{cell doubling time}}}\right)\right)^{\text{time point}}}{1-\left(1-\left(\frac{1}{e^{\log(2)/\text{cell doubling time}}}\right)\right)^{\text{time point}}}$$

Kinetic modeling of mtRNA degradation

Based on the maximum a posteriori (MAP) values obtained from GRAND-SLAM for each time point of the 4sU pulse, we tested two simple models to estimate the degradation rates from the time t dependence of newly synthesized Total mtRNA $T(t)$ [13,18,45].

Model 1: exponential (one-state) decay - RNA is produced with a transcription (production) rate k_p and eventually degraded at the rate k_{deg} (Fig S1F): $\frac{dT(t)}{dt} = k_p - k_{deg}T(t)$. To solve this ordinary differential equation analytically as described previously [13,18,45], we set all newly synthesized RNA levels to zero at t=0, i.e. before any 4sU pulse as the boundary condition. Next, the integrating factor method was used to obtain the solution: $T(k_p, k_{deg}, t)$. Since the observed quantities are fraction of new RNA, rather than RNA levels we derived the model fraction of newly synthesized RNA: $\Lambda_T^{model\,1}(k_{deg}, t) = \frac{T(t)}{T_{ss}}$, with $T_{ss} = \frac{k_p}{k_{deg}}$ the steady state levels of all (newly synthesized and pre-existing unlabeled) RNA, which equals the steady state solution to the ODE: $\frac{dT(t)}{dt} = 0$. Both T_{ss} and $T(t)$ are linear in k_p, so $\Lambda_T^{model\,1}(k_{deg}, t)$ no longer depends on k_p. So, we arrive at our solution:

$$\Lambda_T^{model\,1}(t) = 1 - e^{-k_{deg}(t)} \quad .$$

Model 2: non-exponential (two-state) decay - RNA first assumes state 1 $S_1(t)$ upon which it can either undergo exponential decay at the rate k_{d1} or it can transition to state 2 $S_2(t)$ with rate k_r from which it exponentially degrades with rate k_{d2} (Fig S1F):

$$\frac{dS_1(t)}{dt} = k_p - (k_r + k_{d1})\, S_1(t)$$

$$\frac{dS_2(t)}{dt} = k_r\, S_1(t) - k_{d2}\, S_2(t)$$

Here, the total amount of mtRNA is then given by the sum of both states: $T(t) = S_1(t) + S_2(t)$. Using the same procedure as described above for the 1 state model, the fraction new total RNA then becomes (equation 1)[45]:

$$\Lambda_T^{model\,2}(t) = \frac{S_1(t)+S_2(t)}{S_{1,ss}+S_{2,ss}} = \frac{k_{d2}(k_{d1}-k_{d2})}{k_{d2}(k_{d1}-k_{d2})+k_r(k_{d1}-k_r)}\left(1-e^{-(k_r+k_{d1})t}\right) + \frac{k_r(k_{d1}-k_r)}{k_{d2}(k_{d1}-k_{d2})+k_r(k_{d1}-k_r)}\left(1-e^{-k_{d2}t}\right).$$

To estimate the degradation parameters in model 1 (one parameter) and model 2 (three parameters), we used a non-linear least square fitting framework (Python v3.7.4, package lmfit v1.0.2, function minimize). Note, for all experiments we directly used the MAP values from GRAND-SLAM for the modeling except for the LRPPRC KO cell where 4sU incorporation was very low and turnover very high (and thus the precision in the measurement was low). In this case, if the measured values equaled 100% new at the earliest time point we decreased it to 95% new to allow the fitting to proceed smoothly.

To select the model to make predictions about the RNA decay parameters, we used the Akaike information criterion $AIC = N * ln(\frac{\chi 2}{N}) + 2\xi$, where N denotes the number of data points, χ^2 is residual square sum and ξ is the number of variable parameters. To further estimate which model has a higher probability we can define model weights to define the probability of model 1 over model 2. The difference in AIC for a given model *i* is given by $\Delta AIC_i = AIC_i - AIC_{max}$, and AIC_{min} is the minimum AIC across all models. Based on this Akaike weights are estimated by $W_i(AIC) = \frac{e^{-\frac{1}{2\Delta_i(AIC)}}}{\sum_{k=1}^{K} e^{-\frac{1}{2\Delta_k(AIC)}}}$, where K is the number of models. Since we are comparing two models we defined the normalized probability as π by $\Pi = \frac{W_i(AIC)}{W_i(AIC) + W_j(AIC)}$

Kinetic modeling of mtRNA association to the mitoribosome

To estimate the ribosome association rate, we reused the two-state model as described above, with the free mtRNA state, i.e. not bound to the ribosome, $F(t) = S_1(t)$ and the ribosome-bound mtRNA state $RB(t) = S_2(t)$. The ribosome association rate is then the transfer rate k_r from the free state to

the ribosome-bound state. In addition to the total RNA (equation 1 above), we now also fit the experimentally observed fraction of newly synthesized ribosome bound mtRNA:

$$\Lambda_{RB}(t) = \frac{RB(t)}{RB_{ss}} = 1 - \frac{k_{d2}}{k_{d2} - (k_{d1} + k_r)} e^{-(k_r + k_{d1})t} + \frac{k_{d1} + k_r}{k_{d2} - (k_{d1} + k_r)} e^{-k_{d2}t}.$$

We can further simplify these equations by making the assumption that $k_{deg} = k_{d1} = k_{d2}$:

$$\Lambda_T^{simplified}(t) = 1 - e^{-k_{deg}(t)}$$

$$\Lambda_{RB}^{simplified}(t) = 1 + \frac{k_d e^{-(k_r + k_{deg})t} - (k_r + k_{deg}) e^{-k_{deg}t}}{k_r}$$

To estimate the degradation and mitoribosome association rates in the simplified model (two parameters) and full model (three parameters), we used a non-linear least square fitting framework (Python v3.7.4, package lmfit v1.0.2, function minimize) to fit the total and ribosome-bound mtRNA. To select the model to make predictions about the mtRNA kinetic parameters, we used the Akaike information criterion as discussed above.

Lastly, to calculate the fraction of nuclear-encoded mRNA that was bound by the cytosolic polysome, we derived the following equation, with nuclear residence k_{nuc}, cytoplasmic turnover k_{cyto} and cyto-polysome entry rates k_r as determined in[13]:

$$fraction\ cyto\ polysome\ bound = max\left(\frac{k_{cyto}^{-1} - k_r^{-1}}{k_{nuc}^{-1} + k_{cyto}^{-1}},\ 0\right)$$

Transcription elongation model

In order to model the mitochondrial transcription elongation process, we first processed the Nanopore sequencing reads according to the assumption that total RNA consists of nascent and mature RNAs. Reads with a 3 prime end within a 100nt window of annotated mature RNA 3 prime ends contribute to the mature RNA fraction with coverage across the whole of the gene. Reads with ‘3 ends within the annotated genes, but outside of the 3 prime end window contribute to the nascent RNA fraction with coverage across the whole genome up to the from the single TSS in the

mitochondrial genome, i.e. the HSP2 3 prime end site. The sum of nascent and mature RNA fractions makes up the total RNA coverage across the genome. The model mature RNA fractions equal $\frac{F}{d_i}$ and the model nascent RNA fraction equals $\frac{F}{v}(x - TSS)$, with x the distance (unit: nt) downstream of the TSS, as previously described in[20,21]. F is the transcriptional initiation (Firing) rate, v the MT Polymerase elongation rate, and d_i the mature RNA degradation rate, specific to gene i, as estimated above through TimeLapse sequencing. The initiation rate F can then be estimated from the observed mature RNA counts from each gene i: mature $RNA_i = F_i / d_i$. This leads to an ensemble of F estimates: $\{F_i\}$. Since RNR2 is the most upstream and highly expressed, we used this "best estimate": $\widehat{F}$. We also calculate the $F_{min} = min_{i \in genes}\{F_i\}$ and $F_{max} = max_{i \in genes}\{F_i\}$ as the estimate bounds. To estimate v, we use linear regression to fit the nascent RNA fraction model, whilst utilizing our $\{F_i\}$, to get an elongation rate ensemble $\{v_i\}$ (Python package: lmfit, function: minimize, velo_seed = 3.81 * 60, equalling the previous estimate by[25], velo_max = 100*1000 #units: nt / min) from RNR2 3 prime end to the 3 prime end of the genome. $\widehat{v}$ is then the elongation rate paired with $\widehat{F}$, and $v_{max} = max_{i \in genes}\{v_i\}$, likewise for v_{min}. Altogether, our estimates are then used to predict total RNA coverage with $(\widehat{F},\widehat{v})$ as best estimate (red lines) and (F_{min}, v_{min}) and (F_{max}, v_{max}) as error bands (Figure 1H).

Ribosome profiling of K562 cells

For mitoribosome profiling $5x10^6$ K562 cells were grown at a concentration of $5x10^5$ cells/mL. Upon removing media, cells were rinsed with ice-cold PBS and resuspended in 600µL of mitoribosome lysis buffer (0.25% lauryl maltoside, 50 mM NH_4Cl, 20 mM $MgCl_2$, 0.5 mM DTT, 10 mM Tris, pH 7.5, and 1× EDTA-free protease inhibitor cocktail (Roche)). Before nuclease digestion, 5% NIH-3T3 cell lysate, based on OD_{260} equivalence, was spiked-in for normalization. The lysates were RNA digested using 8 units/µL of RNase If (NEB) for 30 min at room temperature to generate mitoribosome footprints. The digestion was then inhibited by the addition of 80 units SUPERaseIn (Thermo Fisher

Scientific) followed by centrifugation at 10,000 RPM for 5 min to clarify the lysate. For isolating mitoribosomes, 450 µL of clarified lysate was added on to 12 mL of 10-50% linear sucrose gradient and centrifuged at 40,000 RPM for 3h at 4°C using an SW41Ti rotor. The gradient was fractionated into 800 µL fractions using a BioComp instrument and the fractions containing mitoribosomes were identified by western blotting against MRPL12 and MRPS18B. From the monosome fraction, RNA was extracted using 1:1 phenol/chloroform extraction and separated on 15% polyacrylamide TBE-urea gel to collect RNA fragments between 28-40 nucleotide size. Libraries were prepared as previously described along with the modification for mitoribosome library preparations[5]. Sequencing was performed on an Illumina Next-seq 500 at the Bauer sequencing facility (Harvard University). For cytosolic ribosome profiling $5x10^6$ K562 cells were grown at $5x10^5$ cells/mL. Upon removing the media, cells were rinsed with ice-cold PBS and resuspended in 600µL of cytoribosome lysis buffer as previously described[46]. Cytoribosome footprints were generated by digesting the RNA using 8×10^{-4} units/µL of RNase I (Epicentre) at room temperature for 30 min, followed by the addition of 80 units of SUPERaseIn (Thermo Fisher). The cytosolic monosome was isolated using differential ultracentrifugation using 10-50% sucrose gradient at 40,000g Sw 41Ti for 2hr 30min. The fractions containing the cytosolic monosome were detected by their UV absorbance at 254 nm. Cytoribosome footprints were recovered and gel size was purified in the same manner as for the mitoribosome profiling but the gel was excised at 27–33 nucleotides. Library preparation and sequencing were the same as for mitoribosome profiling.

Ribosome footprinting analysis

Mitochondrial ribosome profiling data from above and from[5] (GEO accession number GSE173283) was initially processed as previously described[5]. To analyze 5 prime and 3 prime-end processing status of ribosome-loaded mt-mRNA, the position of both the 5 prime and 3 prime ends of each read were recorded. Soft-clipped bases were ignored for recording positions but flagged if they were determined to come from a 3 prime poly(A) tail (>80% of soft-clipped bases are A). A read was considered processed if its terminus is within two nucleotides of the processed transcript terminus. A

read was considered unprocessed if it spans at least three nucleotides beyond a processed transcript terminus and was not flagged as containing a poly(A) tail.

MitoStrings experiments

mt-RNA turnover in HeLa cells was determined by measuring the decrease of the old/unlabeled mtRNA after a 4sU-pulse as described in[13]. In short, HeLa cells were labeled with 500uM 4sU for 0, 60, or 120 min and cells were lysed in Trizol LS and RNA was extracted according to the manufacturer's protocol. A spike-in of *in vitro* transcribed ERCC-000148 was added to all the cell lysates during the RNA-extraction step. The RNA was denatured at 60C for ten minutes before being biotinylated by the addition of 5ug/ml biotin-MTS (Biotium) in 20% dimethylformamide (Sigma), 20mM Hepes pH 7.4 and 1 mM EDTA for 30 min at room temperature. Free biotin was removed using phase-lock heavy gel tubes (5prime) following the manufacturer's instructions. Biotinylated RNA was removed by incubation with streptavidin beads from the uMACS Streptavidin kit (Miltenyi Biotec) for 15 min at room temp. Beads and RNA were loaded onto a uMacs column. Beads were washed with 100mM Tris pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween 20, and unlabeled RNA was collected in the flowthrough. RNA in the flow-through was purified using the miRNeasy Nano kit following the manufacturer's instructions including the DNAse treatment step (Qiagen). 30 ng of RNA was incubated for 16h at 67°C with the XT Tagset-24 (NanoString Technologies) and with DNA-probes specific for mitochondrial RNA (Original Mitostring-probes[19] were modified as in[47]) in hybridization buffer (NanoString Technologies) according to the manufacturer's protocol before being loaded onto a nCounter Sprint Cartridge and quantified using the nCounter SPRINT Profiler (NanoString Technologies) at the Boston Children's Hospital Molecular Genetics Core.

For direct measurements of processing kinetics HeLa cells were labeled with 500uM 4sU for 0, 1.875, 3.75, 7.5, 15, 30, 60, 120, or 240 min. RNA was extracted as described above with the addition of a second *in vitro* transcribed 4sU-labeled (10% of UTP was exchanged for 4sUTP in the T7 RNA polymerase *in vitro* reaction (NEB)) spike-in ERCC-00136. Biotinylation and streptavidin-bead binding were performed as above with the exception that in this case the

flowthrough was discarded and bead-bound nascent RNA was eluted by incubating the beads at 2 x 5 min at 65C in 100mM TCEP (Thermo Scientific), 10mM EDTA, 500mM Tris at a final pH7. Eluted RNA was purified, hybridized with Mitostring probes, and analyzed on a nCounter Sprint Profiler as above.

Direct nanopore RNA sequencing

RNA was extracted as described above using Trizol LS and chloroform. Extracted RNA was rRNA-depleted by hybridization with a custom set of biotinylated-DNA probes complementary only to cytosolic rRNA (riboPOOL, siTOOLS). In short, 6ug of RNA was mixed with 5uM riboPOOL probes in 20 uL of 2.5mM Tris-HCl (pH 7.5), 0.25mM EDTA, and 500uM NaCl and hybridized by incubating at 68°C for 10 min followed by a slow cooling step decreasing the temperature to 37°C. 20uL of hybridized samples were mixed with Dynabeads MyOne Streptavidin C1 beads (ThermoFisher) in 80uL of 5mM Tris-HCl (pH 7.5), 0.5mM EDTA and 1M NaCl. Dynabeads had been pretreated with sodium hydroxide according to the manufacturer's instructions to minimize RNase contamination. The sample and bead mix was incubated at 37°C for 15 min followed by a 5 min step at 50°C before beads were collected on a magnet and the supernatant was transferred to a new tube with beads and the incubation steps were repeated. RNA collected in the supernatant from the second round of bead depletion was cleaned up by isopropanol precipitation.

To include both non-polyadenylated and polyadenylated transcripts, 500 ng of rRNA-depleted RNA was polyadenylated *in vitro* using *E. coli* poly(A) polymerase as outlined in[48]. As an alternative strategy, we ligated a DNA linker (/5rApp/(N)6CTGTAGGCACCATCAAT, IDT) to the 3 prime-end of total RNA as previously described[49]. Briefly, 1 ug of total RNA was denatured for 2 minutes at 80°C and incubated with 500 ng DNA linker, 8 uL 50% PEG8000, 2 uL DMSO, 2 uL of T4 RNA Ligase Buffer (NEB) and 1 uL T4 RNA Ligase 2 (made in-house) in a 20 uL reaction for 16 hours at 16°C. The RNA was purified with the Zymo RNA Clean & Concentrator Kit, including the optional removal of molecules < 200 nt to exclude the leftover linker. Direct RNA library preparation was performed using the kit SQK-RNA002 (Oxford Nanopore Technologies) with 500 ng of poly(A)-tailed or 3

prime-end ligated RNA according to manufacturer's instructions with the following exceptions: the RCS was omitted and replaced with 0.5 uL water and the ligation of the reverse transcription adapter (RTA) was performed for 15 minutes. For 3 prime-end ligated samples, the standard RTA was replaced by a custom adapter that was generated by annealing two oligonucleotides (IDT): /5PHOS/GGCTTCTTCTTGCTCTTAGGTAGTAGGTTC (ONT_oligoA) and GAGGCGAGCGGTCAATTTTCCTAAGAGCAAGAAGAAGCCGATTGATGGT (ONT_oligoB_linker), where the underlined bases are complementary to the DNA linker. Oligonucleotides were diluted to 1.4 μM in 10 mM Tris-HCl pH 7.5, 50 mM NaCl and annealed by heating to 95°C and slowly cooling to room temperature.

Poly(A)+ RNA from myoblasts and HeLa samples were purified using the Dynabeads mRNA purification kit (ThermoFisher, 61006) according to the manufacturer's instructions, starting with up to 75 ug of total RNA. Direct RNA library preparation was performed as described above with 500-700 ng of poly(A)+ RNA. Direct total poly(A)+ RNA sequencing data from K562 cells were obtained from[13]. Libraries were sequenced on a MinION device (Oxford Nanopore Technologies) for up to 72 hours.

Nanopore sequencing data analysis

Live base-calling of nanopore sequencing data was performed with MinKNOW (release 20.10.3 or later). Reads with a base-calling threshold > 7 were converted into DNA sequences by substituting U to T bases prior to alignment. Reads were aligned to the Hela-specific single nucleotide polymorphism-corrected reference hg38 genome (see below) using minimap2[50] with parameters -ax splice -uf -k14. Multi-mapping reads were included in all downstream analyses. All subsequent analyses were performed using custom python scripts.

For measuring transcript abundance, the alignment BAM file was converted to a BED file using pybedtools[51] bamtobed with options cigar=True, tag='NM'. Reads mapping to the mitochondrial genome were extracted using pybedtools intersect with options wo=True, s=True, and a BED file containing the coordinates of all mitochondrial transcripts. Read/transcript pairs that contained the

cigar string “N” within 25 nt of the transcript start or end were excluded. Unique reads intersecting each transcript by at least 25 nt were counted. For measuring the abundance of reads mapping to the last 100 nt of the transcript, the same strategy was used with a BED file containing the coordinates of the 3 prime-end most 100 nt for each transcript. For transcripts that are frequently detected in the unprocessed state (e.g. *ATP8-6/COX3*), the abundance of the upstream transcript was corrected to take into account reads that are not long enough to reach it. For each read, the genomic coordinates of the read start and the read end were extracted and intersected with the following genomic features within the transcripts of interest using pybedtools with options wo=True, s=True: the transcript start region (0 to +25 nt transcript start), the transcript end region (-25 to +25 nt from the transcript end), the gene body (+25 nt from transcript start to –25 nt from transcript end) and the 5 prime upstream region (-1500 to 0 nt from transcript start). For the downstream transcript (e.g. *COX3*), the 5 prime-end fraction unprocessed was calculated as the number of reads mapping to this transcript that start in the 5 prime upstream region divided by the number of reads mapping to this transcript that start in the 5 prime upstream region or in the transcript start region. Reads mapping to this transcript that start in the gene body were defined as those lacking information about the 5 prime-end and that potentially extend into the upstream transcript (e.g. *ATP8-6*). The abundance of the upstream transcript was corrected using the following formula: reads mapping to upstream transcript + (fraction unprocessed downstream transcript * number of reads lacking information about 5 prime-end downstream transcripts).

For analysis of processing, the genomic coordinates of the read start and the read end were extracted and intersected with the following genomic features using pybedtools with options wo=True, s=True: the transcript start region (-15 to +50 nt from transcript start), the transcript end region (-15 to +15 nt from transcript end) and the gene body (transcript start to transcript end). Next, the entire reads were intersected with the following genomic features using pybedtools with options wo=True, s=True: the upstream region (-60 to -15 nt from transcript start), the downstream region (+15 to +60 nt from transcript end), the transcript end region (-15 to +15 nt from transcript end) and the gene body (transcript start to transcript end). Reads/feature pairs that overlapped by less than 15

nt or that contained the cigar string “N” within 15 nt of the feature start or end were excluded. Reads that mapped to a protein-coding gene and that ended in the transcript end region of that gene were classified as “processed” at the 3 prime-end. Reads that mapped to a protein-coding gene and to the downstream region of that gene were classified as “unprocessed” at the 3 prime-end. Reads that mapped to a protein-coding gene and that started in the transcript start region of that gene were classified as “processed” at the 5 prime-end. Reads that mapped to a protein-coding gene and to the upstream region of that gene were classified as “unprocessed” at the 5 prime-end.

Poly(A) tail lengths were estimated using nanopolish[52]. For 3 prime-end ligated samples, reads were classified as non-polyadenylated if the qc_tag from nanopolish was “NO_REGION” or polyadenylated if the qc_tag was “PASS” or “ADAPTER”. For poly(A)+, only reads classified as polyadenylated were analyzed. Only reads with the 3 prime-end processing status “processed” were included in poly(A) tail length analyses.

For modeling transcription, the genomic coordinates of the read 3 prime-end were extracted. Using pybedtools with options wo=True, s=True, read ends were classified as “processed” if they intersected with a 25 nt window around the transcript end site or “actively transcribing” if they mapped more than 25 nt upstream of the transcript end site.

Quantitative gene expression models

To estimate the importance of each of the levels of gene expression that we had acquired kinetic measurements we developed a quantitative gene-expression model (Fig 5) that we defined as follows: $Protein\ synthesis\ (RPF) = \frac{mRNA\ k_{synthesis}}{mRNA\ k_{deg}} * Fraction\ mRNA\ in\ ribosome * TE$

Where the $mRNA\ k_{synthesis} = 1/(T_{initiation} + T_{elongation})$ from the transcription elongation model defined above, “fraction mRNA in ribosome” as defined above and the

$TE = RPF/(\frac{mRNA\ k_{synthesis}}{mRNA\ k_{deg}} * Fraction\ mRNA\ in\ ribosome)$.

The coefficient of determination from correlating the predicted RPF values from the model to that of measured RPF values was used to measure the predictive power of the model. The model per definition explains all of the variability in RPF. The importance of each gene-expression level was determined by setting the succeeding parameters to 1. E.g. to estimate the predictive value of RNA levels(= $\frac{mRNA\ k_{synthesis}}{mRNA\ k_{deg}}$) we set both $Fraction\ mRNA\ in\ ribosome\ and\ the\ TE$ to 1.

We also developed a simpler minimal gene-expression model to directly compare nuclear and mitochondrial gene-expression control where $Protein\ synthesis\ (RPF) = \frac{mRNA\ k_{synthesis}}{mRNA\ k_{deg}} * TE$ where the $mRNA\ k_{synthesis} = RNA\ levels/mRNA\ k_{deg}$ and $TE = RPF/mRNA\ levels$. RNA-levels were from direct RNA-seq (Fig 1) for mitochondria and TPMs from nuclear-encoded genes.

All the model parameters were estimated using measured rates and RPF data from HeLa cells and were kept constant except for the example of using the $mRNA\ k_{deg}$ from LRPPRC KO cells.

The predictive power was determined as the coefficient of determination from correlating the predicted RPF values from the model to that of measured RPF values, either from the same HeLa data set (Model) or a HeLa replicate (Replicate) or from other cell lines (HEK293T wt, HEK293T LRRPRC KO or K562 cells). The importance of each gene-expression node was again determined by sequentially setting the TE and then $mRNA\ k_{synthesis}$ parameters to 1. All rates, RPF, TPM and direct RNA-seq values can be found in the supplemental tables. For this analysis, we used averaged RPF, half-lives, and RNA abundances between two replicates for each cell line, except for the HeLa model and replicate. For the nuclear side we excluded two outliers, ATP5MC3 and ATP5F1E, out of 75 genes total as they single-handedly depressed the RPF correlation between the different cell lines. HeLa cyto-RPF data was from[53] (GEO accession number GSE69906), K562 mito- and cyto-RPF is from this work while all other RPF data is from[5] (Table S1).

References

1. Balakrishnan, R. *et al.* Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. *Science* **378**, eabk2066 (2022).
2. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
3. Isaac, R. S., McShane, E. & Churchman, L. S. The Multiple Levels of Mitonuclear Coregulation. *Annu. Rev. Genet.* **52**, 511–533 (2018).
4. Couvillion, M. T., Soto, I. C., Shipkovenska, G. & Churchman, L. S. Synchronized mitochondrial and cytosolic translation programs. *Nature* **533**, 499–503 (2016).
5. Soto, I. *et al.* Balanced mitochondrial and cytosolic translatomes underlie the biogenesis of human respiratory complexes. *Genome Biol.* **23**, 170 (2022).
6. Mercer, T. R. *et al.* The human mitochondrial transcriptome. *Cell* **146**, 645–658 (2011).
7. Köhler, F., Müller-Rischart, A. K., Conradt, B. & Rolland, S. G. The loss of LRPPRC function induces the mitochondrial unfolded protein response. *Aging* **7**, 701–717 (2015).
8. Suomalainen, A. & Battersby, B. J. Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat. Rev. Mol. Cell Biol.* (2017) doi:10.1038/nrm.2017.66.
9. Rackham, O. & Filipovska, A. Organization and expression of the mammalian mitochondrial genome. *Nat. Rev. Genet.* (2022) doi:10.1038/s41576-022-00480-x.
10. Tan, B. G. *et al.* The human mitochondrial genome contains a second light strand promoter. *Mol. Cell* **82**, 3646–3660.e9 (2022).
11. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
12. Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C. & Simon, M. D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* **15**, 221–225 (2018).
13. Smalec, B. M. *et al.* Genome-wide quantification of RNA flow across subcellular compartments

reveals determinants of the mammalian transcript life cycle. *bioRxiv* 2022.08.21.504696 (2022) doi:10.1101/2022.08.21.504696.

14. Gelfand, R. & Attardi, G. Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. *Mol. Cell. Biol.* **1**, 497–511 (1981).
15. Chujo, T. *et al.* LRPPRC/SLIRP suppresses PNPase-mediated mRNA decay and promotes polyadenylation in human mitochondria. *Nucleic Acids Res.* **40**, 8033–8047 (2012).
16. Piechota, J. *et al.* Differential stability of mitochondrial mRNA in HeLa cells. *Acta Biochim. Pol.* **53**, 157–168 (2006).
17. Burger, K. *et al.* 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* **10**, 1623–1630 (2013).
18. Jürges, C., Dölken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218–i226 (2018).
19. Wolf, A. R. & Mootha, V. K. Functional genomic analysis of human mitochondrial RNA processing. *Cell Rep.* **7**, 918–931 (2014).
20. Wu, Z. *et al.* Quantitative regulation of FLC via coordinated transcriptional initiation and elongation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 218–223 (2016).
21. Ietswaart, R., Rosa, S., Wu, Z., Dean, C. & Howard, M. Cell-Size-Dependent Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-to-Cell Expression Variation. *Cell Syst* **4**, 622–635.e9 (2017).
22. Ruzzenente, B. *et al.* LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J.* **31**, 443–456 (2012).
23. Singh, V., Itoh, Y., Huynen, M. A. & Amunts, A. Activation mechanism of mitochondrial translation by LRPPRC-SLIRP. *bioRxiv* 2022.06.20.496763 (2022) doi:10.1101/2022.06.20.496763.
24. Sultana, S., Solotchi, M., Ramachandran, A. & Patel, S. S. Transcriptional fidelities of human mitochondrial POLRMT, yeast mitochondrial Rpo41, and phage T7 single-subunit RNA

polymerases. *J. Biol. Chem.* **292**, 18145–18160 (2017).

25. Yu, H. *et al.* TEFM Enhances Transcription Elongation by Modifying mtRNAP Pausing Dynamics. *Biophys. J.* **115**, 2295–2300 (2018).
26. Posse, V., Shahzad, S., Falkenberg, M., Hällberg, B. M. & Gustafsson, C. M. TEFM is a potent stimulator of mitochondrial transcription elongation in vitro. *Nucleic Acids Res.* **43**, 2615–2624 (2015).
27. Costello, M. *et al.* Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* **19**, 332 (2018).
28. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470 (1981).
29. Temperley, R. J., Wydro, M., Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. Human mitochondrial mRNAs--like members of all families, similar but different. *Biochim. Biophys. Acta* **1797**, 1081–1085 (2010).
30. Rackham, O., Mercer, T. R. & Filipovska, A. The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression. *Wiley Interdiscip. Rev. RNA* **3**, 675–695 (2012).
31. Bratic, A. *et al.* Mitochondrial Polyadenylation Is a One-Step Process Required for mRNA Integrity and tRNA Maturation. *PLoS Genet.* **12**, e1006028 (2016).
32. Remes, C. *et al.* Translation initiation of leaderless and polycistronic transcripts in mammalian mitochondria. *Nucleic Acids Res.* (2023) doi:10.1093/nar/gkac1233.
33. Kühl, I. *et al.* POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. *Sci Adv* **2**, e1600963 (2016).
34. Bogenhagen, D. F., Ostermeyer-Fay, A. G., Haley, J. D. & Garcia-Diaz, M. Kinetics and Mechanism of Mammalian Mitochondrial Ribosome Assembly. *Cell Rep.* **22**, 1935–1944 (2018).
35. Ng, K. Y. *et al.* Nonstop mRNAs generate a ground state of mitochondrial gene expression noise. *Sci Adv* **8**, eabq5234 (2022).
36. Irastortza-Olaziregi, M. & Amster-Choder, O. Coupled Transcription-Translation in Prokaryotes:

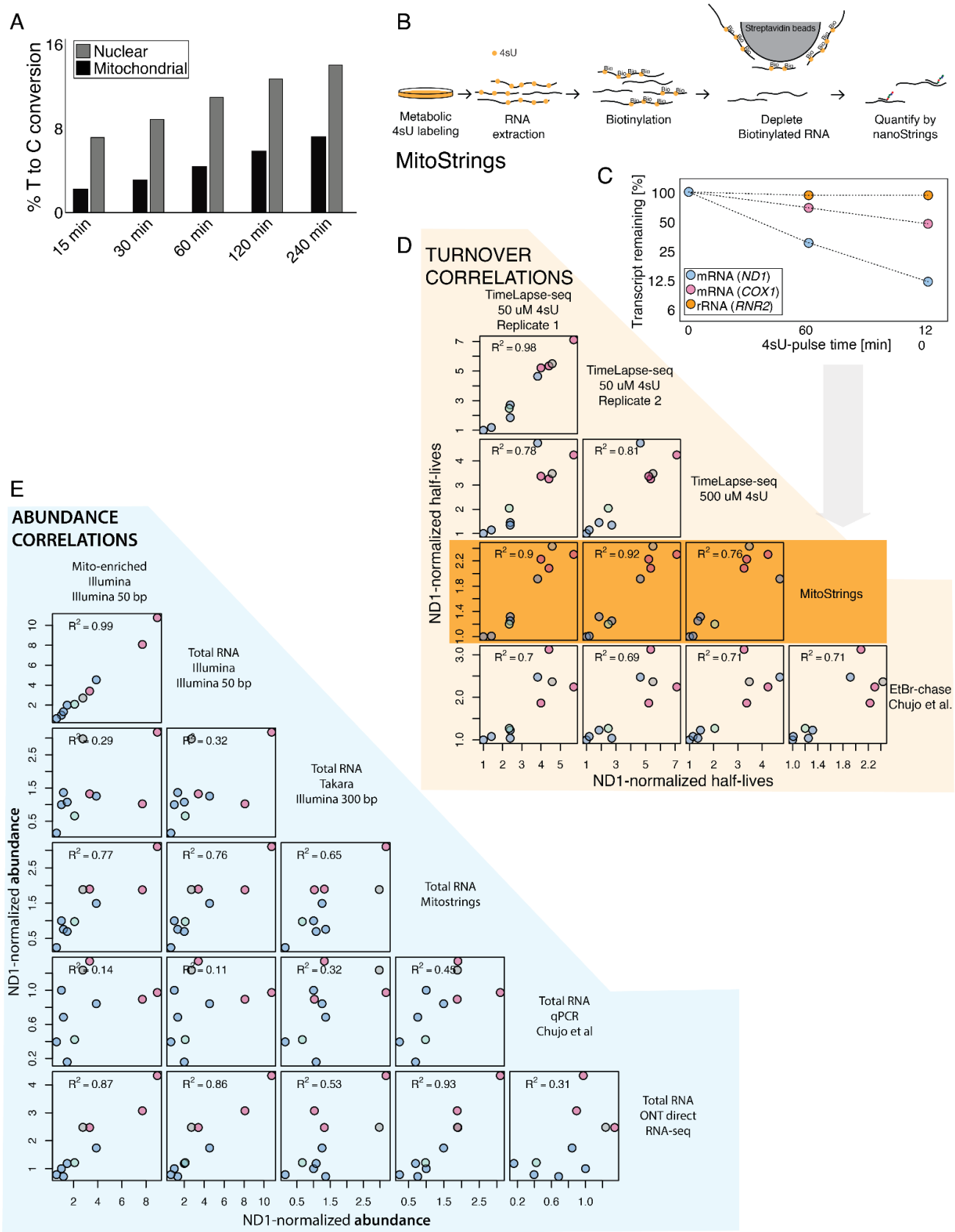
An Old Couple With New Surprises. *Front. Microbiol.* **11**, 624830 (2020).

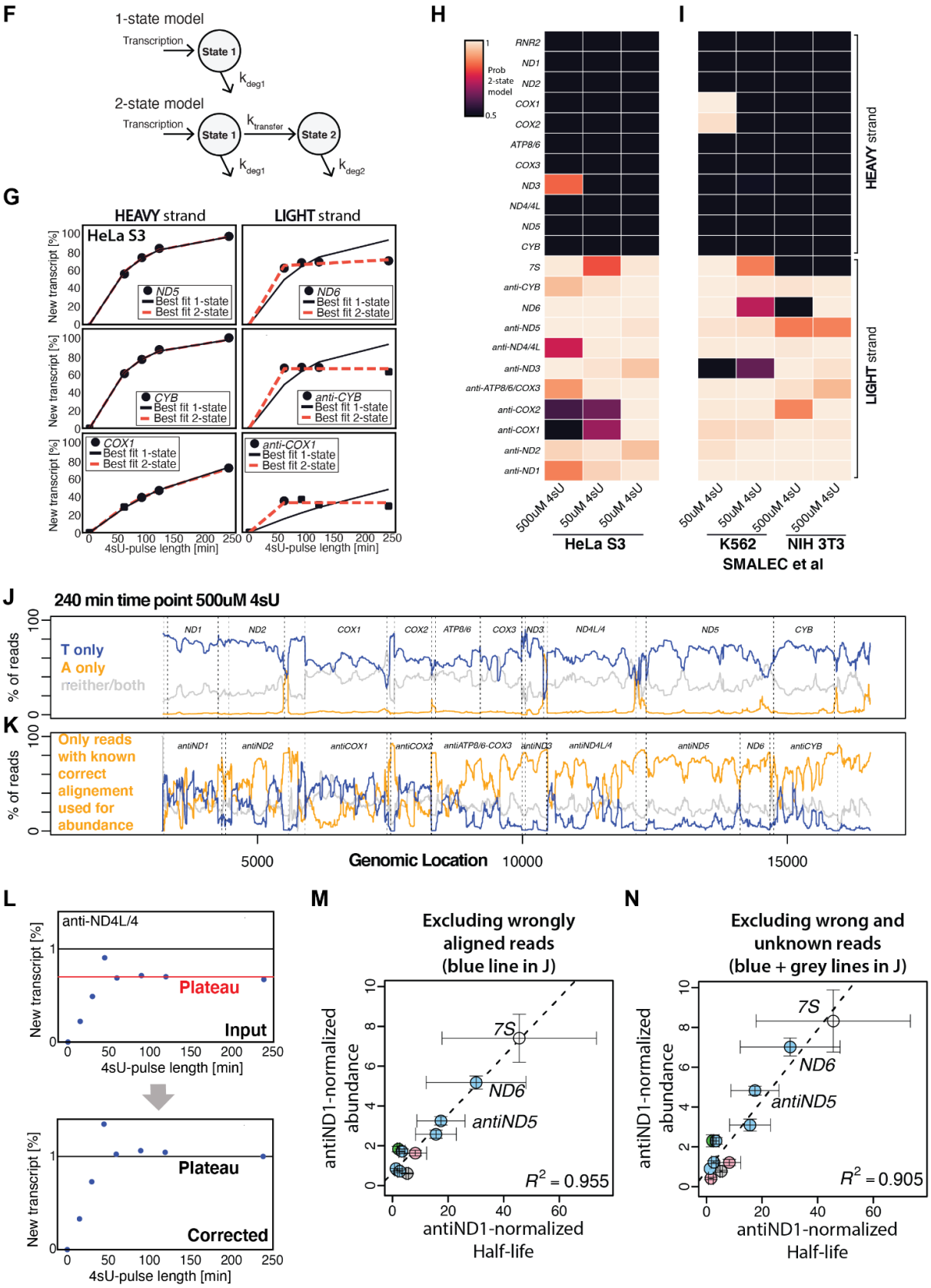
37. Lagouge, M. *et al.* SLIRP Regulates the Rate of Mitochondrial Protein Synthesis and Protects LRPPRC from Degradation. *PLoS Genet.* **11**, e1005423 (2015).
38. Jiang, S. *et al.* TEFM regulates both transcription elongation and RNA processing in mitochondria. *EMBO Rep.* (2019) doi:10.15252/embr.201948101.
39. Webster, M. W. *et al.* Structural basis of transcription-translation coupling and collision in bacteria. *Science* **369**, 1355–1359 (2020).
40. Kummer, E. *et al.* Unique features of mammalian mitochondrial translation initiation revealed by cryo-EM. *Nature* **560**, 263–267 (2018).
41. Ohkubo, A. *et al.* The FASTK family proteins fine-tune mitochondrial RNA processing. *PLoS Genet.* **17**, e1009873 (2021).
42. Morgenstern, M. *et al.* Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metab.* (2021) doi:10.1016/j.cmet.2021.11.001.
43. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).
44. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
45. Sin, C., Chiarugi, D. & Valleriani, A. Degradation Parameters from Pulse-Chase Experiments. *PLoS One* **11**, e0155028 (2016).
46. Shalgi, R. *et al.* Widespread regulation of translation by elongation pausing in heat shock. *Mol. Cell* **49**, 439–452 (2013).
47. Stefan Isaac, R. *et al.* Single-nucleoid architecture reveals heterogeneous packaging of mitochondrial DNA. *bioRxiv* 2022.09.25.509398 (2022) doi:10.1101/2022.09.25.509398.
48. Drexler, H. L. *et al.* Revealing nascent RNA processing dynamics with nano-COP. *Nat. Protoc.* **16**, 1343–1375 (2021).
49. Mayer, A. & Churchman, L. S. Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat. Protoc.*

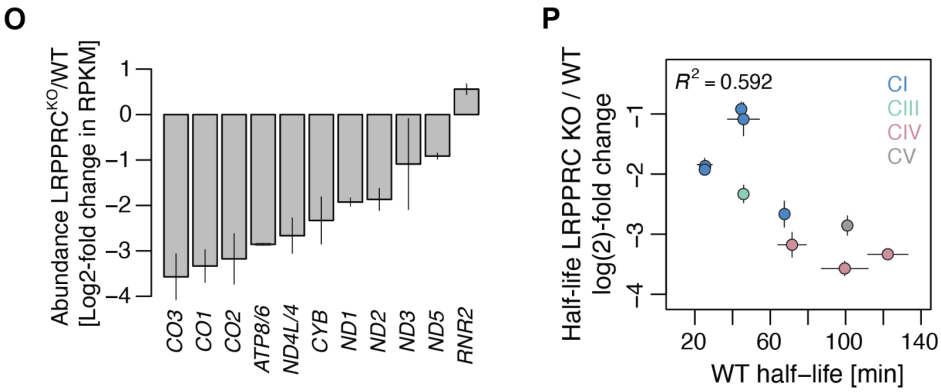
11, 813–833 (2016).

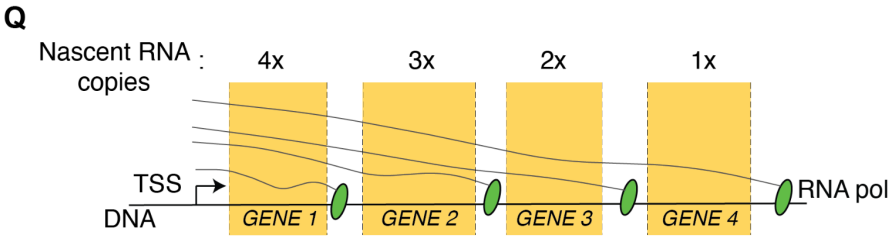
50. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
51. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
52. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
53. Wu, C. C.-C., Zinshteyn, B., Wehner, K. A. & Green, R. High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Mol. Cell* **73**, 959–970.e5 (2019).

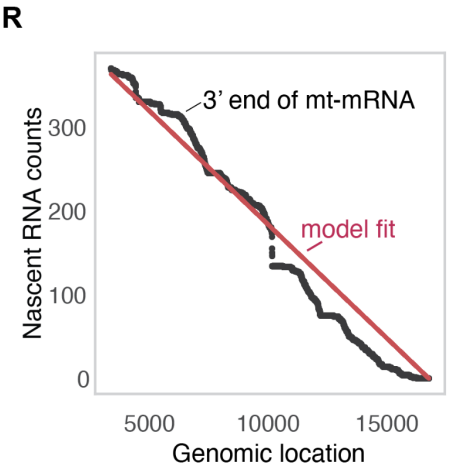
Supplemental figures



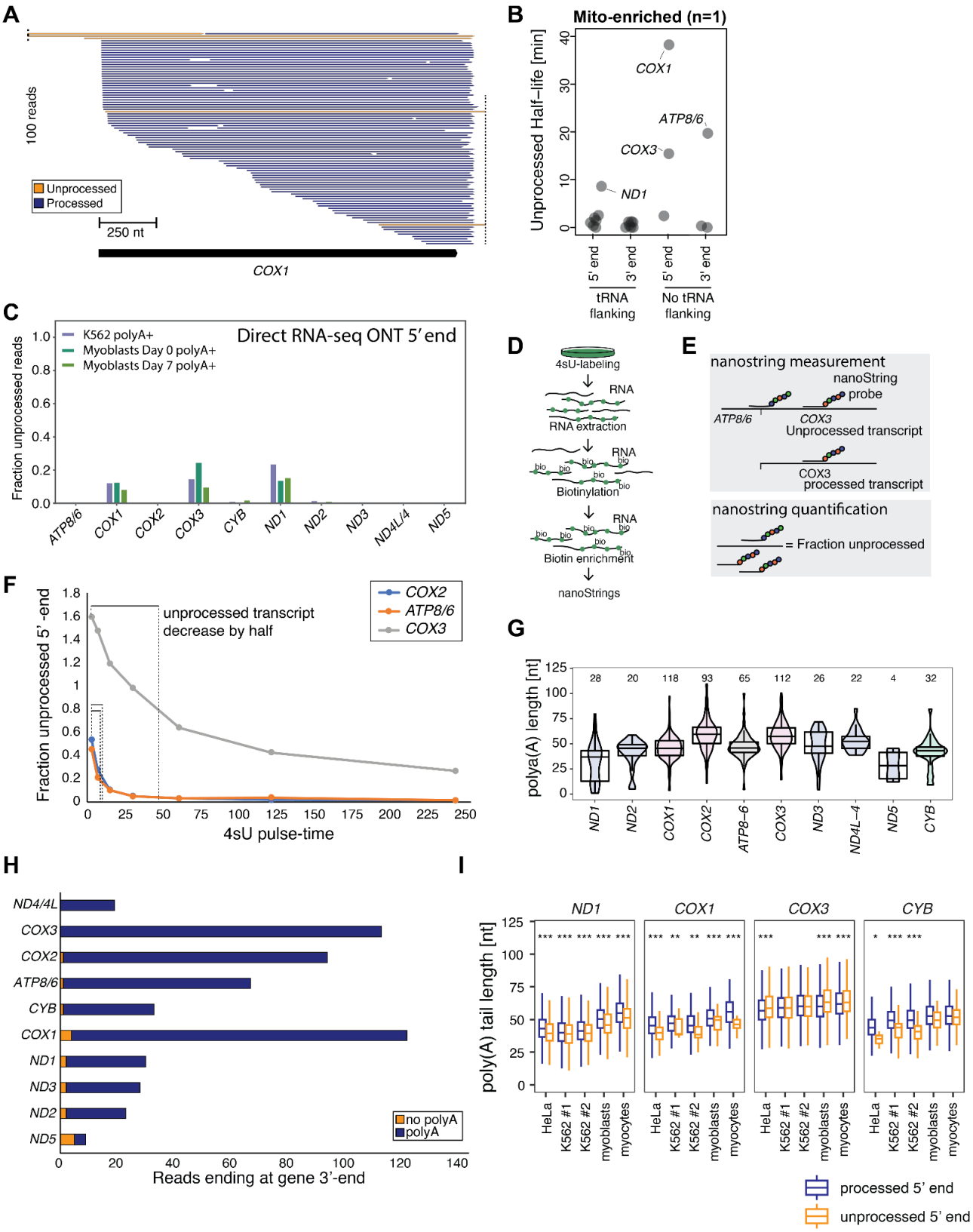

F
1-state model
Transcription
State 1
k_{deg1}
2-state model
Transcription
State 1
$k_{transfer}$
State 2
k_{deg1}
k_{deg2}
G
HEAVY strand
LIGHT strand
HeLa S3
New transcript [%]
ND5
Best fit 1-state
Best fit 2-state
ND6
CYB
anti-CYB
COX1
anti-COX1
4sU-pulse length [min]
H
I
Prob 2-state model
RNR2
ND1
ND2
COX1
COX2
ATP8/6
COX3
ND3
ND4/4L
ND5
CYB
7S
anti-CYB
ND6
anti-ND5
anti-ND4/4L
anti-ND3
anti-ATP8/6/COX3
anti-COX2
anti-COX1
anti-ND2
anti-ND1
HEAVY strand
LIGHT strand
500uM 4sU
50uM 4sU
HeLa S3
K562
NIH 3T3
SMALEC et al
J
240 min time point 500uM 4sU
% of reads
T only
A only
neither/both
ND1
ND2
COX1
COX2
ATP8/6
COX3
ND3
ND4L/4
ND5
CYB
K
Only reads with known correct alignement used for abundance
antiND1
antiND2
antiCOX1
antiCOX2
antiATP8/6-COX3
antiND3
antiND4L/4
antiND5
ND6
antiCYB
Genomic Location
5000
10000
15000
L
anti-ND4L/4
Plateau
Input
Corrected
M
Excluding wrongly aligned reads (blue line in J)
antiND1-normalized abundance
antiND1-normalized Half-life
7S
ND6
antiND5
$R^2 = 0.955$
N
Excluding wrong and unknown reads (blue + grey lines in J)
$R^2 = 0.905$


O
Abundance LRPPRC^KO/WT
[Log2-fold change in RPKM]
1
0
−1
−2
−3
−4
CO3
CO1
CO2
ATP8/6
ND4L/4
CYB
ND1
ND2
ND3
ND5
RNR2
P
Half-life LRPPRC KO / WT
log(2)-fold change
R² = 0.592
CI
CIII
CIV
CV
−1
−2
−3
−4
20
60
100
140
WT half−life [min]

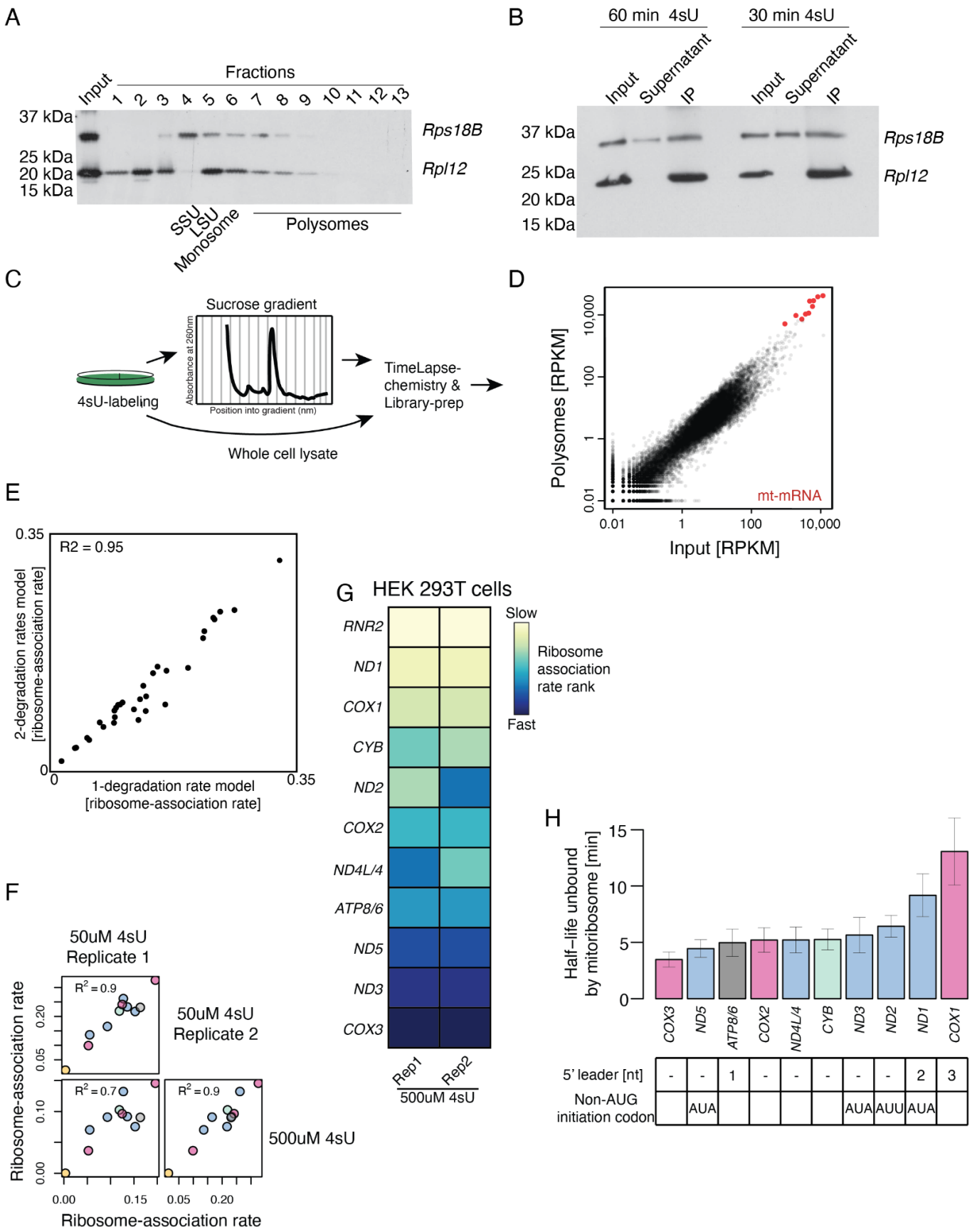

Q
Nascent RNA copies :
4x
3x
2x
1x
TSS
DNA
GENE 1
GENE 2
GENE 3
GENE 4
RNA pol


R
Nascent RNA counts
300
200
100
0
3' end of mt-mRNA
model fit
5000
10000
15000
Genomic location

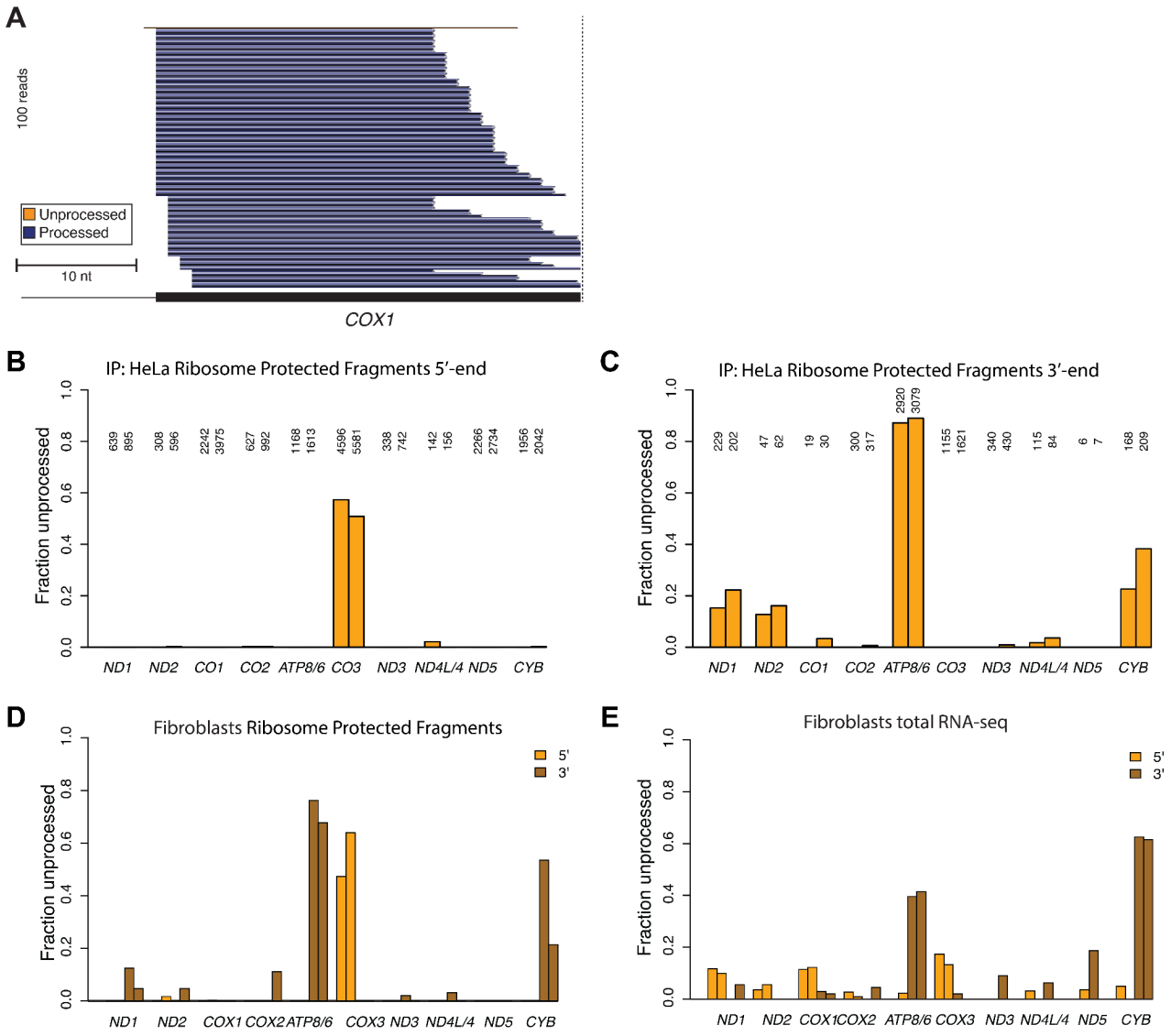
Supplemental Figure 1. A) Barplot showing the fraction T's that have been converted to C's by Time-Lapse chemistry and reverse transcription. Nuclear-encoded transcripts consistently have a higher conversion rate than mitochondrial-encoded transcripts in the same experiment and time point. B) 4sU labeled RNA is biotinylated and then removed after incubation with streptavidin beads leaving only pre-existing "old" RNA. The decrease in native/unlabeled RNA over time can be measured by the MitoStrings assay (see material and methods). C) Representative example profiles from one of three biological replicates. D) Correlation between turnover measurement using MitoStrings (from C), with TimeLapse-seq using 50 or 500uM 4sU or previously published data from HeLa cells deploying ethidium bromide (EtBr) chase experiments with qPCR as a readout[15]. E) Correlation between abundance estimates using different readout methods. F) We deployed a 2-state Markov model to detect RNA with turnover profiles that significantly differed from a simple exponential model. G) Example profiles from light and heavy strand transcripts. Light strand transcript can be seen to plateau before reaching maximum turnover. H) Systematic analysis of the occurrence of a plateau shows that this behavior is consistent but limited to the light-strand encoded transcripts. I) The same behavior can be seen in previously published data sets with RNA from different cell lines deploying different indexing and sequencing chemistries but the same library prep kit[13]. J and K) The plateau coincides with a significant amount of reads aligning to the wrong strand as judged by the 4sU-induced mismatches showing up at the wrong strand. The heavy strand transcriptome J) displays very few incorrectly aligned reads (A to G mismatches - orange line) while the light strand K) shows significant amounts of misaligned reads (T to C mismatches - blue line). L) Light strand turnover profiles can be corrected by normalizing each profile so that the last time point is set to 100% new. This moves the plateau up to where we would expect it to be in case the plateau reflects a fully turned-over transcript (bottom panel). The difference between the top and bottom panels is assumed to be caused by wrongly aligned reads that cannot contain the expected mismatches caused by 4sU. M and N) Using corrected turnover profiles (as in L) to calculate half-lives as well as corrected abundance estimates based on all reads aligning to the light strand minus the known wrongly aligned reads (M, blue line in K) or known and unknown (N, we cannot determine if reads lacking a 4sU induced mismatch (gray line in K) are aligned to the correct strand or not). Shown is the average of *antiND1* normalized values and error bars show standard deviations. O) Bar plot showing that most heavy-strand encoded mt-mRNA have significantly decreased steady-state levels in the absence of LRPPRC. Error bars standard deviation from two replicates. P) Scatter plot showing the negative correlation between mRNA half-lives in HEK293T wild-type cells and change in stability after LRPPRC knockout. The most stable mRNAs are most affected by the loss of LRPPRC. The Pearson R^2 is shown. Q) The nascent (actively transcribed) transcriptome has a TSS-proximal gene bias because, at any one point in time, more RNA polymerases will have traversed proximal genes than distal genes. As the mitochondrial transcriptome is rapidly turned over a significant fraction of all mt-mRNA are nascent and thus this bias could impact overall RNA levels. R) transcription elongation rates were estimated based on the slope of 3 prime ends of mt-mRNA that are actively transcribed. The black line shows reads from nanopore direct RNA sequencing that are considered nascent RNA as their 3 prime ends align to the gene bodies as opposed to the 3 prime cleavage sites across the heavy strand. The red line shows the model fit estimating transcription elongation rate (see material and methods).


A
100 reads
Unprocessed
Processed
250 nt
COX1
B
Mito-enriched (n=1)
Unprocessed Half–life [min]
COX1
ATP8/6
COX3
ND1
5' end
3' end
5' end
3' end
tRNA flanking
No tRNA flanking
C
Direct RNA-seq ONT 5′ end
K562 polyA+
Myoblasts Day 0 polyA+
Myoblasts Day 7 polyA+
Fraction unprocessed reads
ATP8/6
COX1
COX2
COX3
CYB
ND1
ND2
ND3
ND4L/4
ND5
D
4sU-labeling
RNA
RNA extraction
Biotinylation
Biotin enrichment
nanoStrings
E
nanostring measurement
nanoString probe
ATP8/6
COX3
Unprocessed transcript
COX3 processed transcript
nanostring quantification
= Fraction unprocessed
F
Fraction unprocessed 5' -end
unprocessed transcript decrease by half
COX2
ATP8/6
COX3
4sU pulse-time
G
polya(A) length [nt]
28
20
118
93
65
112
26
22
4
32
ND1
ND2
COX1
COX2
ATP8–6
COX3
ND3
ND4L–4
ND5
CYB
H
ND4/4L
COX3
COX2
ATP8/6
CYB
COX1
ND1
ND3
ND2
ND5
no polyA
polyA
Reads ending at gene 3'-end
I
ND1
COX1
COX3
CYB
poly(A) tail length [nt]
HeLa
K562 #1
K562 #2
myoblasts
myocytes
processed 5' end
unprocessed 5' end

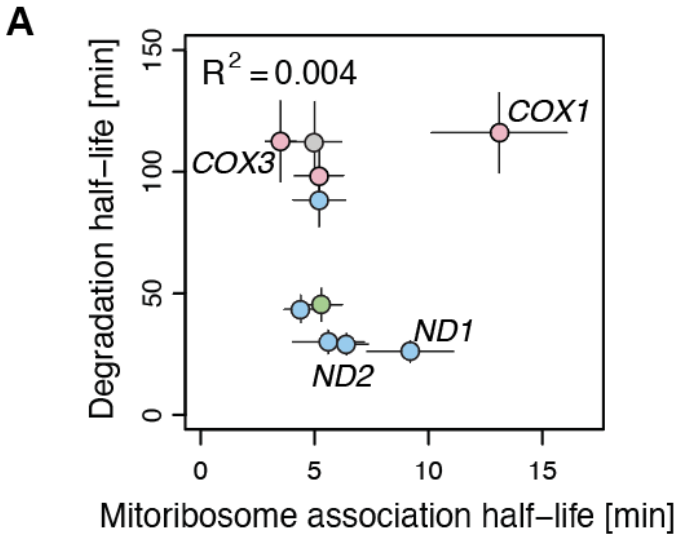
Supplemental Figure 2. A) 100 randomly sampled reads from nanopore direct RNA-seq that align to *COX1*. Orange reads are unprocessed at the 5 prime and/or the 3 prime end, and blue reads are processed or undetermined in cases where the read does not reach the end of the transcript. B) The amount of time required until half of the newly synthesized transcripts have been processed at a specific site using nanopore direct RNA-seq data from a mitochondria-enriched sample. Processing sites are subdivided into 3 prime or 5 prime ends of the mRNA as well as into groups of canonical (tRNA-flanking) or non-canonical. C) Bar plot showing the fraction of unprocessed 5 prime ends in poly(A)-selected RNA sequenced by nanopore direct RNA-seq. Data is from K562 cells and human myoblasts before or after 7 days of differentiation towards myotubes. D) Cartoon showing the experimental procedure for the data shown in E). E) MitoStrings can be used to estimate the fraction of unprocessed transcript by applying probes that bind either gene internally or overlap a processing site. F) Data for three processing junctions using the experimental design as shown in D) and analyzed as in E). G) Poly(A) tail length from nanopore direct RNA-seq experiment from total rRNA depleted RNA. H) All reads that end at the 3 prime end of a gene, from the same experiment as in G), and whether they have been polyadenylated or not. I) Poly(A) tail length as a function of 5 prime end processing status for four cell lines. All data in the figure are from HeLa cells if not stated otherwise.


A
Fractions
Input
1 2 3 4 5 6 7 8 9 10 11 12 13
37 kDa
25 kDa
20 kDa
15 kDa
Rps18B
Rpl12
SSU
LSU
Monosome
Polysomes
B
60 min 4sU
30 min 4sU
Input
Supernatant
IP
Input
Supernatant
IP
37 kDa
25 kDa
20 kDa
15 kDa
Rps18B
Rpl12
C
Sucrose gradient
Absorbance at 260nm
Position into gradient (nm)
4sU-labeling
TimeLapse-chemistry & Library-prep
Whole cell lysate
D
Polysomes [RPKM]
Input [RPKM]
0.01
1
100
10,000
mt-mRNA
E
R2 = 0.95
2-degradation rates model [ribosome-association rate]
1-degradation rate model [ribosome-association rate]
0
0.35
G
HEK 293T cells
RNR2
ND1
COX1
CYB
ND2
COX2
ND4L/4
ATP8/6
ND5
ND3
COX3
Rep1
Rep2
500uM 4sU
Slow
Ribosome association rate rank
Fast
F
50uM 4sU Replicate 1
50uM 4sU Replicate 2
500uM 4sU
$R^2 = 0.9$
$R^2 = 0.7$
$R^2 = 0.9$
Ribosome-association rate
0.00
0.05
0.10
0.15
0.20
H
Half-life unbound by mitoribosome [min]
0
5
10
15
COX3
ND5
ATP8/6
COX2
ND4L/4
CYB
ND3
ND2
ND1
COX1
5' leader [nt]
- - 1 - - - - - 2 3
Non-AUG initiation codon
AUA AUA AUU AUA

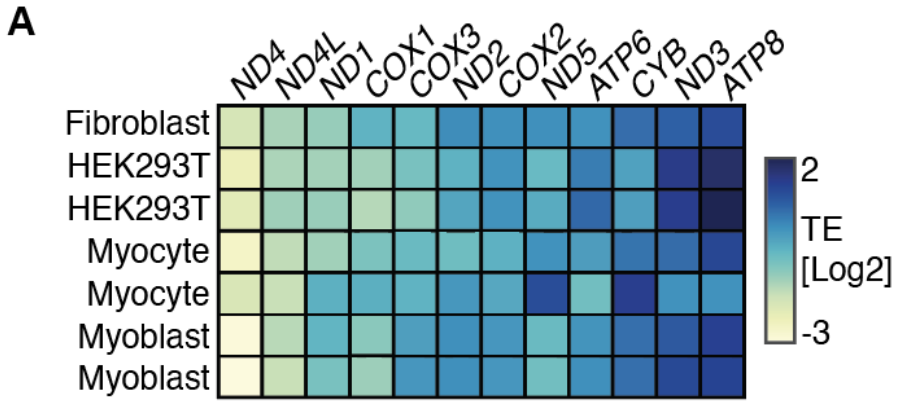
Supplemental Figure 3. A) Western blot of two ribosomal proteins, one small subunit (*Rps18B*) and one large subunit (Rpl12), from proteins extracted out of a sucrose gradient such as in C). The fractions where the majority of the small ribosomal subunit (SSU), large ribosomal subunit (LSU), monosome and polysome reside have been annotated. B) Western Blot for the same ribosomal proteins in A) but here showing the result of immunoprecipitation of the ribosome out of pooled gradient fractions from two 4sU labeling time points. All of Rpl12 (the bait) from the input comes down in the IP fraction and none remains in the supernatant. For Rps18B the majority comes down in the IP while a fraction remains in the supernatant. C) Cartoon of the experimental design using only sucrose gradient to enrich for mitochondrial ribosome-associated RNA. D) Results from the experiment outlined in C). RPKMs from libraries made from RNA in the input or extracted from pooled sucrose gradient fractions (fractions 6-13) show that there is little enrichment for mt-mRNA or mt-rRNA overall cellular mRNA or antisense (non-translated) mt-transcripts. E) Scatter plot comparing the ribosome-association rates (from all three replicates) using either the simpler mathematical model assuming the same degradation rate for both the ribosome-associated RNA and the unbound RNA (1-degradation rate model) or a slightly more complex model allowing for independent degradation rates for the two states (2-degradation rate model, see material and methods). Both models result in similar estimates of the ribosome association rates. F) Same as in Fig 3E) but showing the correlation between ribosome association rates. G) Heatmap showing the transcripts ranked by their ribosome-association half-lives in each replicate using HEK293T cells. H) Barplot showing the half-life of mt-mRNA until they are associated with the mitoribosome bound. Transcripts 5 prime characteristics are listed below the plot. All data in the figure are from HeLa cells if not stated otherwise.



Supplemental Figure 4. A) Subsampling of 100 ribosome-protected fragments aligned to the 5 prime end of COX1. The scale bar shows the distance in nucleotides and the color code refers to if the reads have been called processed or not. Only 1 read is unprocessed for COX1 in the ribosome. B) Quantification of all ribosome-protected fragments that align to the 5 prime end of mitochondrial genes from immunoprecipitated ribosomes[5]. The plot shows the fraction of the reads that are unprocessed (i.e. are orange in A)) over all reads aligning to the 5 prime end of the gene (I.e. blue and orange reads in A)). C) Same as in B) but for 3 prime ends. When compared to the non-immunoprecipitated RPF from HeLa cells in figure 4, there are no major differences. D) Quantification of all RPF, from primary human fibroblasts[5], that align to either the 5 prime or 3 prime end of a mitochondrial gene. Same analysis as in B) and C) but combined into one plot. E) Illumina short-read sequencing of rRNA-depleted total RNA from the same human fibroblasts as in D). Comparing E) and D) it is clear that unprocessed 5 prime (but not 3 prime) ends found in the total RNA are depleted in the ribosome-protected RNA.



Supplemental Figure 5. A) Scatter plot showing the lack of correlation between RNA turnover and ribosome association half-lives.



Supplemental Figure 6. A) Heatmap of translation efficiency (TE) values from different cell lines measured in[5].