Leveraging genomic approaches to characterize mitochondrial RNA biology

Leveraging genomic approaches to characterize mitochondrial RNA biology

A dissertation presented

by
Ashley Robin Wolf

to
The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

in the subject of
Systems Biology

Harvard University
Cambridge, Massachusetts

April 2014

Dissertation advisor: Professor Vamsi K. Mootha Ashley Robin Wolf

Leveraging genomic approaches to characterize mitochondrial RNA biology

Abstract

Transcription and translation of mammalian mitochondrial DNA (mtDNA) occurs within the mitochondrial matrix to produce oxidative phosphorylation subunits required for efficient energy production. These mtDNA-encoded subunits complex with mitochondrial-localized, nuclear-encoded subunits to form the respiratory chain, and aberrant production or function of these subunits can cause devastating human disease. In addition to 13 oxidative phosphorylation subunits, mtDNA encodes 2 rRNAs and 22 tRNAs. All proteins required for mitochondrial RNA transcription, processing, and translation are encoded in the nucleus and translocated into the mitochondria. Here, I characterize over 100 nuclear-encoded mitochondrial proteins with predicted RNA-binding domains. Using RNAi and an RNA profiling approach, MitoString, we further characterize previously identified RNA processing factors and identify the novel regulator FASTKD4, which influences the abundance of a subset of mitochondrial mRNAs. Next, we apply knowledge of the RNA degradation component SUPV3L1 gleaned from our RNAi studies and previous research to test whether a specific set of variants influence the function of this gene in patient fibroblasts. Using MitoString, we find no evidence of pathogenicity of these variants in our fibroblast model. Our approach highlights the value of a thorough understanding of mitochondrial proteins and the

necessity of experimental techniques to validate the effect of variants found in exome-sequencing studies. Finally, we take an unbiased approach to characterizing the mitochondrial transcriptome of mouse liver by sequencing RNA from sequentially enriched mitochondrial fractions. Although we find an abundance of nuclear-encoded *5S* rRNA, consistent with previous research, we fail to identify any imported nuclear-encoded tRNAs. Uniting genomics, biochemistry, and medicine, these findings advance our understanding of mitochondrial RNA biology.

Table of Contents

Acknowledgements

A PhD is a journey, and I am indebted to the incredible support network I have enjoyed along the way. Vamsi has been an enthusiastic and rigorous scientific advisor, as well as a considerate and accessible person. From our first meeting, I knew that his lab would be an excellent choice and this many years later I still agree. The incredible team of researchers he has assembled has made the Mootha lab an exciting place to be. The entire lab has played a role in my education, but in particular I want to thank Olga, for keeping everything running smoothly, Josh, for keeping me laughing and focused, and Robert for his insights and never sugarcoating the truth. Danny and Steve took the plunge into the lab alongside me, and even if they escaped a year earlier, were key comrades in the journey, as was Laura, who has kept me sane and motivated in our final year as graduate students.

I want to thank my committee members, John Rinn, Chris Burge and Mike Springer, for their enthusiasm and valuable suggestions. I appreciate the time each of them dedicated, and the attitude they brought made each occasion constructive and fun.

The Systems Biology PhD program is a special place, and the many students and faculty I have interacted with have inspired me in many ways. In particular, I want to thank the students in my cohort, who have always been forthright and supportive and are now close friends. I also want to thank Galit Lahav and Angela Depace for starting the Women in Systems Biology group, which led to many enlightening discussions and strengthened friendships. Pam Silver, who initially chaired the program, welcomed me into graduate school, and Tim Mitchison and Andrew Murray have since taken up the

mantle, chairing the program with spunk and inspiring my greatest artistic achievement. Finally, I want to thank Sam Reed, whose tireless efforts make our lives so much easier.

Beginning with my high school biology teacher, John Brick, I have been blessed with enthusiastic educators who have encouraged my growth as a scientist. David Botstein, Leonid Kruglyak, Josh Rabinowitz, Manuel Llinas, Coleen Murphy, and Amy Caudy are some of the faculty at Princeton whose passion for science and unwavering encouragement led me to pursue a PhD. Prior to graduate school, I was welcomed into the laboratories of Antonio Planchart, James Allison, and Bonnie Bassler, who each treated me with respect and exposed me to environments that that prepared me for graduate research. Rotations in graduate school with John Rinn, Pardis Sabeti, and Brad Bernstein were also incredible learning experiences that shaped me as a scientist.

Outside of lab, I have been blessed to reside with three of my best friends for over five years. Emily, Julia, and Sarah have been with me every step of the way. As we moved up from New Jersey, were all awarded NSF fellowships on the same day, passed qualifying exams, went on picnics, taught undergraduates, baked endless cookies, wrote papers, defended theses, took day trips to the beach, and searched for postdocs, their support and companionship has been priceless. My other Boston-based friends have also contributed to the joy of living in this wonderful city.

I started graduate school alongside my boyfriend, Max, who, in addition to being a first-rate scientist, is my staunchest supporter and combines all the good qualities of a gentleman with the principles of a feminist. He has patiently stood by me through the good and the bad.

My mother, Robin, introduced me to the fascinating world of genetics and encouraged my interests in math and science. My father, Rodger, set a high bar with his legendary tales of acing math tests. Each of them has always believed in my ability to accomplish anything. My sisters Chelsea, Briana, and Heather have always simultaneously supported me and reminded me that there is more to life than academics. Chelsea has kept me laughing through our joint thesis-writing journey, and commiserated with my pain. My large network of extended family has also supported my journey, including my Boston-area aunts, Amy and Linda. My Grandmother Janet Wolf, despite refusing to comprehend what mitochondria are, is coming to my defense and has inspired me through frequent letters and cookie mailings. Finally, the battles of my Uncles Tom and Rick, and my Grandmother Joan Brubaker, have served as persistent reminders of what scientific research is ultimately for.

Chapter 1
Introduction

Overview

Over 1.5 billion years ago, the mitochondrion's prokaryotic ancestor invaded the proto-eukaryotic cell, ushering in DNA encoding rRNAs, tRNAs, and mRNAs necessary for the energy-generating process we know as oxidative phosphorylation (Sicheritz-Pontén et al., 1998). As oxidative phosphorylation (OXPHOS) became indispensible to its host and the mitochondrion lost autonomy, many of these genes transferred to the nuclear genome and were lost from mitochondrial DNA (mtDNA). In mammalian mitochondria, 13 essential subunits of the respiratory chain remain encoded by mtDNA (Anderson et al., 1981) (Figure 1-1). These mtDNA-encoded genes require nuclear-encoded proteins, of both prokaryotic ancestry and eukaryotic innovation, for their transcription, processing, and translation. Without proper production of these mtDNA-encoded subunits, OXPHOS malfunctions, triggering devastating human diseases.

OXPHOS is the process by which mitochondria synthesize ATP by harvesting the energy from a proton gradient formed by electron transfer through the respiratory chain. Complexes I-IV couple electron transfer to oxygen with formation of a proton gradient, which then powers ATP production by ATP synthase (Complex V). Excepting Complex II, each of these complexes is composed of proteins encoded by both the nuclear and mitochondrial genomes. Nuclear proteins are translocated into the mitochondria, while the mtDNA transcripts are translated by mitochondrial ribosomes within the mitochondrial matrix.

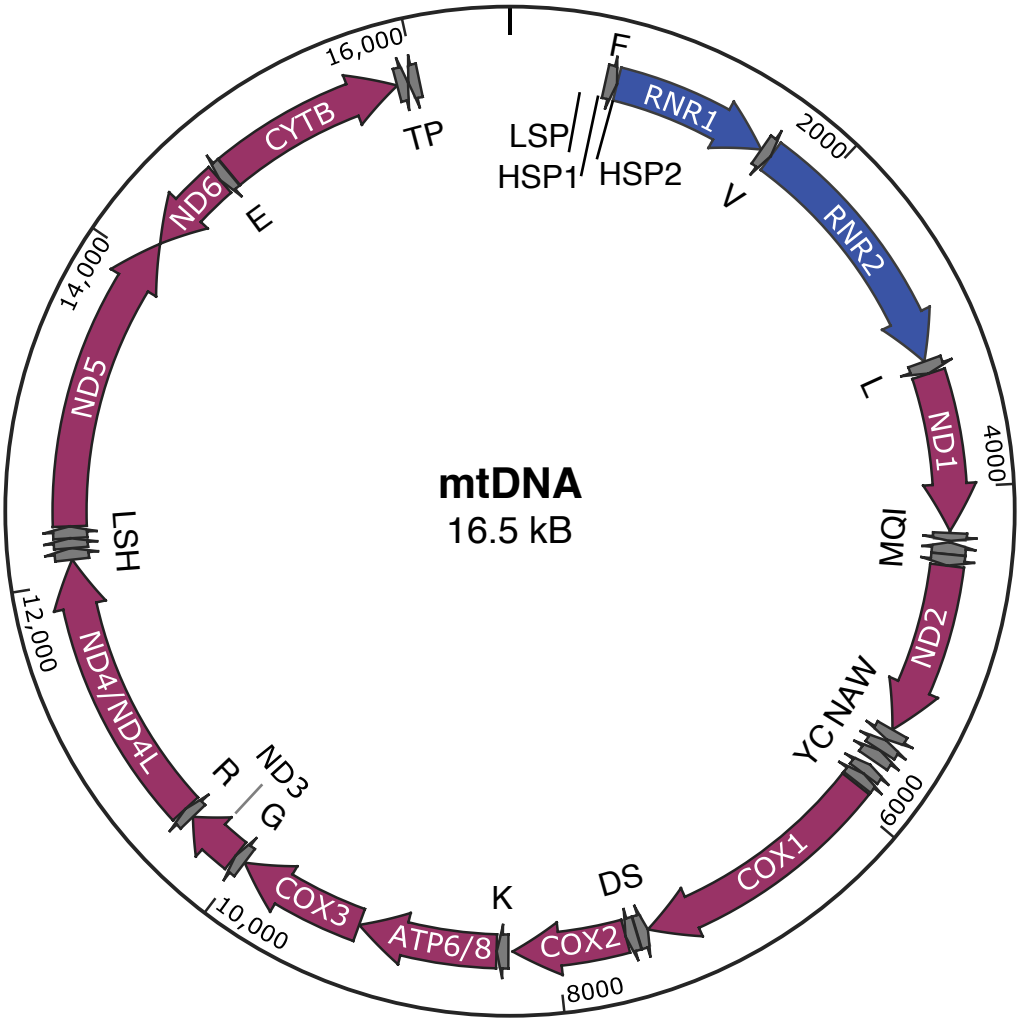


Figure 1-1. Human mtDNA encodes 13 proteins, 22 tRNAs, and 2 rRNAs
The 16.5 kB compact circular mitochondrial genome is shown with rRNAs depicted in blue, mRNAs in purple, and tRNAs in grey (labeled by their 1 letter abbreviations). The genome is transcribed from three promoters: LSP (light strand promoter) initiates transcription of ND6 and the light strand tRNAs; HSP1 (heavy strand promoter 1) transcribes the two rRNAs and the phenylalanine and valine tRNAs; and HSP2 (heavy strand promoter 2) initiates within the phenylalanine tRNA and transcribes the remainder of the heavy strand.

In addition to OXPHOS subunits, roughly 1,100 nuclear-encoded proteins are mitochondrial-localized; the identity of these proteins has been collated into a list termed MitoCarta (Pagliarini et al., 2008). Among these translocated proteins are metabolic enzymes, components of the mtDNA nucleoid, and various RNA processing proteins. The MitoCarta compendium is particularly powerful for prioritized screening applications. When seeking to identify an unknown protein that enables a particular mitochondrial function, starting with MitoCarta immediately narrows the search space. This approach has enabled identification of components of the mitochondrial calcium uniporter (Baughman et al., 2011; Perocchi et al., 2010), and in Chapter 2 enables the identification of a new regulator of mitochondrial RNAs, FASTKD4. Additionally, MitoCarta has enabled prioritization of candidate genes for molecular diagnosis of mitochondrial disorders caused by nuclear mutations (Calvo et al., 2012; Lieber et al., 2013). However, these variants cannot be proven pathogenic without molecular proof, and in Chapter 3 I describe an approach to tackle this challenge using a mitochondrial RNA profiling technique developed in Chapter 2. In Chapter 4, I use sequential enrichment of mitochondria to determine whether any nuclear RNA transcripts are imported into mitochondria along with MitoCarta proteins.

Why does mtDNA persist?

If so many historically mitochondrial genes have been successfully transitioned to the nucleus, why do any remain encoded in mtDNA? Three theories have been posed to answer this question. First, the proteins that remain encoded by mtDNA are extremely hydrophobic, and their efficient translocation into mitochondria may be

inhibited by this property, as only unfolded proteins can be translocated (Popot and de Vitry, 1990). Second, the mtDNA genetic code is slightly different than the nuclear genetic code, and may prevent efficient transfer to the nucleus; in particular, a nuclear stop codon encodes tryptophan within the organelle (Jacobs, 1991). Finally, others have argued that local control of mtDNA allows rapid response of gene expression to local metabolic changes, known as Co-location for redox regulation (CORR) (Allen, 1993). These theories are not mutually exclusive, and highlight the unique nature of the remaining mtDNA-encoded proteins.

mtDNA is transcribed in long polycistrons and cleaved into individual transcripts

MtDNA is transcribed in long polycistronic units by mitochondrial RNA polymerase (POLRMT), which is distantly related to the polymerase from bacteriophage T7. Transcription initiates from three distinct promoters: light strand promoter (LSP), which encodes the mRNA *ND6* and 8 tRNAs, and two heavy strand promoters (HSP1 and HSP2) (Figure 1-1) (Chang and Clayton, 1984; Montoya et al., 1982). HSP1 transcribes the two rRNA units and the first two tRNAs, while HSP2 transcribes the remainder of the heavy strand, producing 2 rRNAs, 10 mRNAs and 13 tRNAs (Montoya et al., 1983). The mRNAs *ATP6/8* and *ND4/4L* are bicistroinc, each containing two overlapping reading frames. The transcription factors TFAM and TFB2M are required for transcription from LSP and HSP1 in vitro (Litonin et al., 2010) (Figure 1-2).

These multigenic precursor strands are cleaved into individual rRNAs, tRNAs, and mRNAs. The secondary structure of the tRNAs, which intersperse the mRNAs, is thought to guide cleavage (Ojala et al., 1981). RNase P cleaves at the 5' end of most

tRNAs, while RNase Z (ELAC2) cleaves at the 3' end (Figure 1-2) (Brzezniak et al., 2011; Rossmanith and Karwan, 1998). Mitochondrial RNase P is a complex of three proteins: mitochondrial RNase P protein (MRPP) 1, MRPP2, and MRPP3 (Holzmann et al., 2008). In addition to this identified protein complex, a mitochondrial *RNase P* RNA has been identified with 5' cleavage activity (Puranam and Attardi, 2001).

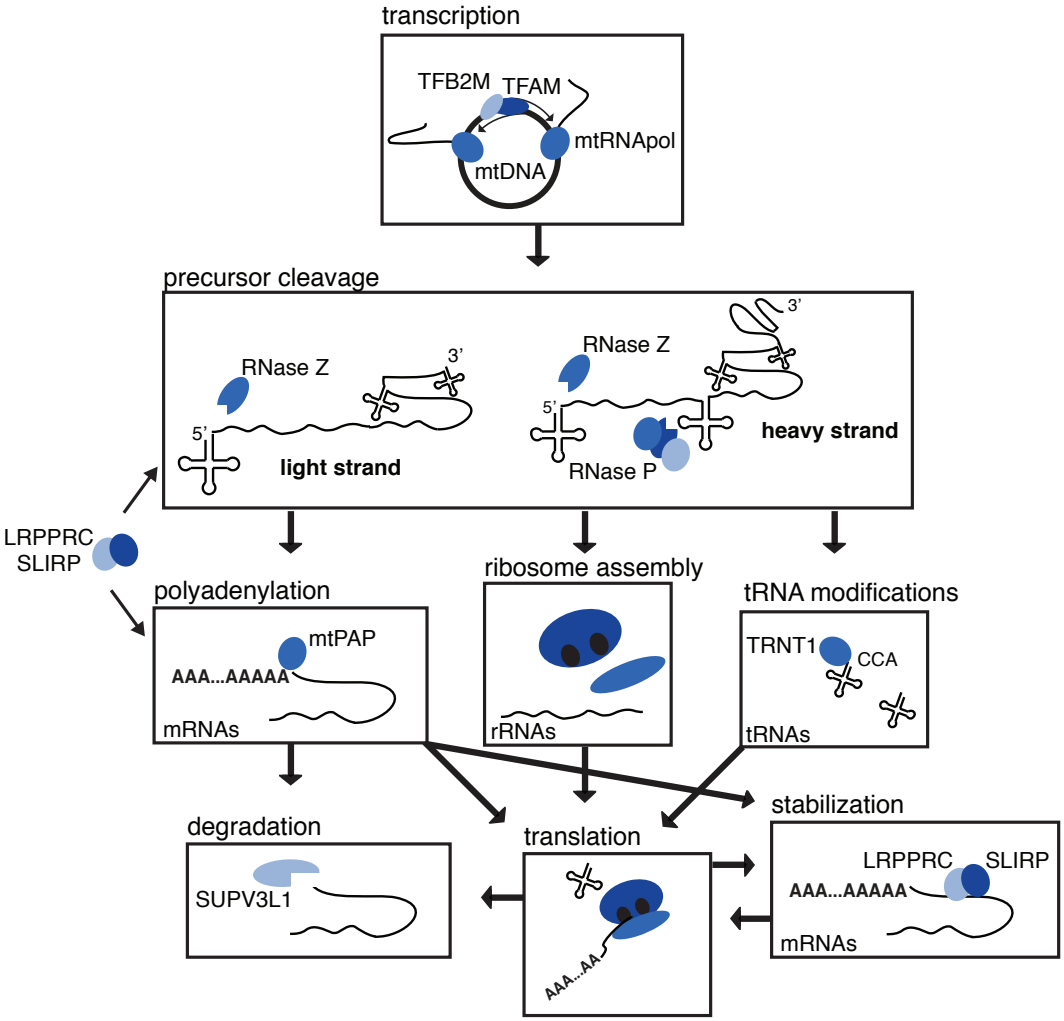


Figure 1-2. Mitochondrial RNA processing
Mitochondrial RNAs are transcribed in long precursor strands, known as the heavy and light strands, which are cleaved by RNase Z and P into individual rRNAs, tRNAs, and mRNAs. These transcripts are further modified and can contribute to translation, be degraded, or stabilized. LRPPRC and SLIRP form a complex that stabilizes mRNAs and has been also implicated in polyadenylation and abundance of the precursors.

However, some junctions do not appear to behave as expected under knockdown of RNase P or Z (Lopez Sanchez et al., 2011).

All mRNAs in mitochondria are 3' polyadenylated with a 40-60 nucleotide tail. For some transcripts, the genome has been so compacted that polyadenylation is necessary to complete a stop codon (Temperley et al., 2010). Knockdown experiments have verified the role of mitochondrial poly(A) polymerase (MTPAP) in 3' adenylation, however in these reports a ~10 nucleotide poly(A) tail remains, leading some to hypothesize that an additional polymerase is involved in the primary addition of adenosines, with MTPAP playing a role in extension of the poly(A) tail (Slomovic et al., 2005; Slomovic and Schuster, 2008). Knockdown of MTPAP affects the stability of some, but not all of the mitochondrial mRNAs, and severely compromises mitochondrial membrane potential and oxygen consumption. Furthermore, MTPAP does not contain an RNA-binding domain, and is expected to cooperate with an unidentified partner in order to bind RNA (Bobrowicz et al., 2008). Internal poly(A) sites have also been identified, suggesting that this polyadenylation may be part of a degradation pathway (Slomovic et al., 2005). However, it is unclear how the two poly(A) signals are differentiated and which enzymes are responsible for the internal cleavage and subsequent polyadenylation.

LRPPRC, SLIRP, and GRSF1 have important but murky roles in mitochondrial RNA regulation

The LRPPRC-SLIRP complex is essential for mitochondrial mRNA abundance and sediments with a subset of mitochondrial mRNAs when separated by a density

gradient (Ruzzenente et al., 2012). However, the exact molecular role remains contested. Leucine-rich PPR-motif containing protein (LRPPRC) binds mitochondrial polyadenylated RNAs in vivo and may play a role in nuclear RNA processing (Mili and Piñol-Roma, 2003). LRPPRC contains the pentatricopeptide repeat (PPR) RNA-binding motif, which is prevalent in hundreds of proteins active in plant organelle RNA processing, but is only present in five mammalian proteins, all of which are mitochondrial-targeted (Lightowlers and Chrzanowska-Lightowlers, 2008). SLIRP is found in complex with LRPPRC, and its depletion results in decreased RNA transcript levels and oxygen consumption (Baughman et al., 2009; Sasarman et al., 2010). LRPPRC has been linked to the transcription, polyadenylation, translation and degradation of mitochondrial mRNAs (Chujo et al., 2012; Gohil et al., 2010; Ruzzenente et al., 2012; Sasarman et al., 2010; Xu et al., 2004). How one protein can play so many roles has yet to be established, but it has been hypothesized that the mitochondrial precursor is coated with the LRPPRC-SLIRP complex from a very early stage and continues to chaperone the mRNAs through polyadenylation, translation and degradation (Chujo et al., 2012; Ruzzenente et al., 2012). Recently it has been suggested that mRNAs with longer half-lives are more affected by knockdowns of these proteins, but the correlation is imperfect (Chujo et al., 2012; Ruzzenente et al., 2012).

Recent studies have found a role for GRSF1 in regulation of mitochondrial RNAs. GRSF1 localizes to RNA granules, punctate loci within the mitochondrial matrix, alongside RNase P, and is required for precursor cleavage (Jourdain et al., 2013), and it preferentially associates with light strand transcripts (Antonicka et al., 2013). GRSF1

depletion also depleted rRNAs and some mRNAs and impaired ribosomal assembly (Antonicka et al., 2013; Jourdain et al., 2013)

Mitochondrial RNA translation and degradation

Each mitochondrial mRNA is translated by the mitochondrial ribosome, which is composed of the two mtDNA-encoded rRNAs and a number of nuclear-encoded proteins. The mitochondrial mRNAs were initially believed to be translated with just 22 mitochondrial-encoded tRNAs, employing a slightly modified coding language that makes generous use of the wobble base (Crick, 1966). These tRNAs are excised from the precursor as described above, and modified by enzymes including TRNT1, which adds a CCA to the end of each tRNA (Nagaike et al., 2001).

E. coli and yeast mitochondria contain nonhomologous protein complexes termed degradosomes for destroying RNA (Carpousis, 2007; Malecki et al., 2007). Human mitochondria contain homologs of degradosome components including polynucleotide phosphorylase (PNPase encoded by *PNPT1*) (*E. coli*) and the helicase SUPV3L1 (yeast); these human proteins have recently been found to form a complex in vitro (Wang et al., 2009). Further, knockdowns of both *SUPV3L1* and *PNPT1* disrupt mitochondrial function (Khidr et al., 2008; Slomovic and Schuster, 2008). However, PNPase is located in the intermembrane space, which has led to the hypothesis that it affects degradation of poly(A) tails through indirect means such as altering free nucleotide levels (Slomovic and Schuster, 2008). No other proteins have been implicated in degradation thus far, but the unclear roles of SUV3 and PNPase suggest that other proteins may be involved.

Import of nuclear RNAs into the mitochondria

Interestingly, PNPase is also required for import of nuclear RNAs into mitochondria (Wang et al., 2010). Mitochondrial import has been described for *5S* rRNA, which associates with the mitochondrial ribosome, *MRP* RNA, which couples with a protein component to cleave RNA primers required for DNA synthesis, and *RNase P* RNA, described above (Li et al., 1994; Puranam and Attardi, 2001; Smirnov et al., 2011; Yoshionari et al., 1994). PNPase was found to stimulate import of both the *MRP* and *RNase P* RNAs, and *5S* import has been biochemically characterized to require a specific RNA sequence and multiple cytoplasmic proteins (Smirnov et al., 2010; Smirnov et al., 2011; Smirnov et al., 2008; Wang et al., 2010). Although the 22 tRNAs present in the human mitochondria are sufficient to decode the mitochondrial genome with wobble of the third anticodon base, nuclear tRNAs have recently been identified within the mitochondria (Chomyn and Attardi, 2009; Mercer et al., 2011; Rubio et al., 2008). The purpose and necessity of these RNAs remains unclear, and in Chapter 4, we describe our approach to identify and study these nuclear-encoded mitochondrial-imported RNAs using an unbiased approach.

Mutations in genes required for mitochondrial RNA processing cause hereditary mitochondrial disorders

Inherited disorders affecting mitochondrial function can be caused by mutations in both mtDNA and the nuclear genome. These disorders can have diverse presentation, but usually affect multiple organ systems, including the nervous system, musculoskeletal system, and gastrointestinal tract (reviewed in Vafai and Mootha,

2012). These disorders can present in patients from infancy to adulthood and range in severity from mild to lethal.

Mitochondrial disorders can result from a malfunctioning respiratory chain due to a plethora of causes (Calvo and Mootha, 2010). Most directly, mutations in proteins that are subunits of the respiratory chain or chaperone proteins required for proper OXPHOS function can impact respiratory chain function. Mutations in proteins required for mtDNA maintenance and protein translation can also result in decreased OXPHOS function by impacting production of mtDNA-encoded respiratory chain subunits. Proteins responsible for RNA processing and abundance have also been implicated.

In 2003, mutations in *LRPPRC* were identified as a cause of Leigh Syndrome, French Canadian type (LSFC), a severe infantile disorder characterized by neurodegeneration and severe episodes of metabolic acidosis (Mootha et al., 2003). LSFC patients had low cytochrome *c* oxidase activity, but even after identification of causal mutations in *LRPPRC*, the exact role of the protein remained unclear. Subsequent studies, described above, have shown that *LRPPRC* plays a role in regulation of mitochondrial RNAs, characterizing it as the first mitochondrial disease gene affecting mitochondrial RNAs (Gohil et al., 2010; Sasarman et al., 2010).

Subsequently, disease-causing mutations have been found in genes encoding the RNA processing factors mitochondrial poly-A polymerase (*MTPAP*) (Crosby et al., 2010) and RNase Z (*ELAC2*) (Haack et al., 2013). Defects in *MTPAP* were found in an Amish family with several siblings affected by slowly progressive neurodegenerative condition marked by cerebellar ataxia, spastic lower limb movements, dysarthria

(difficulty speaking), optic atrophy and impaired learning. In the affected siblings, but not the parents, improper polyadenylation of mitochondrial RNAs was observed, confirming that suspected variants in *MTPAP* cause a dysregulation of mitochondrial RNA processing, and likely the disease observed. *ELAC2* mutations were identified in unrelated individuals with infantile hypertrophic cardiomyopathy, lactic acidosis, and isolated complex I deficiency in skeletal muscle (Haack et al., 2013). In this case, measurement of precursor RNAs by qPCR and RNA-sequencing identified aberrantly increased precursor RNAs in the affected individuals, suggesting that the *ELAC2* mutations are causal.

Although exome-sequencing has become widespread for many inherited disorders, proving pathogenicity of identified variants of unknown significance is a challenge. Recently, Chakravarti et al. proposed a version of Koch's postulates for complex human disease, including sporadic monogenic cases. The authors require proof at two levels: demonstration that a mutant allele causes a phenotype in a model system and demonstration that the model phenotype is equivalent to the human disease, using four crisp postulates (Chakravarti et al., 2013). For both the *MTPAP* and *ELAC2* cases, an understanding of the biology of these enzymes enabled determination of the likely pathogenicity of the identified variants using experimental techniques. With the benefit of Mendelian inheritance patterns, experimental proof was less necessary. However, noting how each falls short of the rigorous standards proposed by Chakravarti et al. highlights the challenges of meeting this high standard of proof. Both papers use family history to prove the enrichment of mutant alleles within patients (Postulate 1) and

fibroblast models to demonstrate a phenotype associated with disrupted gene function (Postulate 2). However, only in the *ELAC2* case do the authors reintroduce the wild type allele (Postulate 3), and neither case re-introduces the mutant allele (Postulate 4). Following these principles, in Chapter 3 we attempt to develop a model in fibroblasts for a patient with candidate variants in a gene encoding a component of the mitochondrial RNA degradation machinery, *SUPV3L1*. We were aided in our approach by the body of knowledge surrounding this protein and a known function, described above, but still faced a formidable challenge. For genes that are less well characterized, developing a model system will be even more challenging.

In a recent MitoExome sequencing study, molecular diagnoses were established for just 22% of suspected mitochondrial disease patients, while variants of unknown significance were prioritized for an additional 26% (Lieber et al., 2013). For many of these variants of unknown significance, the tools to develop and assess an appropriate model system are lacking. This emphasizes both the necessity of understanding the roles various mitochondrial enzymes play, as well as the need for molecular techniques that can probe the function of specific enzymes. Towards this effort, in Chapter 2 I develop a molecular tool for probing mitochondrial RNA expression and characterize over 100 mitochondrial proteins, including 13 known disease genes.

References

Allen, J.F. (1993). Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. J Theor Biol *165*, 609-631.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F.*, et al.* (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457-465.

Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V., and Shoubridge, E.A. (2013). The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. Cell Metab *17*, 386-398.

Baughman, J.M., Nilsson, R., Gohil, V., Arlow, D.H., Gauhar, Z., and Mootha, V. (2009). A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. PLoS Genetics *5*, e1000590.

Baughman, J.M., Perocchi, F., Girgis, H.S., Plovanich, M., Belcher-Timme, C.A., Sancak, Y., Bao, X.R., Strittmatter, L., Goldberger, O., Bogorad, R.L.*, et al.* (2011). Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. Nature *476*, 341-345.

Bobrowicz, A.J., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z. (2008). Polyadenylation and degradation of mRNA in mammalian mitochondria: a missing link? Biochemical Society Transactions *36*, 517-519.

Brzezniak, L.K., Bijata, M., Szczesny, R.J., and Stepien, P.P. (2011). Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. RNA biology *8*, 616-626.

Calvo, S.E., Compton, A.G., Hershman, S.G., Lim, S.C., Lieber, D.S., Tucker, E.J., Laskowski, A., Garone, C., Liu, S., Jaffe, D.B.*, et al.* (2012). Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Sci Transl Med *4*, 118ra110.

Calvo, S.E., and Mootha, V.K. (2010). The mitochondrial proteome and human disease. Annu Rev Genomics Hum Genet *11*, 25-44.

Carpousis, A.J. (2007). The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. Annu Rev Microbiol *61*, 71-87.

Chakravarti, A., Clark, A.G., and Mootha, V.K. (2013). Distilling pathophysiology from complex disease genetics. Cell *155*, 21-26.

Chang, D.D., and Clayton, D.A. (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell *36*, 635-643.

Chomyn, A., and Attardi, G. (2009). Mitochondrial Gene Products. Current topics in bioenergetics *15*, 295-329.

Chujo, T., Ohira, T., Sakaguchi, Y., Goshima, N., Nomura, N., Nagao, A., and Suzuki, T. (2012). LRPPRC/SLIRP suppresses PNPase-mediated mRNA decay and promotes polyadenylation in human mitochondria. Nucleic Acids Res *40*, 8033-8047.

Crick, F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. J Mol Biol *19*, 548-555.

Crosby, A.H., Patel, H., Chioza, B.A., Proukakis, C., Gurtz, K., Patton, M.A., Sharifi, R., Harlalka, G., Simpson, M.A., Dick, K.*, et al.* (2010). Defective mitochondrial mRNA maturation is associated with spastic ataxia. Am J Hum Genet *87*, 655-660.

Gohil, V.M., Nilsson, R., Belcher-Timme, C.A., Luo, B., Root, D.E., and Mootha, V.K. (2010). Mitochondrial and nuclear genomic responses to loss of LRPPRC expression. J Biol Chem *285*, 13742-13747.

Haack, T.B., Kopajtich, R., Freisinger, P., Wieland, T., Rorbach, J., Nicholls, T.J., Baruffini, E., Walther, A., Danhauser, K., Zimmermann, F.A.*, et al.* (2013). ELAC2 mutations cause a mitochondrial RNA processing defect associated with hypertrophic cardiomyopathy. Am J Hum Genet *93*, 211-223.

Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell *135*, 462-474.

Jacobs, H.T. (1991). Structural similarities between a mitochondrially encoded polypeptide and a family of prokaryotic respiratory toxins involved in plasmid maintenance suggest a novel mechanism for the evolutionary maintenance of mitochondrial DNA. J Mol Evol *32*, 333-339.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab *17*, 399-410.

Khidr, L., Wu, G., Davila, A., Procaccio, V., Wallace, D., and Lee, W.H. (2008). Role of SUV3 helicase in maintaining mitochondrial homeostasis in human cells. J Biol Chem *283*, 27064-27073.

Li, K., Smagula, C.S., Parsons, W.J., Richardson, J.A., Gonzalez, M., Hagler, H.K., and Williams, R.S. (1994). Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis. The Journal of cell biology *124*, 871-882.

Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R., Berry, G.T.*, et al.* (2013). Targeted exome sequencing of suspected mitochondrial disorders. Neurology *80*, 1762-1770.

Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2008). PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. Biochemical Journal *416*, e5-6.

Litonin, D., Sologub, M., Shi, Y., Savkina, M., Anikin, M., Falkenberg, M., Gustafsson, C.M., and Temiakov, D. (2010). Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro. J Biol Chem *285*, 18129-18133.

Lopez Sanchez, M.I.G., Mercer, T.R., Davies, S.M.K., Shearwood, A.J., Nygård, K.K.A., Richman, T.R., Mattick, J.S., Rackham, O., and Filipovska, A. (2011). RNA processing in human mitochondria. Cell Cycle *10*, 2904-2916.

Malecki, M., Jedrzejczak, R., Stepien, P.P., and Golik, P. (2007). In vitro reconstitution and characterization of the yeast mitochondrial degradosome complex unravels tight functional interdependence. J Mol Biol *372*, 23-36.

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

Mili, S., and Piñol-Roma, S. (2003). LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. Mol Cell Biol *23*, 4972-4982.

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M., and Attardi, G. (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci U S A *79*, 7195-7199.

Montoya, J., Gaines, G.L., and Attardi, G. (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell *34*, 151-159.

Mootha, V., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F.*, et al.* (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc Natl Acad Sci USA *100*, 605-610.

Nagaike, T., Suzuki, T., Tomari, Y., Takemoto-Hori, C., Negayama, F., Watanabe, K., and Ueda, T. (2001). Identification and characterization of mammalian mitochondrial tRNA nucleotidyltransferases. The Journal of biological chemistry *276*, 40041-40049.

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470-474.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K.*, et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112-123.

Perocchi, F., Gohil, V.M., Girgis, H.S., Bao, X.R., McCombs, J.E., Palmer, A.E., and Mootha, V.K. (2010). MICU1 encodes a mitochondrial EF hand protein required for Ca(2+) uptake. Nature *467*, 291-296.

Popot, J.L., and de Vitry, C. (1990). On the microassembly of integral membrane proteins. Annu Rev Biophys Biophys Chem *19*, 369-403.

Puranam, R.S., and Attardi, G. (2001). The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. Mol Cell Biol *21*, 548-561.

Rossmanith, W., and Karwan, R.M. (1998). Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. Biochem Biophys Res Commun *247*, 234-241.

Rubio, M.A., Rinehart, J.J., Krett, B., Duvezin-Caubet, S., Reichert, A.S., Soll, D., and Alfonzo, J. (2008). Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. Proc Natl Acad Sci USA *105*, 9186-9191.

Ruzzenente, B., Metodiev, M.D., Wredenberg, A., Bratic, A., Park, C.B., Camara, Y., Milenkovic, D., Zickermann, V., Wibom, R., Hultenby, K.*, et al.* (2012). LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. EMBO J *31*, 443-456.

Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T., Shoubridge, E.A., and Consortium, L. (2010). LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria. Mol Biol Cell *21*, 1315-1323.

Sicheritz-Pontén, T., Kurland, C.G., and Andersson, S.G. (1998). A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. Biochim Biophys Acta *1365*, 545-551.

Slomovic, S., Laufer, D., Geiger, D., and Schuster, G. (2005). Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. Molecular and cellular biology *25*, 6427-6435.

Slomovic, S., and Schuster, G. (2008). Stable PNPase RNAi silencing: its effect on the processing and adenylation of human mitochondrial RNA. RNA *14*, 310-323.

Smirnov, A., Comte, C., Mager-Heckel, A.M., Addis, V., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I. (2010). Mitochondrial enzyme rhodanese is essential for 5 S ribosomal RNA import into human mitochondria. J Biol Chem *285*, 30792-30803.

Smirnov, A., Entelis, N., Martin, R.P., and Tarassov, I. (2011). Biological significance of 5S rRNA import into human mitochondria: role of ribosomal protein MRP-L18. Genes Dev *25*, 1289-1305.

Smirnov, A.V., Entelis, N.S., Krasheninnikov, I.A., Martin, R., and Tarassov, I.A. (2008). Specific features of 5S rRNA structure - its interactions with macromolecules and possible functions. Biochemistry BiokhimiiÕ°a *73*, 1418-1437.

Temperley, R.J., Wydro, M., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2010). Human mitochondrial mRNAs--like members of all families, similar but different. Biochim Biophys Acta *1797*, 1081-1085.

Vafai, S.B., and Mootha, V.K. (2012). Mitochondrial disorders as windows into an ancient organelle. Nature *491*, 374-383.

Wang, D.D., Shu, Z., Lieser, S.A., Chen, P.L., and Lee, W.H. (2009). Human mitochondrial SUV3 and polynucleotide phosphorylase form a 330-kDa heteropentamer to cooperatively degrade double-stranded RNA with a 3'-to-5' directionality. J Biol Chem *284*, 20812-20821.

Wang, G., Chen, H.W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., McCaffery, J.M.*, et al.* (2010). PNPASE regulates RNA import into mitochondria. Cell *142*, 456-467.

Xu, F., Morin, C., Mitchell, G., Ackerley, C., and Robinson, B.H. (2004). The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA. The Biochemical journal *382*, 331-336.

Yoshionari, S., Koike, T., Yokogawa, T., Nishikawa, K., Ueda, T., Miura, K., and Watanabe, K. (1994). Existence of nuclear-encoded 5S-rRNA in bovine mitochondria. FEBS Lett *338*, 137-142.

Chapter 2
Functional genomic analysis of human mitochondrial RNA processing

Contributors: Ashley R. Wolf and Vamsi K. Mootha

Supplementary information can be found in Appendix A
This work has been accepted for publication at Cell Reports

Summary

Both strands of human mitochondrial DNA (mtDNA) are transcribed in continuous, multi-genic units that are cleaved into the mature rRNAs, tRNAs, and mRNAs required for respiratory chain biogenesis. We sought to systematically identify nuclear-encoded proteins that contribute to processing of mitochondrial RNAs (mt-RNAs) within the organelle. First, we devised and validated a multiplex "MitoString" assay that quantitates 27 mature and precursor mtDNA transcripts. Second, we applied MitoString profiling to evaluate the impact of silencing each of 107 mitochondrial-localized, predicted RNA-binding proteins. With the resulting dataset, we rediscover the roles of recently identified RNA processing enzymes, detect unanticipated roles of known disease genes in RNA processing, and identify new regulatory factors. We demonstrate that one such factor, FASTKD4, modulates half-lives of a subset of mt-mRNAs and associates with mitochondrial RNAs *in vivo*. MitoString profiling may be useful in diagnosing and deciphering the pathogenesis of mtDNA disorders.

Introduction

Human mitochondrial DNA (mtDNA) encodes 13 protein subunits of the respiratory chain, as well as the 2 rRNAs and 22 tRNAs required for their translation. All protein factors required for mtDNA replication, transcription, and translation are nuclear-encoded and imported into the mitochondrion. Mutations in either nuclear- or mitochondrial-encoded genes cause a range of heritable disorders with overlapping phenotypes. Expression of nuclear and mtDNA-encoded oxidative phosphorylation (OXPHOS) complex proteins is coordinated by nuclear transcription factors and co-activators (Scarpulla et al., 2012). However, local control of mtDNA expression is also possible, as suggested by by the finding that chemical perturbations are capable of decoupling the expression of the two genomes (Wagner et al., 2008).

The 16.5 kB human mitochondrial genome is highly compact and transcribed by mitochondrial RNA polymerase as two continuous polycistrons, one for each strand. The “heavy strand” expresses 2 rRNAs, 14 tRNAs and 12 mRNAs, while the “light strand” expresses the *ND6* mRNA and 8 tRNAs (Figure 2-1AB). These precursor RNAs primarily contain mt-mRNAs “punctuated” by tRNAs, whose structure is proposed to guide the cleavage responsible for liberating individual mt-mRNAs and tRNAs (Anderson et al., 1981; Ojala et al., 1981). Cleavage at the 5’ end of tRNAs is catalyzed by the RNase P complex, which is comprised of three recently identified proteins: mitochondrial RNase P protein 1 (MRPP1), MRPP2, and MRPP3 (Holzmann et al., 2008). An alternative RNase P containing an imported catalytic RNA has also been

described (Puranam and Attardi, 2001). Cleavage at the 3' end of tRNAs is catalyzed by the nuclease ELAC2 (Brzezniak et al., 2011; Lopez Sanchez et al.,

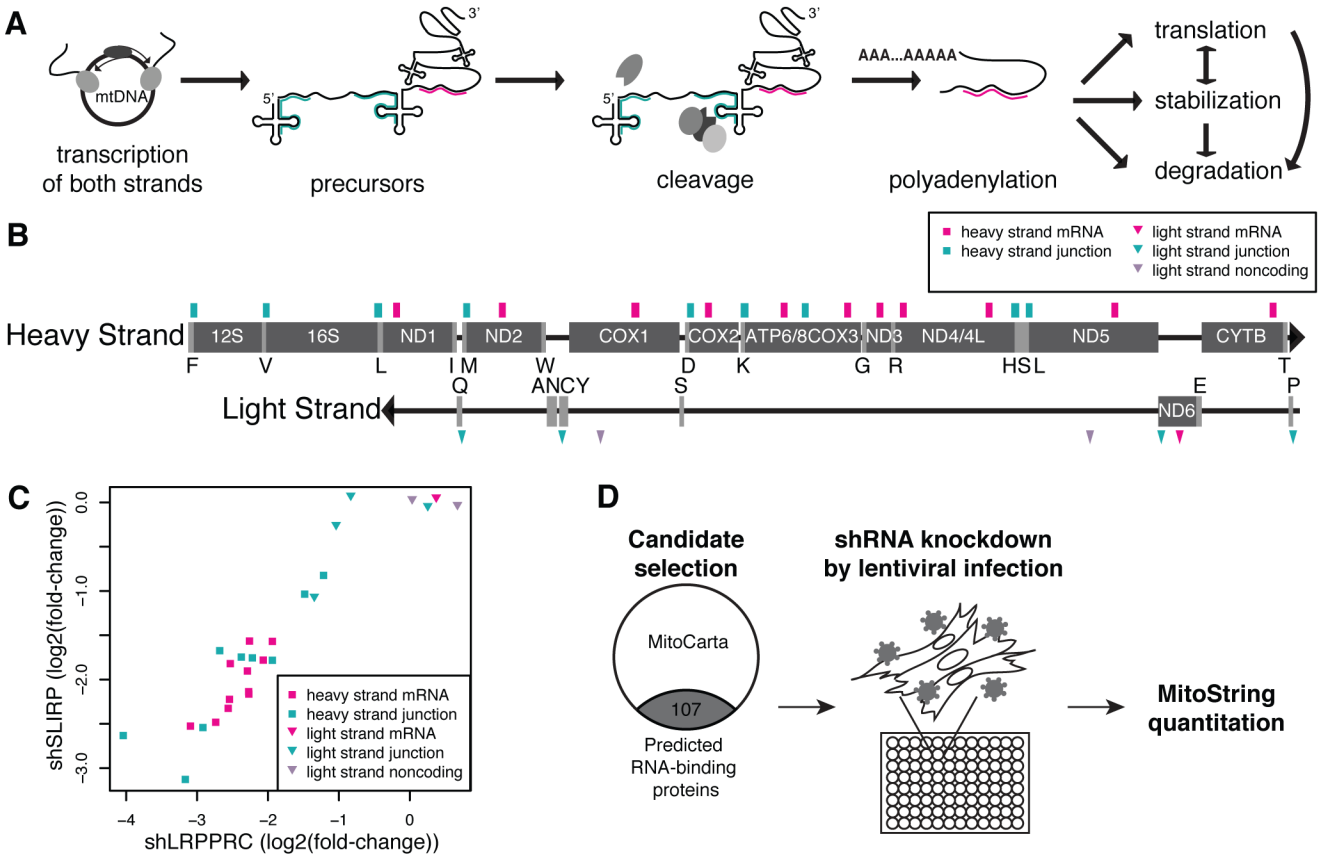


Figure 2-1. MitoString screen for regulators of mitochondrial RNA processing
(A) Schematic depicting mtDNA transcription (in two continuous units, the heavy and light strands), followed by cleavage into individual mRNAs, tRNAs, and rRNAs. Pink and turquoise lines represent MitoString probes targeting mRNAs and junctions, respectively. (B) Location of MitoString probes on the mtDNA heavy and light strands. rRNAs and mRNAs encoded by the mtDNA are labeled in white text. tRNAs are demarcated by their one letter symbols. The location of mRNA probes are noted in pink, junction probes in turquoise, and noncoding probes in grey. Rectangles indicate probes targeting the heavy strand, while triangles indicate probes targeting the light strand. (C) Comparison of the effects of *shLRPPRC* and *shSLIRP* on MitoString probes, shown as log2(fold-change) with respect to *shGFP*. Probe colors and shapes are depicted as in (B). (D) Overview of MitoString screening approach. Candidates for lentiviral knockdown were selected by identifying MitoCarta proteins containing RNA-binding domains. Candidate genes were knocked down in WI-38 fibroblasts and the mitochondrial RNA levels were assessed six days later by the MitoString probes described in (B).

2011). After cleavage, the mitochondrial RNA poly-A polymerase (MTPAP) polyadenylates the mt-mRNAs (Nagao et al., 2008; Piechota et al., 2006). Mt-mRNA abundance is regulated by the SLIRP- LRPPRC complex, although the exact mechanism is debated. LRPPRC has been implicated in mt-mRNA transcription, polyadenylation, translation and degradation suppression (Baughman et al., 2009; Chujo et al., 2012; Gohil et al., 2010; Liu et al., 2011; Ruzzenente et al., 2012; Sasarman et al., 2010).

Despite the concurrent transcription of heavy strand genes, their cognate mt-mRNAs reach distinct steady-state levels. These ten transcripts have distinct half-lives that fall into two categories: short-lived and long-lived. The complex I transcripts *ND1–ND3* and *ND5*, and the complex III transcript *CYTB* are short-lived ($t_{1/2}$ = 68–94 min), whereas the complex IV transcripts cytochrome *c* oxidase 1 (*COX1*), *COX2*, and *COX3*; complex V bicistronic transcript *ATP6/8*, and complex I bicistronic transcript *ND4/4L* are long-lived ($t_{1/2}$= 138-231 min) (Nagao et al., 2008). These differential mt-mRNA stabilities are consistent with early observations (Gelfand and Attardi, 1981) and recent RNA sequencing analysis (Mercer et al., 2011), but remain unexplained by transcript length, polyadenylation, known degradation pathways, or characterized stability factors. Although the purpose of these differential half-lives is unexplored, we note that concentrations of OXPHOS protein complexes (reviewed by Lenaz and Genova, 2010) tend to correlate with reported mt-mRNA half-lives (Nagao et al., 2008). Complex I is the least abundant complex and contains subunits encoded by the short-lived mt-mRNAs.

Thus, differences in mt-mRNA abundance may help establish OXPHOS protein stoichiometry.

Because all mt-mRNA transcripts but one originate from a single heavy strand promoter, the observed steady-state levels are expected to be highly dependent on transcript degradation rates (Chujo et al., 2012). The helicase SUPV3L1, in complex with polynucleotide phosphorylase (PNPT1), has been implicated in the degradation of the light-strand transcripts, and a dominant negative form of either gene stabilizes light strand noncoding transcripts and some short-lived heavy strand mt-mRNAs (Borowski et al., 2013; Szczesny et al., 2010). However, both RNAi and dominant negative experiments targeting *SUPV3L1* or *PNPT1* actually decrease levels of the long-lived mt-mRNA *COX1,* indicating additional degradation or feedback mechanisms may exist.

Our goal was to systematically identify mitochondrial proteins that contribute to mitochondrial RNA processing. We begin with a scalable, accurate method to measure multiple mitochondrial RNAs throughout the processing stages. Past approaches to measuring mitochondrial RNA levels, such as Northern blots, quantitative PCR, and GE-HTS (Wagner et al., 2008), have been valuable, but limited by scalability, strand specificity, or dynamic range, respectively. Here, we report simultaneous, strand-specific measurement of multiple precursor and mature mtDNA-encoded RNAs following stable genetic silencing of nuclear factors predicted to play a role in mitochondrial RNA biology. We produce a focused compendium of mitochondrial RNA expression across a set of targeted genetic perturbations, which we mine to probe the identity and role of nuclear-encoded factors in mitochondrial RNA processing. In the

process, we identify *FASTKD4*, a novel factor that regulates the stability of a subset of mt-mRNAs.

Results

MitoString: a multiplexed assay for precursor and mature mt-RNAs

We developed a "MitoString" assay to interrogate four types of transcripts using the nCounter Analysis System (Geiss et al., 2008), in which fluorescent RNA probes quantitate unamplified RNA within a crude cell lysate sample. First, we designed a probe targeting each of the highly abundant mt-mRNA transcripts (Figure 2-1B), achieving a coefficient of variation (CV) of 7-14% for all but one probe across 10 independent infections of the control hairpin *shGFP* (Figure 2-S1AB). Second, we designed probes to two regions of the light strand precursor that are transcribed, but are not believed to encode a functional protein. Third, we designed probes overlapping the junctions of two adjacent genes. These junction probes only produce signal when bound to the unprocessed precursor transcript. Probes assessing the precursor strands have lower signal at baseline, and are thus noisier (Figure 2-S1C), but their levels can be reproducibly induced in response to perturbation. In addition, we designed probes to detect nuclear-encoded genes important to mitochondrial function, as well as the nuclear-encoded candidate genes, with the vast majority having a CV of <20% (Figure 2-S1D).

To verify that mtRNA perturbations are reliably measured, we profiled cells following transduction with *PGC-1α*, which induces mitochondrial biogenesis, and *shLRPPRC*, which depletes mtRNAs. *PGC-1α* overexpression increases mature and

immature transcripts (Figure 2-S2AB), as well as nuclear-encoded, mitochondrial-localized transcripts (Figure 2-S2C), as expected. *LRPPRC* silencing depletes heavy strand mt-mRNAs as found previously (Figure 2-1C). We further note that knockdowns of the individual components of the LRPPRC-SLIRP complex have well-correlated effects on all transcripts measured by MitoString (Figure 2-1C), suggesting that hierarchical clustering of MitoString profiles may be valuable for predicting gene function.

Prioritizing candidate mt-RNA binding proteins for knockdown

With a facile assay for quantifying immature and mature mitochondrial transcripts in hand, we proceeded to select genes that might be involved in mitochondrial RNA processing for RNAi-based silencing. We focused on members of MitoCarta, a high-confidence collection of mitochondrial-localized proteins (Pagliarini et al., 2008), and prioritized proteins with known or predicted RNA-binding domains, based on Pfam, GO annotation, or manual curation (Figure 2-1D) (Finn et al., 2008). We excluded components of the mitochondrial ribosome and in total prioritized 107 candidates.

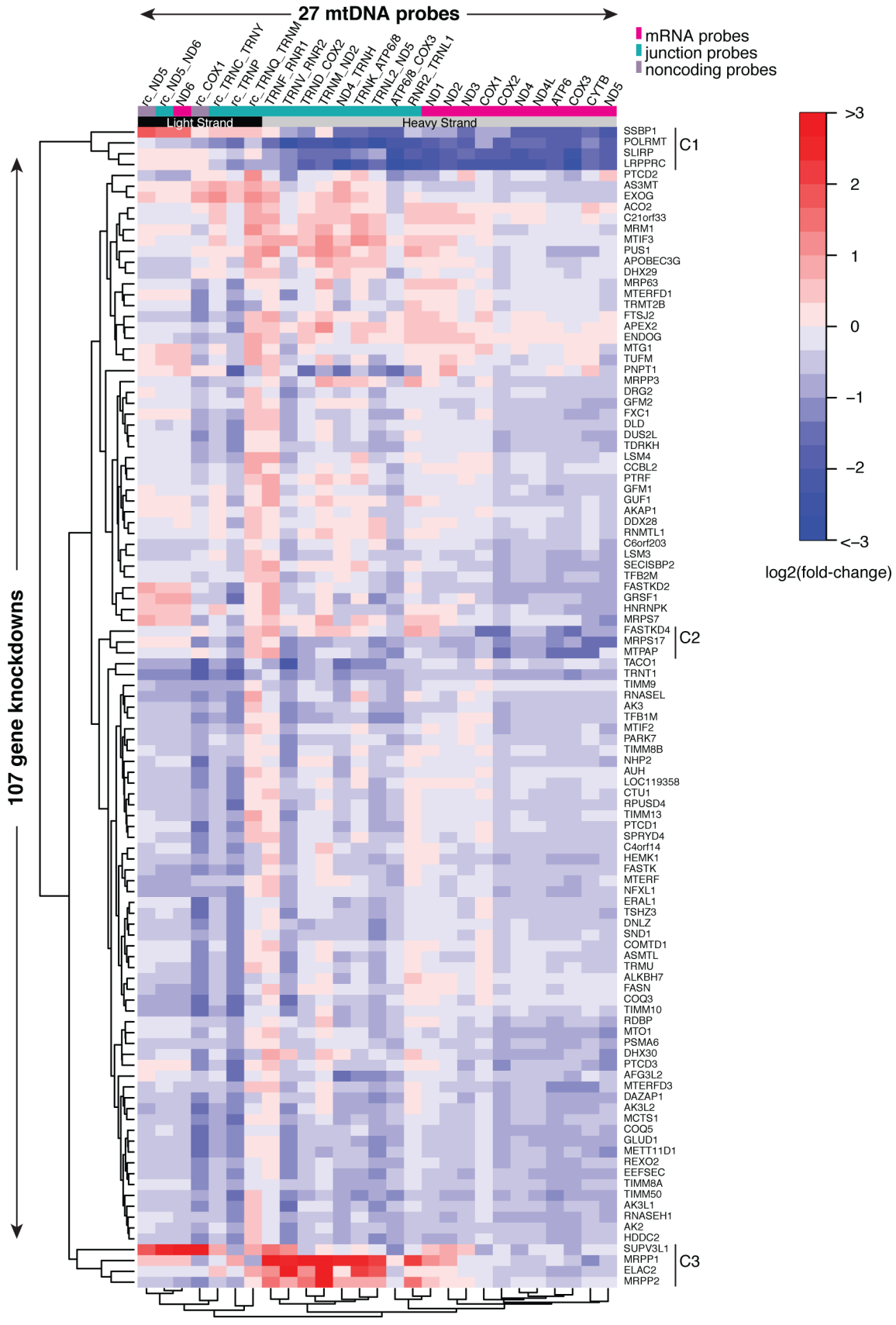
For each of these genes, we selected the three most effective lentiviral short hairpin RNAs (shRNAs) available from the RNAi Consortium (Table 2-S1) (Methods). We screened WI-38 fetal lung fibroblasts, which are untransformed and viable after six days of strong *LRPPRC* knockdown. Cells were infected with hairpins in duplicate, and lysed for MitoString analysis after six days of antibiotic selection. Using our nuclear NanoString probes, we were able to measure knockdown efficiency for most hairpins within the same assay (Table 2-S2). On each 96-well screening plate, we included five hairpins targeting non-human sequences as RNAi controls, a hairpin targeting *LRPPRC*

known to deplete mt-mRNAs (Figure 2-1C), and a *PGC-1α* overexpression construct to stimulate mitochondrial biogenesis (Figure 2-S2A-C).

A compendium of perturbational profiles for mitochondrial RNA

MitoString quantitation of knockdown perturbations provides a valuable compendium for gene function prediction and mtDNA transcript characterization. Each expression value is normalized to a set of endogenous controls and calculated as a fold-change compared to *shGFP*, one of our controls targeting non-human sequences. Since *PGC-1α* induction and *LRPPRC* knockdown produce the expected mt-mRNA perturbations (Figure 2-1C, S2A-C), we have confidence that our dataset can identify novel regulators. By hierarchically clustering the average perturbational profile across hairpins for each of 107 gene knockdowns along each dimension, we can identify genes and mitochondrial transcripts that may have similar functions (Figure 2-2, Table 2-S3,S4). Unsupervised hierarchical clustering automatically groups *LRPPRC* and *SLIRP* together with *POLRMT*, which encodes the RNA polymerase responsible for mtDNA transcription, and *SSBP1*, which encodes a protein required for mtDNA replication (Figure 2-2, cluster C1). Separately, genes with roles in mtRNA precursor cleavage, including *ELAC2* and *MRPP1*, also form a strong group based on the signature of junction probe enrichment (Figure 2-2, cluster C3). Unsupervised hierarchical clustering of the MitoString probes results in automated segregation of light strand probes, heavy strand mRNAs, and heavy strand junctions (Figure 2-2). Principal component analysis further demonstrates the power of our dataset to differentiate these distinct transcript types (Figure 2-S1F). In particular, the heavy strand junction and mRNA probes are

Figure 2-2. MitoString profiles for 107 gene knockdowns
Heatmap depicts the log2(fold-change) expression level of each of 27 mtDNA probes across 107 gene knockdowns. mtDNA probes are noted along the top, and ordered by hierarchical clustering (Euclidean distance). Each gene was targeted by three distinct lentiviral hairpins and measured in duplicate. The mean of all hairpins for a given gene is displayed and the genes are hierarchically clustered (Euclidean distance). Red represents increased and blue represents decreased expression with respect to *shGFP*. Identifiers starting with 'TRN' denote tRNAs. Probe colors are as in Figure 2-1B.


27 mtDNA probes
mRNA probes
junction probes
noncoding probes
rc_ND5
rc_ND5_ND6
ND6
rc_COX1
rc_TRNC_TRNY
rc_TRNP
rc_TRNQ_TRNM
TRNF_RNR1
TRNV_RNR2
TRND_COX2
TRNM_ND2
ND4_TRNH
TRNK_ATP6/8
TRNL2_ND5
ATP6/8_COX3
RNR2_TRNL1
ND1
ND2
ND3
COX1
COX2
ND4
ND4L
ATP6
COX3
CYTB
ND5
Light Strand
Heavy Strand
107 gene knockdowns
SSBP1
POLRMT
SLIRP
LRPPRC
C1
PTCD2
AS3MT
EXOG
ACO2
C21orf33
MRM1
MTIF3
PUS1
APOBEC3G
DHX29
MRP63
MTERFD1
TRMT2B
FTSJ2
APEX2
ENDOG
MTG1
TUFM
PNPT1
MRPP3
DRG2
GFM2
FXC1
DLD
DUS2L
TDRKH
LSM4
CCBL2
PTRF
GFM1
GUF1
AKAP1
DDX28
RNMTL1
C6orf203
LSM3
SECISBP2
TFB2M
FASTKD2
GRSF1
HNRNPK
MRPS7
FASTKD4
MRPS17
MTPAP
C2
TACO1
TRNT1
TIMM9
RNASEL
AK3
TFB1M
MTIF2
PARK7
TIMM8B
NHP2
AUH
LOC119358
CTU1
RPUSD4
TIMM13
PTCD1
SPRYD4
C4orf114
HEMK1
FASTK
MTERF
NFXL1
ERAL1
TSHZ3
DNLZ
SND1
COMTD1
ASMTL
TRMU
ALKBH7
FASN
COQ3
TIMM10
RDBP
MTO1
PSMA6
DHX30
PTCD3
AFG3L2
MTERFD3
DAZAP1
AK3L2
MCTS1
COQ5
GLUD1
METT11D1
REXO2
EEFSEC
TIMM8A
TIMM50
AK3L1
RNASEH1
AK2
HDDC2
SUPV3L1
MRPP1
ELAC2
C3
MRPP2
>3
2
1
0
−1
−2
<−3
log2(fold-change)

separated by Principal Component 2, which explains 19% of the variance. The lone mRNA-mRNA junction, *ATP6/8_COX3*, is the sole outlier junction, as it behaves like its mature products, *ATP6/8* and *COX3*. These unsupervised analyses indicate that many genes targeted for RNAi are grouped by function and that mtDNA transcripts respond to perturbations in modules.

Knockdown of LRPPRC or SLIRP depletes all heavy strand mt-RNAs

By mining the compendium, we gain insight into the action of LRPPRC and SLIRP, which are the strongest regulators tested. Knockdown of *LRPPRC* or *SLIRP* results in depletion of the heavy strand precursor and all its resulting mt-mRNAs (Figure 2-2, cluster C1). In addition to *LRPPRC* and *SLIRP*, *POLRMT* and *SSBP1* also universally deplete heavy strand mt-RNAs when knocked down (Figure 2-2, C1), consistent with our expectation that depletion of both precursor and mature transcripts should reflect reduced transcription or mtDNA content. SLIRP and LRPPRC form a complex known to bind mitochondrial RNA and affect mt-mRNA levels (Mili and Piñol-Roma, 2003; Sasarman et al., 2010), but reports both supporting and contesting a role for LRPPRC in transcription exist (Harmel et al., 2013; Sondheimer et al., 2010). Although all heavy strand probes are impacted by *shLRPPRC* and *shSLIRP*, some junction probes are less strongly affected than the mRNA probes, suggesting a dual role for these regulatory proteins in post-transcriptional stabilization and transcription (Figure 2-1C). One theory consistent with our results is that the complex stabilizes the nascent transcript, as suggested elsewhere (Harmel et al., 2013). Because we could quantitate

junction probes strand-specifically, we were able to observe that *ND6* and most light strand transcripts are unaffected by *LRPPRC* or *SLIRP* knockdown.

MitoString highlights distinct regulatory mechanisms for coding and noncoding antisense transcripts

Querying the MitoString compendium for specific transcripts, such as the heavy strand mRNA *COX1*, uncovers distinct regulation for heavy strand coding and light strand noncoding RNAs. We identify nuclear genes whose knockdown alters *COX1* expression relative to the control hairpins, using a rank-sum statistic at a nominal p-value <0.01 (Figure 2-3A). On the left tail of the distribution, we recover the aforementioned *LRPPRC* and *SLIRP*. *COX1* is also depleted by knockdown of *FASTKD4* (also known as *TBRG4*), an uncharacterized protein containing a RAP domain with predicted RNA-binding abilities (Lee and Hong, 2004). Hairpins targeting the uncharacterized nucleoid component *DHX30* also deplete *COX1.* Interestingly, depletion of *SUPV3L1*, which is implicated in mtRNA degradation, also decreases *COX1* levels, possibly due to decreased mtDNA copy number, as described previously (Khidr et al., 2008). Silencing of *MTPAP* (mitochondrial poly-A polymerase) also diminishes *COX1* levels markedly, with the exception of one hairpin (Figure 2-3A). No knockdowns significantly increased *COX1* expression, perhaps due to its extremely high baseline abundance.

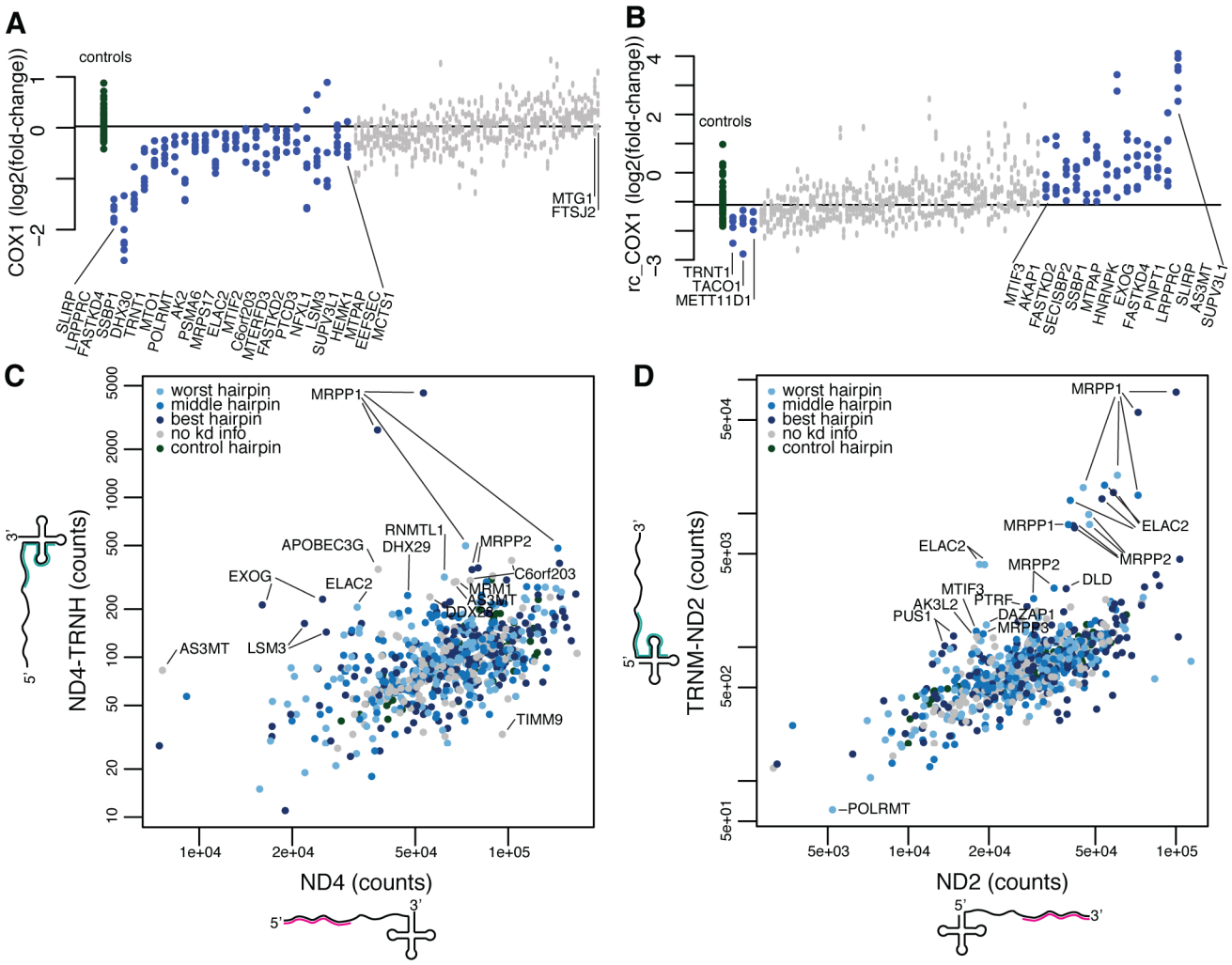


Figure 2-3. Discordant expression of junction and noncoding probes highlights genes required for processing and degradation

(A) Expression of the *COX1* probe as a log2(expression/*shGFP* expression) value is plotted for each control (green) and knockdown hairpin (blue denotes nominal p-value <0.01 by Mann-Whitney rank-sum test when comparing six gene-targeted hairpins to the control hairpins shown, grey p-value >0.01). The knockdown of *LRPPRC*, *SLIRP* and *FASTKD4* caused the strongest depletion of *COX1*. (B) Expression of the *rc_COX1* probe, which targets a noncoding region on the light strand is plotted as in (A). (C) The probe count for each knockdown hairpin was plotted for the *ND4* mRNA and *ND4_TRNH* junction probes. Genes that disproportionally affect one probe are found offset from the diagonal and labeled (Supplemental experimental procedures). Hairpins are colored based on knockdown strength (light blue for worst hairpin, blue for middle hairpin, dark blue for best hairpin, grey for no knockdown information, and green for control). (D) The probe count for each knockdown shRNA hairpin plotted for the *ND2* mRNA and *TRNM_ND2* junction probe as in (C)(see also Figure 2-S3).

A similar analysis, focusing on the light strand probe rc_COX1, reveals that the majority of significant knockdowns increase *rc_COX1* expression (Figure 2-3B). Since the noncoding portion of the light strand exists transiently, measuring these regions can identify degradation machinery. For the *rc_COX1* probe, knockdown of *SUPV3L1* induces a strong increase in transcript levels, confirming the role of *SUPV3L1* in this process (Figure 2-3B). However, *SUPV3L1* silencing actually decreases the level of some mt-mRNA transcripts, most notably *COX1*, *COX3*, and *ND4* (Figure 2-2, cluster C3), suggesting that its role in degradation may be selective. The response of *COX1* and *rc_COX1* transcripts to systematic RNAi perturbation highlights the distinct regulation of heavy strand and light strand transcripts shown in Figure 2-2.

Identifying negative regulators of mt-mRNA abundance

We next sought to identify factors that function to repress the abundance of mt-mRNAs at steady state, in contrast to LRPPRC and SLIRP, which stabilize transcripts, by further mining our compendium. Loss of *MTG1* (Mitochondrial GTPase 1) expression results in an increase of all mt-mRNAs except *ND5* (Figure 2-S2D), and it is the second strongest scoring gene on the right tail of the *COX1* distribution (Figure 2-3A). Although it is not significant when summarized by three hairpins for *COX1* individually (Figure 2-3A), measuring multiple genes results in an intriguing phenotype (Figure 2-S2D). *MTG1* has been linked to respiration and translation in human cell lines, but up to now, the effect of *MTG1* knockdown on mt-mRNA levels had not been measured (Barrientos et al., 2003; Kotani et al., 2013). Knockdown of the uncharacterized gene *C21orf33* increases the mt-mRNAs *ND1-3*, with excellent knockdown-phenotype correlation, but

has minimal effect on the other transcripts (Figure 2-S2E). In a complementary analysis using the GNF Mouse GeneAtlas (Lattin et al., 2008), mRNA expression of the *C21orf33* mouse homolog is well-correlated with nuclear-encoded OXPHOS gene expression (Figure 2-S2F). Our data offers tantalizing clues, but more research is required to elucidate the exact roles of these genes.

Identifying factors that cleave precursor transcripts

A key advantage of the MitoString approach is that it simultaneously monitors both precursor and mature transcripts across a battery of perturbations, allowing identification of specific cleavage and processing factors. Within our dataset, known cleavage factors form a distinct cluster marked by increased junction expression concomitant with stable mt-mRNA expression (Figure 2-2, cluster C3). These recently discovered factors include mitochondrial RNase P proteins 1 and 2 (*MRPP1* and *MRPP2*), which encode subunits of the tRNA 5' end cleavage machinery, and ELAC2, which cleaves the tRNA 3' end (Brzezniak et al., 2011; Holzmann et al., 2008; Lopez Sanchez et al., 2011).

By comparing the abundance of specific mt-mRNAs to their unprocessed precursor, over all perturbations, we observe a background distribution from which outliers represent candidate cleavage factors (Figure 2-3CD). The activity of mitochondrial RNase P is estimated by comparing the *ND4* transcript with its neighboring *ND4_TRNH* transcript (Figure 2-3C). If the junction is less efficiently cleaved, due to gene silencing, more unprocessed RNA will remain (Figure 2-3C). Hairpins targeting *MRPP1* and *MRPP2* have this effect and are consistently distinct

from the background distribution in this case and in others tested (Figure 2-3C, S3A-C). We investigate the 3' cleavage site similarly, by comparing *ND2* counts to *TRNM_ND2* counts (Figure 2-3D). In this case, we identify *ELAC2*, *MRPP1*, and *MRPP2*, suggesting that knockdown of RNase P components can disrupt 3' cleavage, as found previously for some junctions (Brzezniak et al., 2011; Lopez Sanchez et al., 2011).

We also interrogate junctions cleaved by non-canonical mechanisms. We find that silencing *MRPP1-3* or *ELAC2* does not influence the *ATP6/8_COX3* junction cleavage (Figure 2-2, cluster C3 and S3D-F). Further, outlier hairpins for this distribution are driven by mt-mRNA depletion, not *ATP6/8_COX3* transcript accumulation, leaving the cleavage factor for this junction unidentified (Figure 2-S3D-F). The 5' junction of *TRNQ* on the light strand abuts noncoding DNA and is unexpectedly more strongly affected by knockdown of *ELAC2* than knockdown of *MRPP1/2* (Figure 2-2, cluster C3). Conversely, *TRNF_RNR1* is more strongly regulated by silencing of *MRPP1/2* and *SUPV3L1* than silencing of *ELAC2*, although one *ELAC2* hairpin still has a strong effect (Figure 2-2, C3). Both cases suggest noncanonical processing.

Impact of known disease genes on mt-mRNA levels

Our knockdown inventory includes twelve genes implicated in Mendelian mitochondrial disease (Table 2-S5). Disease genes *MTO1*, *PUS1*, and *TRMU* encode enzymes with well-established roles in tRNA modification and efficient translation, but to our knowledge, their role in regulating human mt-mRNA abundance has never been tested (Patton et al., 2005; Wang et al., 2010b; Yan and Guan, 2004; Zeharia et al., 2009). For the strongest *MTO1* knockdown hairpin (13% remaining), our compendium

shows depletion of multiple mt-mRNAs (Figure 2-2), confirming similar findings in yeast (Wang et al., 2010b). Similarly, strong *PUS1* knockdown decreased *ATP6*, *COX3,* and *CYTB* mt-mRNA abundance (Figure 2-2). On the other hand, silencing of *TRMU* has a negligible effect on mt-mRNAs (Figure 2-2). In combination, these data suggest that some tRNA modifications may be necessary for proper mt-mRNA stability. Disease mechanisms may include both improper tRNA modification and mt-mRNA depletion.

Our compendium also expands our understanding of a number of other disease genes (Table 2-S5). A mutation in *FASTKD2* has been implicated in cytochrome *c* oxidase (COX) deficiency (Ghezzi et al., 2008), and we now find a 2-fold decrease of *COX2-3,CYTB, ATP6/8, and ND3-4/4L-5* mt-mRNA expression under *FASTKD2* silencing by at least one hairpin (Figure 2-2). PNPT1, which is implicated in both RNA degradation (Borowski et al., 2013) and *5S rRNA*, *RNase P* RNA and *MRP* RNA import (Wang et al., 2010a), has also been recently associated with hearing loss and respiratory chain deficiency (Vedrenne et al., 2012; von Ameln et al., 2012). Our data reveals moderate increases in short-lived transcripts *CYTB* and *ND2* as well as moderate decreases in long-lived transcripts *COX1-3* (Figure 2-2) in *PNPT1* silenced cells, consistent with previous reports and highlighting the complexity of this protein's role in the cell (Borowski et al., 2013). Knockdown of TIMM8A, a known inner-mitochondrial membrane translocase component targeted in the current study because it contains a zinc finger-like motif, is correlated with moderately decreased mt-mRNA levels for almost all transcripts (Figure 2-2), and has been previously implicated in deafness-dystonia syndrome (Aguirre et al., 2006; Swerdlow et al., 2004).

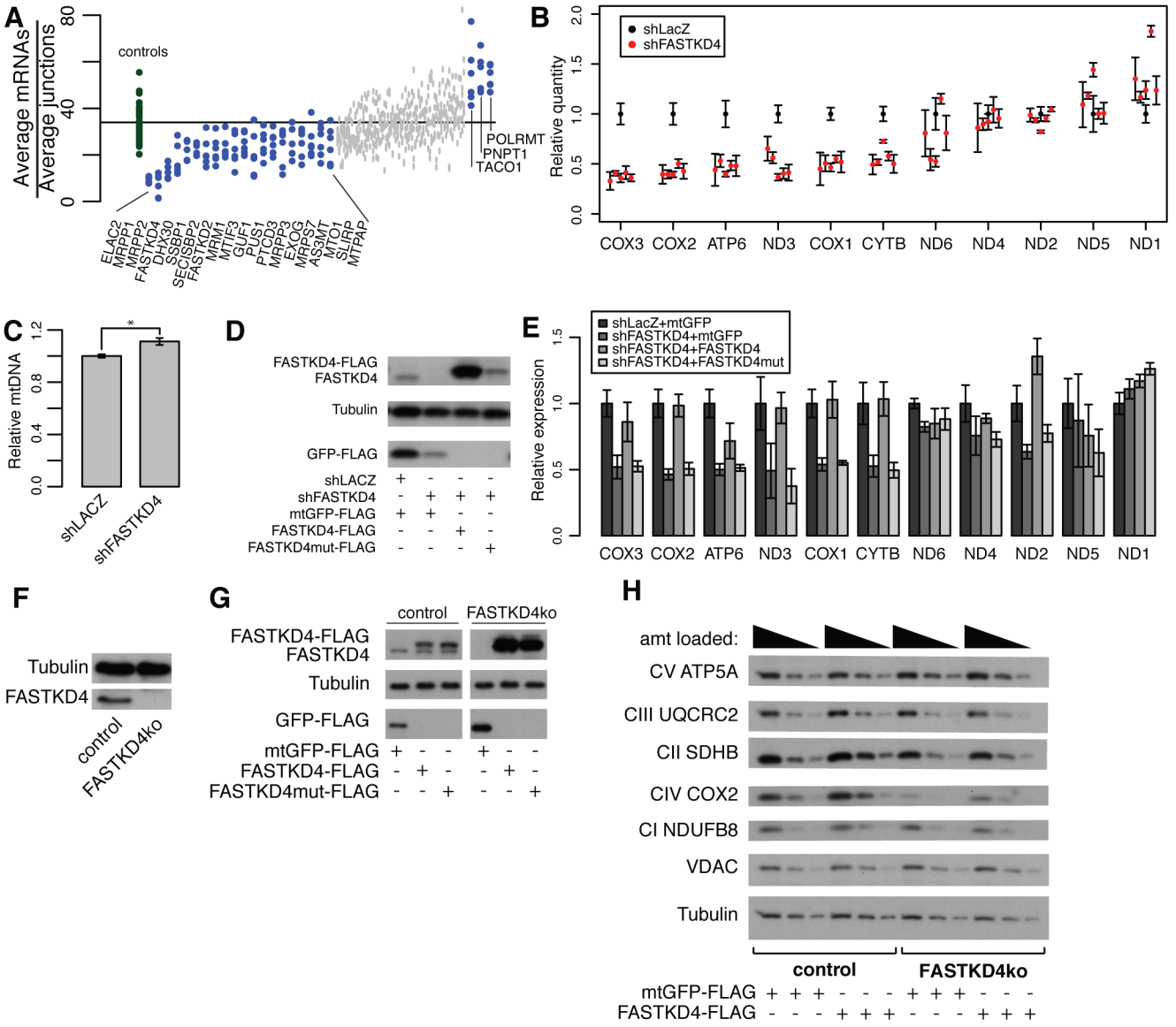


Figure 2-4. Loss of FASTKD4 leads to loss of a subset of steady-state mtDNA gene products

(A) The geometric mean of heavy strand mRNA probe counts divided by the geometric mean of heavy strand junction probe counts is plotted for each hairpin. Control hairpins are in green and knockdown hairpins are in blue or grey (blue represents nominal p-value <0.01 by Mann-Whitney rank-sum test). (B) MitoString results normalized to *shLacZ* for five distinct *shFASTKD4* hairpins in HEK-293T cells. Control hairpin shown in black, *FASTKD4* hairpins shown in red. (C) Relative mtDNA content in *shLACZ* knockdown cells compared to *shFASTKD4* knockdown cells as measured by qPCR, * indicates p <.05 (two-tailed unpaired t-test). (D) Western blot showing expression of FASTKD4-FLAG and FASTKD4 (detected by FASTKD4 antibody, top), tubulin (loading control, middle), and GFP-FLAG (overexpression control, bottom) in stable HEK-293T

Figure 2-4 (Continued)
cell lines. Cells are overexpressing RNAi resistant *FASTKD4-FLAG* and *FASTKD4mut-FLAG* (with RAP domain point mutations), or *mitoGFP-FLAG* in the presence of *shLACZ* (control) or *shFASTKD4* knockdown, as indicated. (E) Relative expression of each mt-mRNA probe quantified by MitoString, shown for the four cell lines depicted in (D). (F) Western blot showing expression of FASTKD4 and tubulin in the *FASTKD4ko* CRISPR/Cas9 cell line. (G) Western blot showing expression of FASTKD4-FLAG and FASTKD4 (detected by FASTKD4 antibody, top), tubulin (loading control, middle), and GFP-FLAG (overexpression control, bottom) in control and *FASTKD4ko* cell lines with overexpression of *FASTKD4-FLAG*, *FASTKD4mut-FLAG*, and *mtGFP-FLAG* as indicated (samples run on same gel with irrelevant lane removed). (H) Western blot showing respiratory complex protein expression in control and *FASTKD4ko* cell lines with *mtGFP-FLAG* or *FASTKD4-FLAG* overexpression as indicated. Three concentrations of each cell lysate was loaded as indicated. In all panels, error bars represent s.e.m., n=3.

Identifying genes that influence mature mt-mRNA abundance

To identify genes that specifically stabilize mature heavy strand mt-mRNAs, we developed an ordered list that incorporates all of our heavy strand probe data. We examined the ratio of the average heavy strand mature mt-mRNA probes to the average heavy strand precursor probes for each knockdown gene (Figure 2-4A). Using this measurement, we ordered each gene based on the rank-sum of all hairpins targeting that gene (Figure 2-4A). This procedure is complementary to Figure 2-3CD, which focuses on two individual junctions. Proteins identified through this method might play a role in post-transcriptional processing, stability, and degradation.

POLRMT, which encodes the mitochondrial mtRNA polymerase, is amongst the highest scoring hits, causing the ratio of mature mt-mRNA to precursor RNA to increase when silenced (Figure 2-4A). POLRMT is well-studied *in vitro*, but to our knowledge, our study is the first to perturb it *in vivo*. This result suggests that when mtDNA transcription

is reduced, the cell's response is to stabilize mature mt-mRNAs, via an uncharacterized regulatory mechanism.

A number of genes, when silenced, decrease the ratio of mt-mRNAs to precursor transcripts, suggesting that either the nascent strand is stabilized and/or the mt-mRNAs are destabilized. As expected, RNase P- and RNase Z-encoding genes (*MRPP1*, *MRPP2*, *ELAC2*) are represented here because they stabilize nascent transcripts. Knockdown of an uncharacterized gene, FASTKD4, destabilizes mature transcripts and has an effect size similar to that produced by knockdown of known regulators (Figure 2-4A). *FASTKD4* clusters in our study with *MTPAP* and the ribosomal protein-encoding *MRPS17* (Figure 2-2, cluster C2). In general, silencing of genes in this cluster affects the mt-mRNA probes more strongly than the junction probes and does not affect all mt-mRNAs uniformly. As FASTKD4 has not been previously linked to mt-mRNA expression, we pursued its function in more depth.

Silencing of FASTKD4 depletes some, but not all, mt-mRNAs

FASTKD4 belongs to a metazoan-specific cohort of proteins, defined by the presence of FAST (Fas-activated serine/threonine kinase-like) and RAP (RNA-binding domain abundant in Apicomplexans) domains. The human genome encodes six of these proteins (FASTK, FASTKD1-5), all of which localize to the mitochondrion (Simarro et al., 2010), although not all were identified in the initial MitoCarta compendium (Pagliarini et al., 2008). Mutations in *FASTKD2* are associated with cytochrome *c* oxidase deficiency (Ghezzi et al., 2008) and *FASTKD3* knockdown inhibits respiration in cultured cells (Simarro et al., 2010). *FASTK* encodes a kinase, whose mitochondrial

localization is disrupted during UV- or anti-Fas antibody-induced apoptosis (Li et al., 2004). FASTK also plays a role in alternative splicing in the nucleus, which requires the RAP domain (Izquierdo and Valcárcel, 2007; Simarro et al., 2007). Unlike FASTK, FASTKD4 does not have an identifiable kinase active site.

To confirm our screening result that *shFASTKD4* depletes a subset of mt-mRNAs, we began by validating the *FASTKD4* knockdown phenotype in an independent cell line (HEK-293T). We used five distinct hairpins against *FASTKD4* and used MitoString to measure expression of the mitochondrial transcripts after eight days of stable knockdown (Figure 2-4B). In agreement with our initial screen, only a subset of the mt-mRNA transcripts were affected (*COX1-3*, *ATP6/8*, *CYTB*, *ND3*), and we show that mtDNA is actually increased, ruling out mtDNA depletion as a mechanism (Figure 2-4C). *ND1* was also consistently upregulated in both cases, while *ND2*, *ND4/4L* and *ND5* remained unchanged. Thus, the impact of *FASTKD4* silencing on mt-mRNA abundance is robust across five distinct hairpin sequences and two cell lines. To confirm that this is an on-target effect of *FASTKD4* knockdown, we overexpressed a FLAG-tagged RNAi-resistant *FASTKD4* (*FASTKD4-FLAG*), as well as *FASTKD4* with four characteristic RAP domain residues mutated to alanines (*FASTKD4mut-FLAG*) (Figure 2-4DE). *FASTKD4-FLAG* expression, but not *FASTKD4mut-FLAG* expression, is accompanied by recovery in RNA levels for affected transcripts (Figure 2-4E), proving that the *FASTKD4* knockdown is responsible for the observed phenotype.

We next used a complementary gene knockout strategy to perturb FASTKD4 and interrogate its impact on OXPHOS protein production. Although the knockdown

produced a reliable RNA phenotype, residual FASTKD4 RNA is translated. We generated a *FASTKD4* knockout (*FASTKD4ko*) in HEK-293T cells using the CRISPR/Cas9 system (Hsu et al., 2013), which enabled assays in the complete absence of FASTKD4. In our *FASTKD4ko* cell line, we identified three indels at the FASTKD4 locus that generate protein-terminating frame shifts, resulting in the absence of full-length protein (Figure 2-4F). In *FASTKD4ko*, we observe reduced abundance of complex IV subunit COX2 protein that is rescued by *FASTKD4-FLAG* overexpression (Figure 2-4GH). Complex I, II, III, and V nuclear subunits were unaffected, consistent with our screening results and the dependence of these complexes on mtDNA-encoded subunits (Figure 2-4H). Thus, two independent genetic strategies indicate that FASTKD4 is required for the proper expression of a specific subset of mitochondrial RNAs, and in the case of COX2, the protein product.

Stability of a subset of mt-RNA transcripts requires FASTKD4

Assuming all heavy strand transcripts are transcribed concurrently, we hypothesized that the differential abundance of mt-mRNAs following silencing of *FASTKD4* is due to differential degradation rates. To measure RNA stability within the mitochondrion, we blocked POLRMT-mediated transcription with media containing a high concentration of ethidium bromide (ETBR) (Nagao et al., 2008). We quantitated the mt-mRNAs using qPCR following six hours of ETBR treatment, during which time period mitochondrial transcription is suspended. The fraction of RNA remaining can be determined by comparing the ETBR-treated RNA to the untreated RNA. Of the six transcripts downregulated at steady state in FASTKD4 knockdown cells (Figure 2-4B),

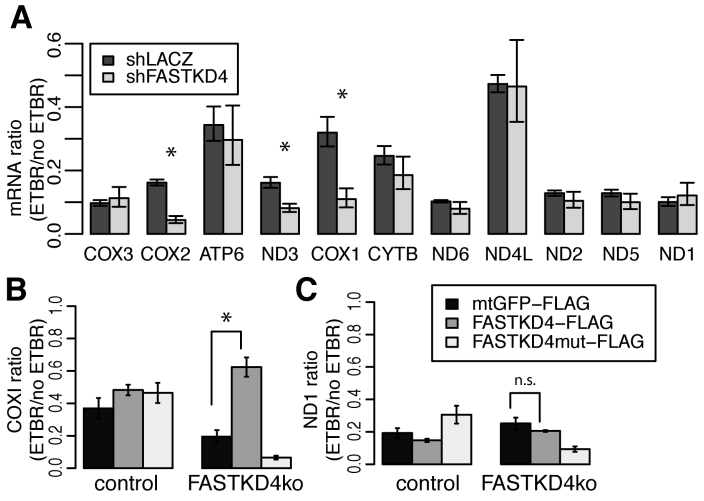


Figure 2-5. FASTKD4 is required for stability of a subset of mt-mRNAs

(A) The fraction of RNA remaining after ethidium-bromide (ETBR) transcription inhibition for 6-hours is plotted for each mt-mRNA as measured by qPCR in *shLACZ* and *shFASTKD4* cell lines. This fraction is also plotted specifically for (B) *COX1* and (C) *ND1* in control or knockout cell lines expressing FLAG-tagged *mtGFP*, *FASTKD4*, and *FASTKD4mut*. Error bars represent s.e.m., n=3. * represents p<=0.05 (two-tailed unpaired t-test).

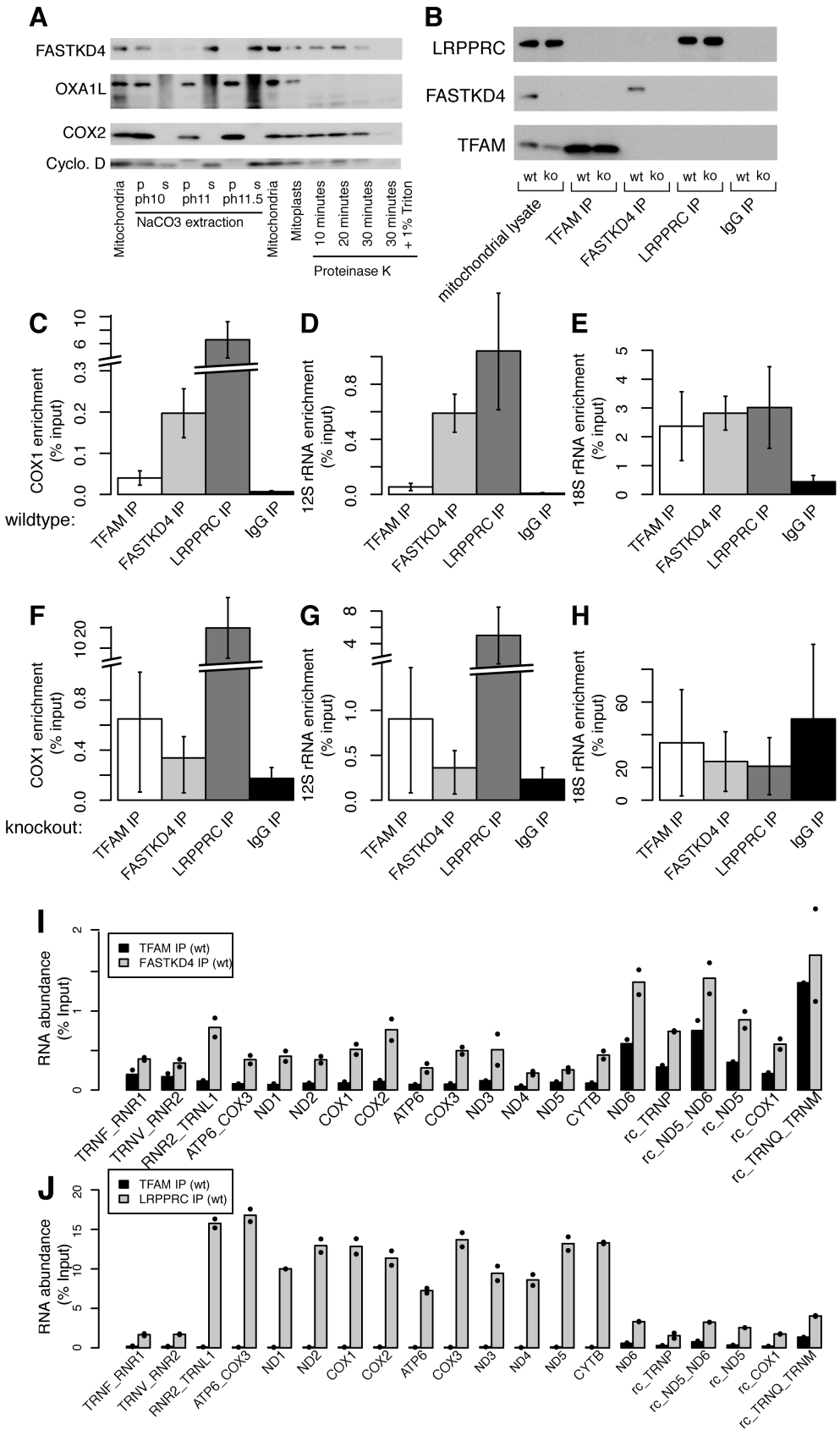
we find that *COX2*, *ND3*, and *COX1* also have a strong decrease in the fraction of RNA remaining under ETBR treatment (Figure 2-5A), suggesting an increased degradation rate. *CYTB* also trends downward, but not significantly. The only transcripts that are depleted (Figure 2-4B) but not destabilized in our assay are *COX3* and *ATP6/8* (Figure 2-5A): regulation of these two transcripts, which are part of the stable *ATP6/8_COX3* precursor, appears to be more complicated. By performing the same assay in our *FASTKD4ko*, we observed destabilized *COX1*, and unchanged *ND1*, consistent with *shFASTKD4* (Figure 2-5BC). Further, to rule out off-target RNAi effects, we rescued the *COX1* destabilization with *FASTKD4-FLAG* overexpression, but not with overexpression of *FASTKD4mut-FLAG* (Figure 2-5B).

FASTKD4 associates with mitochondrial RNAs in the matrix

We sought to define the sub-organellar localization of FASTKD4 relative to the mt-mRNAs. FASTKD4, alongside soluble protein Cyclophilin D, was extractable with

Figure 2-6. FASTKD4 is localized to the matrix and physically associates with mitochondrial RNAs

(A) Western blot showing the presence of FASTKD4, OXA1L, COX2, and Cyclophilin D in isolated mitochondria, with carbonate extraction, and with proteinase K digestion of mitoplasts. (B) Western blot showing presence of LRPPRC, FASTKD4, and TFAM after immunoprecipitation of each protein or IgG control in control (wt) and *FASTKD4ko* (ko) cells. (C-E) qPCR was used to measure enrichment of (C) *COX1* mRNA and (D) *12S* rRNA, and (E) *18S* rRNA with immunoprecipitation of TFAM, FASTKD4, LRPPRC, and IgG from wild-type cells. (F-H) qPCR-measured enrichment of transcripts measured in C-E with identical immunoprecipitation conditions, but in *FASTKD4ko* cells. Error bars represent s.e.m., n=3. (I) RNA abundance (% input) after FASTKD4 immunoprecipitation or TFAM immunoprecipitation in wild type cells measured by MitoString. (J) RNA abundance (as % input) after LRPPRC immunoprecipitation or TFAM immunoprecipitation in wild type cells. Points represent each duplicate.


A
FASTKD4
OXA1L
COX2
Cyclo. D
Mitochondria
p ph10
s
p ph11
s
p ph11.5
s
NaCO3 extraction
Mitochondria
Mitoplasts
10 minutes
20 minutes
30 minutes
30 minutes + 1% Triton
Proteinase K
B
LRPPRC
FASTKD4
TFAM
wt ko
mitochondrial lysate
TFAM IP
FASTKD4 IP
LRPPRC IP
IgG IP
C
COX1 enrichment (% input)
wildtype:
D
12S rRNA enrichment (% input)
E
18S rRNA enrichment (% input)
F
COX1 enrichment (% input)
knockout:
G
12S rRNA enrichment (% input)
H
18S rRNA enrichment (% input)
I
RNA abundance (% Input)
TFAM IP (wt)
FASTKD4 IP (wt)
TRNF_RNR1
TRNV_RNR2
RNR2_TRNL1
ATP6_COX3
ND1
ND2
COX1
COX2
ATP6
COX3
ND3
ND4
ND5
CYTB
ND6
rc_TRNP
rc_ND5_ND6
rc_ND5
rc_COX1
rc_TRNQ_TRNM
J
RNA abundance (% Input)
TFAM IP (wt)
LRPPRC IP (wt)

alkaline carbonate treatment of isolated mitochondria (Figure 2-6A). Additionally, FASTKD4 in mitoplasts was resistant to proteinase K treatment, alongside matrix protein Cyclophilin D (Figure 2-6A). All measured proteins are substrates for Proteinase K. Collectively, these studies indicate that FASTKD4 is a soluble protein residing within the mitochondrial matrix.

Because FASTKD4 appears to stabilize mt-mRNAs, we investigated its binding to mtRNA by immunoprecipitating endogenous FASTKD4 from mitochondrial lysates and testing for mitochondrial RNA enrichment using qPCR. As controls, we immunoprecipitated TFAM, a known DNA-binding protein which is not expected to bind to RNA, and LRPPRC, a known RNA-binding protein, alongside FASTKD4 in both wildtype and *FASTKD4ko* HEK-293T cells (Figure 2-6B). In agreement with published results (Chujo et al., 2012; Sasarman et al., 2010), LRPPRC is associated with *COX1* mRNA (Figure 2-6C). In addition, LRPPRC associates with *12S* rRNA but not nuclear *18S* rRNA (Figure 2-6DE). In comparison, TFAM is not appreciably enriched for either RNA (Figure 2-6C and 2-6D). Similar to LRPPRC, FASTKD4 immunoprecipitation enriches both *COX1* mRNA and *12S* rRNA, as compared to TFAM immunoprecipitation, in wild type cells, but does not enrich nuclear *18S* rRNA (Figure 2-6CDE). Intriguingly, FASTKD4 seems to bind *12S* almost as strongly as LRPPRC, while it is a weaker binder of *COX1*. In *FASTKD4ko* cells, RNA enriched from FASTKD4 immunoprecipitation is indistinguishable from that enriched for TFAM and IgG immunoprecipitation, proving that the RNA enrichment of *COX1* and *12S* rRNA is

specifically due to FASTKD4 immunoprecipitation and is not an antibody artifact (Figure 2-6FGH).

In addition to qPCR, we assessed our TFAM, FASTKD4, and LRPPRC immunoprecipitations with MitoString (Figure 2-6IJ). MitoString is valuable in this context as it allows strand-specific multiplexed measurement with minimal sample. We found that the majority of heavy strand abundant transcripts were enriched in both LRPPRC and FASTKD4 immunoprecipitations in comparison to TFAM immunoprecipitation (Figure 2-6IJ). In contrast, light strand probes targeting *ND6* and light strand intergenic space (*rc_ND5*, *rc_ND5_ND6*, and *rc_TRNQ_TRNM*) had less specific enrichment (Figure 2-6I). The heavy strand junction probes spanning the *RNR2_TRNL1* junction and the *ATP6/8_COX3* junction are enriched in both FASTKD4 and LRPPRC immunoprecipitations. These results for LRPPRC are well correlated with previous findings using qPCR (Chujo et al., 2012). In addition, our approach interrogates new junction transcripts, especially on the light strand, which do not appear enriched relative to TFAM immunoprecipitation. *RNR2_TRNL1*, which we measure for the first time, is enriched in both LRPPRC and FASTKD4 immunoprecipitations at the same level as many of the mRNA genes. Our distinct approach and the inclusion of TFAM as a negative control demonstrate that FASTKD4 associates with mitochondrial RNAs and further validates LRPPRC as a mitochondrial RNA-binding protein.

Discussion

The expression of mtDNA requires nuclear-encoded proteins, yet at present, we lack a full molecular description of this system. Previous approaches to mitochondrial

RNA measurement, including northern blots, qPCR, and next-generation sequencing, have been valuable, but are limited by lack of strand-specificity or sample throughput. In the current study, we establish a facile assay, MitoString, to systematically monitor precursor and mature mitochondrial RNA species following silencing of over 100 candidate nuclear-encoded factors. MitoString enables us to identify and classify mitochondrial RNA regulators based on their roles in transcription, cleavage, and stability. Through this compendium, we have expanded our understanding of previously characterized mitochondrial RNA processing proteins and implicated new players in this pathway, including the novel regulator FASTKD4.

Our approach, which measures both mature and precursor transcripts (Figure 2-1), implicates the LRPPRC-SLIRP complex in production or stability of the heavy strand precursor. Previous reports have implicated LRPPRC in promoting transcription (Liu et al., 2011; Sondheimer et al., 2010) or the stability of individual mt-mRNAs (Chujo et al., 2012; Ruzzenente et al., 2012). We find that silencing of *SLIRP* or *LRPPRC* primarily affects abundance of the heavy strand precursor, indicating a role for the complex in either transcription or stabilization of the precursor (Figure 2-2). This decrease is accompanied by a more pronounced downregulation of mature heavy strand mt-mRNAs, which may suggest a secondary role in mature transcript stability. In contrast, Gohil et al. (2010) measured precursors in *LRPPRC*-silenced cells by qPCR, which is not strand specific, and found no change in precursor abundance. Our new method assesses just the heavy strand and finds a 4-fold decrease in the same *TRNM_ND2* junction upon *LRPPRC* silencing (Figure 2-2). *LRPPRC* or *SLIRP* silencing has no effect on the light

strand mRNA *ND6* by our assay (Figure 2-2), consistent with strand-specific Northern blot measurements (Ruzzenente et al., 2012) and in contrast with strand-insensitive qPCR results (Gohil et al., 2010). Recently, GRSF1 has been implicated as a binder and regulator of *ND6* and its precursor strand, suggesting a distinct mechanism operates on the light strand (Antonicka et al., 2013; Jourdain et al., 2013). Additionally, we and others find that LRPPRC binds heavy strand transcripts at a higher abundance and specificity than light strand transcripts (Chujo et al., 2012), which is consistent with our steady-state mtRNA findings.

We confirm the role of proteinaceous RNase P in the cleavage of tRNA 5' junctions and identify noncanonical cleavage sites. The MitoString profiles of RNase P subunits *MRPP1* and *MRPP2* knockdowns are well correlated, while the effect of *MRPP3* is weaker (Figure 2-2,3CD). Our comprehensive approach builds on earlier findings focused on *MRPP1* and a subset of junctions, (Holzmann et al., 2008) as well as work focused on *MRPP1 and MRPP3* (Lopez Sanchez et al., 2011). We also identified a few exceptions to this canonical cleavage process. On the light strand, *ELAC2* plays a role in the 5' cleavage of *TRNQ,* which excises it from surrounding noncoding DNA (Figure 2-2, cluster C3). Also, *TRNF_RNR1*, which we expect to be regulated by ELAC2, is more strongly regulated by silencing of *MRPP1/2* and *SUPV3L1* (Figure 2-2, cluster C3). Consistent with this, this same *TRNF_RNR1* junction was previously reported to be unaffected by *ELAC2* siRNA knockdown (Brzezniak et al., 2011). Sequencing results had suggested that knockdown of *ELAC2* and RNase P components did not affect all tRNA junctions (Lopez Sanchez et al., 2011), but for all of

our tested junctions on the heavy strand (excepting *ATP6/8_COX3*), either one or both were involved (Figure 2-2). The discrepancy may stem from the prior study's exclusion of large precursor fragments due to size selection for tRNA-sized transcripts (Lopez Sanchez et al., 2011). Our assay found no obvious candidate for processing of the *ATP6/8_COX3* junction, as all outlier hairpins are driven by a decrease in mt-mRNA levels, not an increase in junction content (Figure 2-S3D-F).

Although SUPV3L1 clearly plays a role in the degradation of light strand transcripts, we find it has contrasting effects on different classes of mt-mRNAs (Figure 2-2). The short-lived *ND1* and *ND2* transcripts are increased, whereas the long-lived *COX1-3* transcripts are actually decreased, suggesting that SUPV3L1 may not be responsible for all degradation within the mitochondria. These contrasting effects have been previously demonstrated (Szczesny et al., 2010), but only for a subset of mt-mRNAs. Here we obtained data for all transcripts simultaneously. For the *TRNF_RNR1* and *TRNV_RNR2* junctions, which we expect to be a readout for transcriptional initiation, hairpins targeting *SUPV3L1* are the second strongest inducers of expression (Figure 2-2,C3). This increased transcription may pertain only to early termination transcripts, which end after the second rRNA, since the *RNR2_TRNL1* junction and other heavy strand junctions are not increased (Figure 2-2, cluster C3). Therefore, *SUPV3L1* depletion may increase ribosome production.

We find that FASTKD4 regulates a specific subset of mt-mRNAs. *FASTKD4* silencing in general depletes those transcripts that are long-lived while sparing the precursor strand (Figure 2-4B). The sole exceptions are *ND3*, a short-lived depleted

transcript, and *ND4/4L*, a long-lived unaffected transcript. Most of the affected mt-mRNAs also have an increased degradation rate in both *FASTKD4* knockdown or knockout cells (Figure 2-5AB). Thus, we hypothesize that FASTKD4 may be serving to shield a specific set of mt-mRNAs from degradation by SUPV3L1 or an as-yet unidentified degradation machinery. *ATP6* and *COX3* unexpectedly do not have an increased degradation rate by our assay (Figure 2-5A), however the *ATP6/8_COX3* precursor is the sole mtDNA gene-gene junction and is relatively stable compared to other precursors, as estimated by probe counts. The factor responsible for *ATP6/8_COX3* cleavage remains undetermined and we expect a full understanding of this mechanism will resolve the discrepancy we observe.

We cement a role for FASTKD4 in mitochondrial RNA metabolism based upon the above findings, past studies and our immunoprecipitation of FASTKD4-associated mt-RNA. FASTKD4 contains a RAP domain in common with experimentally-validated RNA-binding proteins (Lee and Hong, 2004) and is found in a genome-wide mRNA-binding resource (Castello et al., 2012). Here, we show that FASTKD4 is localized to the mitochondrial matrix (Figure 2-6A) and associates with mitochondrial RNAs in a specific way (Figure 2-6C-I). Our immunoprecipitation assay does not distinguish among the heavy-strand transcripts in terms of strength of association with FASTKD4, because this protein associates with all heavy-strand transcripts. However, mtRNA susceptibility to degradation may be influenced in part by the activity of another, as yet unidentified protein that requires FASTKD4 binding.

Why the mitochondrion maintains mt-mRNAs at such disparate levels is unknown, but we propose that it may provide a mechanism for ensuring proper OXPHOS complex stoichiometry. Across tissues, the molar ratios of different complexes is robust, but the mechanism that generates this stability is unknown (Benard et al., 2006). Complex I is consistently the least abundant among the OXPHOS complexes (Lenaz and Genova, 2010), and with the exception of the bicistronic *ND4/4L* transcript, mtDNA-encoded Complex I transcripts have half-lives of less than 90 minutes (Nagao et al., 2008). In contrast, Complex IV and V are present at 3-12 times the concentration of Complex I, a stoichiometry that we suggest could be metered in part by higher levels of Complex IV and V component transcripts (*COX1-3*, *ATP6/8*), whose half-lives range 138-231 minutes. Thus, the precise regulation of mt-mRNA transcript levels by factors like FASTKD4 and SUPV3L1 may be important for ensuring proper OXPHOS complex stoichiometry and function.

In the future we anticipate that MitoString may prove useful for interrogating the role of mt-mRNA processing directly in tissues from patients with mitochondrial disease. Our current study includes the analysis of 12 previously established disease genes (Table 2-S5). Of these genes, three had previously been linked to aberrant expression of mt-mRNAs in humans. Our current compendium demonstrates that *MTO1*, *PUS1*, and *FASTKD2* loss of function may also cause aberrant expression of mitochondrial RNAs. Our results point to additional modes of pathogenesis that may be at play in these disorders. As new genetic variants in mitochondrial RNA processing factors are identified in patients, a key challenge will lie in proving their pathogenicity. MitoString is

a highly quantitative and systematic approach for measuring mitochondrial gene expression that in principle can be easily applied to patient biopsy specimens. MitoString may therefore be a useful companion diagnostic for detecting aberrant expression of the mtDNA genome in human diseases.

Experimental Procedures

Cell culture

WI-38 fibroblasts and HEK-293T cells were cultured in high glucose DMEM (Invitrogen Cat. No. 11995) supplemented with 10% fetal bovine serum (Sigma Cat. No. F6178) and 1X penicillin, streptomycin, and glutamine (Invitrogen Cat. No. 10378-016) at 37° C with 5% CO_2.

Immunoblotting

Protein was electrophoresed on a 12% Novex Tris-Glycine polyacrylamide gel at constant voltage (125V). The gel was transferred to a polyvinylidene difluoride membrane (Trans-Blot Turbo Transfer System). Membranes were blocked for one hour at room temperature in tris-buffered saline with 0.1% Tween-20 and 5% BSA (TBST-BSA). Membranes were incubated with primary antibody in TBST-BSA overnight at 4°C (Table 2-S6). Secondary antibody was used at (1:5,000) for one hour at room temperature. Membranes were developed using WesternLightning Plus-ECL.

Screen

The three best hairpins per gene were selected, and produced and arrayed by the RNAi consortium (TRC) as previously described (Moffat et al., 2006).

WI-38 cells were seeded, infected after 24-hours, and selected 24-hours later. 6 days post-infection the cells were lysed with RLT buffer (Invitrogen) and β-Mercaptoethanol (1:100) and frozen. RLT lysates were processed with a custom MitoString codeset from NanoString technologies per manufacturer's instructions (Geiss et al., 2008). Screen details and data processing were as described in supplemental experimental procedures.

Protein expression

plexFLAG was generated from plex-TRC-983 (TRC) by molecular cloning. pDONR223-FASTKD4 from a previously described cDNA library (Pagliarini et al., 2008) was mutagenized via QuikChange mutagenesis (Agilent) using the primers described (Table 2-S7). mtGFP was a gift from Y. Sancak. All cDNAs were cloned into plexFLAG via the Gateway system (Invitrogen).

Lentiviral production and infection

For follow-up experiments, shRNA lentiviral vectors (TRC) and FLAG-tagged expression constructs were used to produce virus and infect HEK-293T cells as previously described (Baughman et al., 2009). 4 ug/ml puromycin or 5 ug/ml blasticidin was used for selection.

mtDNA quantitation

mtDNA quantification was as previously described (Baughman et al., 2009)

Mitochondrial Isolation / Na_2CO_3 extraction / Proteinase K protection

HEK-293T cells were suspended in isolation buffer (220 mM mannitol, 70 mM sucrose, 5mM HEPES, 1 mM EGTA, pH 7.4 + cOmplete, EDTA-free Protease Inhibitor Cocktail

(Roche)) and lysed using a Cell Disruption Vessel (Parr Instrument Company) pressurized to 800 psi with Nitrogen followed by homogenization with five strokes of a Teflon Potter Elvehjem homogenizer. Crude mitochondria were isolated by differential centrifugation at 600g and 8000g. Mitoplasts were created by swelling with H_2O on ice and were stabilized in 137 mM KCl, 2.5mM $MgCl_2$, 10 mM HEPES pH 7.2, pelleted at 8000g and resuspended in isolation buffer. Na_2CO_3 extraction was performed on crude mitochondria and proteinase K protection was performed on mitoplasts as previously described (Ryan et al., 2001).

Measurement of RNA half-lives

HEK-293T cell lines were seeded the day before exposure to media with and without 1 ug/ml ethidium bromide. After six hours, the cells were lysed with RLT (Qiagen) for RNA extraction.

RNA isolation and qPCR

RNA was isolated using the RNeasy protocol with DNase I treatment (Qiagen). cDNA was transcribed using Superscript III First-Strand Synthesis Supermix (Invitrogen). qPCR was performed with a 7500-FAST ABI instrument and 2X Taqman FAST Advanced master-mix (Applied Biosystems) with mt-mRNA probes as described previously (Gohil et al., 2010).

RNA Immunoprecipitation

Isolated crude mitochondria were lysed in lysis buffer (50 mM HEPES (pH 7.4), 150 mM NaCl, 2 mM EDTA, 1% Triton, RNase Inhibitor). Lysates were treated with Turbo DNase for 1 hour at 37C. After spinning at 16000g 10 min at 4C, the supernatant was

precleared with Protein A agarose beads for 40 min at 4C with rotation. The beads were spun out and the supernatant incubated with antibody overnight at 4C with rotation. Fresh Protein A agarose beads were blocked overnight with 1% BSA, 100 ug/ml Heparin, 100 ug/ml yeast tRNA in Lysis buffer. The next day, the beads were washed 3X with lysis buffer, and added to the lysate-antibody mix, and incubated with rotation for 2 hours at 4C. The bead complex was washed 3X with lysis buffer, and 3X with 500 mM NaCl lysis buffer. RNA samples were eluted with 300 ul 1M NaHCO3, 1% SDS at room temperature for 15 minutes. Beads were spun down and the supernatant was added to 12 ul 5M NaCl. 700 ul of RLT buffer was added for RNA isolation. Protein samples were boiled in sample buffer for Immunoblotting.

CRISPR knockout cell line generation

sgRNA and U6 primers (Table 2-S7) were used to amplify the pX064 plasmid (gift from F. Zhang), as previously described (Hsu et al., 2013). This PCR product was cotransfected with the Cas9 plasmid (gift from F. Zhang) into HEK-293T cells with XTreme Gene transfection reagent per manufacturer's instructions. Cells were single-cell cloned and screened by Western blot for FASTKD4 protein. The identified knockout was verified by PCR followed by subcloning and Sanger sequencing.

Acknowledgements

We thank J. Baughman, X.R. Bao, S. Calvo, Y. Sancak, L. Strittmatter, I. Jain, N. Delaney, E. Kovacs-Bogdan, A. Li, S. Vafai, M. Staller and N.M. Cabili for comments and helpful discussions. We thank the Broad Institute RNAi Platform for shRNA reagents and S. Silver and S. Gopal for advice; R. Boykin and G. Geiss of Nanostring

Technologies for technical assistance; the Regev and Shamji groups for access to the nCounter Analysis system. We thank M. Guttman, D. Shechner, J. Rinn, D. Scott, and F. Zhang for guidance on experimental protocols. This work was supported by an NSF graduate research fellowship to A.R.W. and an NIH grant (GM077465) to V.K.M.

References

Aguirre, L.A., del Castillo, I., Macaya, A., Medá, C., Villamar, M., Moreno-Pelayo, M.A., and Moreno, F. (2006). A novel mutation in the gene encoding TIMM8a, a component of the mitochondrial protein translocase complexes, in a Spanish familial case of deafness-dystonia (Mohr-Tranebjaerg) syndrome. Am J Med Genet A *140*, 392-397.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F.*, et al.* (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457-465.

Antonicka, H., Sasarman, F., Nishimura, T., Paupe, V., and Shoubridge, E.A. (2013). The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. Cell Metab *17*, 386-398.

Barrientos, A., Korr, D., Barwell, K.J., Sjulsen, C., Gajewski, C.D., Manfredi, G., Ackerman, S., and Tzagoloff, A. (2003). MTG1 codes for a conserved protein required for mitochondrial translation. Mol Biol Cell *14*, 2292-2302.

Baughman, J.M., Nilsson, R., Gohil, V., Arlow, D.H., Gauhar, Z., and Mootha, V. (2009). A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. PLoS Genetics *5*, e1000590.

Benard, G., Faustin, B., Passerieux, E., Galinier, A., Rocher, C., Bellance, N., Delage, J.P., Casteilla, L., Letellier, T., and Rossignol, R. (2006). Physiological diversity of mitochondrial oxidative phosphorylation. Am J Physiol Cell Physiol *291*, C1172-1182.

Borowski, L.S., Dziembowski, A., Hejnowicz, M.S., Stepien, P.P., and Szczesny, R.J. (2013). Human mitochondrial RNA decay mediated by PNPase-hSuv3 complex takes place in distinct foci. Nucleic Acids Res *41*, 1223-1240.

Brzezniak, L.K., Bijata, M., Szczesny, R.J., and Stepien, P.P. (2011). Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. RNA biology *8*, 616-626.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M.*, et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Chujo, T., Ohira, T., Sakaguchi, Y., Goshima, N., Nomura, N., Nagao, A., and Suzuki, T. (2012). LRPPRC/SLIRP suppresses PNPase-mediated mRNA decay and promotes polyadenylation in human mitochondria. Nucleic Acids Res *40*, 8033-8047.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.*, et al.* (2008). The Pfam protein families database. Nucleic Acids Res *36*, D281-288.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T.*, et al.* (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol *26*, 317-325.

Gelfand, R., and Attardi, G. (1981). 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.

Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V., Elpeleg, O., and Zeviani, M. (2008). FASTKD2 nonsense mutation in an infantile mitochondrial encephalomyopathy associated with cytochrome c oxidase deficiency. Am J Hum Genet *83*, 415-423.

Gohil, V.M., Nilsson, R., Belcher-Timme, C.A., Luo, B., Root, D.E., and Mootha, V.K. (2010). Mitochondrial and nuclear genomic responses to loss of LRPPRC expression. J Biol Chem *285*, 13742-13747.

Harmel, J., Ruzzenente, B., Terzioglu, M., Spahr, H., Falkenberg, M., and Larsson, N.G. (2013). The leucine-rich pentatricopeptide repeat-containing protein (LRPPRC) does not activate transcription in mammalian mitochondria. J Biol Chem *288*, 15510-15519.

Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell *135*, 462-474.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Izquierdo, J.M., and Valcárcel, J. (2007). Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. J Biol Chem *282*, 1539-1543.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab *17*, 399-410.

Khidr, L., Wu, G., Davila, A., Procaccio, V., Wallace, D., and Lee, W.H. (2008). Role of SUV3 helicase in maintaining mitochondrial homeostasis in human cells. J Biol Chem *283*, 27064-27073.

Kotani, T., Akabane, S., Takeyasu, K., Ueda, T., and Takeuchi, N. (2013). Human G-proteins, ObgH1 and Mtg1, associate with the large mitochondrial ribosome subunit and are involved in translation and assembly of respiratory complexes. Nucleic Acids Res *41*, 3713-3722.

Lattin, J.E., Schroder, K., Su, A.I., Walker, J.R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C.K., Hume, D.A., Kellie, S.*, et al.* (2008). Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res *4*, 5.

Lee, I., and Hong, W. (2004). RAP--a putative RNA-binding domain. Trends Biochem Sci *29*, 567-570.

Lenaz, G., and Genova, M.L. (2010). Structure and organization of mitochondrial respiratory complexes: a new understanding of an old subject. Antioxid Redox Signal *12*, 961-1008.

Li, W., Simarro, M., Kedersha, N., and Anderson, P. (2004). FAST is a survival protein that senses mitochondrial stress and modulates TIA-1-regulated changes in protein expression. Mol Cell Biol *24*, 10718-10732.

Liu, L., Sanosaka, M., Lei, S., Bestwick, M.L., Frey, J.H., Surovtseva, Y.V., Shadel, G.S., and Cooper, M.P. (2011). LRP130 protein remodels mitochondria and stimulates fatty acid oxidation. J Biol Chem *286*, 41253-41264.

Lopez Sanchez, M.I.G., Mercer, T.R., Davies, S.M.K., Shearwood, A.J., Nygård, K.K.A., Richman, T.R., Mattick, J.S., Rackham, O., and Filipovska, A. (2011). RNA processing in human mitochondria. Cell Cycle *10*, 2904-2916.

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

Mili, S., and Piñol-Roma, S. (2003). LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. Mol Cell Biol *23*, 4972-4982.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K.*, et al.* (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. Cell *124*, 1283-1298.

Nagao, A., Hino-Shigi, N., and Suzuki, T. (2008). Measuring mRNA decay in human mitochondria. Methods Enzymol *447*, 489-499.

Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470-474.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K.*, et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112-123.

Patton, J.R., Bykhovskaya, Y., Mengesha, E., Bertolotto, C., and Fischel-Ghodsian, N. (2005). Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation

in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation. J Biol Chem *280*, 19823-19828.

Piechota, J., Tomecki, R., Gewartowski, K., Szczesny, R., Dmochowska, A., Kudła, M., Dybczyńska, L., Stepien, P.P., and Bartnik, E. (2006). Differential stability of mitochondrial mRNA in HeLa cells. Acta Biochim Pol *53*, 157-168.

Puranam, R.S., and Attardi, G. (2001). The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. Mol Cell Biol *21*, 548-561.

Ruzzenente, B., Metodiev, M.D., Wredenberg, A., Bratic, A., Park, C.B., Camara, Y., Milenkovic, D., Zickermann, V., Wibom, R., Hultenby, K.*, et al.* (2012). LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. EMBO J *31*, 443-456.

Ryan, M.T., Voos, W., and Pfanner, N. (2001). Assaying protein import into mitochondria. Methods Cell Biol *65*, 189-215.

Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T., Shoubridge, E.A., and Consortium, L. (2010). LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria. Mol Biol Cell *21*, 1315-1323.

Scarpulla, R.C., Vega, R.B., and Kelly, D.P. (2012). Transcriptional integration of mitochondrial biogenesis. Trends Endocrinol Metab *23*, 459-466.

Simarro, M., Gimenez-Cassina, A., Kedersha, N., Lazaro, J.B., Adelmant, G.O., Marto, J.A., Rhee, K., Tisdale, S., Danial, N., Benarafa, C.*, et al.* (2010). Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. Biochem Biophys Res Commun *401*, 440-446.

Simarro, M., Mauger, D., Rhee, K., Pujana, M.A., Kedersha, N.L., Yamasaki, S., Cusick, M.E., Vidal, M., Garcia-Blanco, M.A., and Anderson, P. (2007). Fas-activated serine/threonine phosphoprotein (FAST) is a regulator of alternative splicing. Proc Natl Acad Sci U S A *104*, 11370-11375.

Sondheimer, N., Fang, J.-K., Polyak, E., Falk, M.J., and Avadhani, N.G. (2010). Leucine-Rich Pentatricopeptide-Repeat Containing Protein Regulates Mitochondrial Transcription. Biochemistry *49*, 7467-7473.

Swerdlow, R.H., Juel, V.C., and Wooten, G.F. (2004). Dystonia with and without deafness is caused by TIMM8A mutation. Adv Neurol *94*, 147-154.

Szczesny, R.J., Borowski, L.S., Brzezniak, L.K., Dmochowska, A., Gewartowski, K., Bartnik, E., and Stepien, P.P. (2010). Human mitochondrial RNA turnover caught in flagranti: involvement of hSuv3p helicase in RNA surveillance. Nucleic Acids Res *38*, 279-298.

Vedrenne, V., Gowher, A., De Lonlay, P., Nitschke, P., Serre, V., Boddaert, N., Altuzarra, C., Mager-Heckel, A.M., Chretien, F., Entelis, N.*, et al.* (2012). Mutation in PNPT1, which encodes a polyribonucleotide nucleotidyltransferase, impairs RNA import into mitochondria and causes respiratory-chain deficiency. Am J Hum Genet *91*, 912-918.

von Ameln, S., Wang, G., Boulouiz, R., Rutherford, M.A., Smith, G.M., Li, Y., Pogoda, H.M., Nürnberg, G., Stiller, B., Volk, A.E.*, et al.* (2012). A mutation in PNPT1, encoding mitochondrial-RNA-import protein PNPase, causes hereditary hearing loss. Am J Hum Genet *91*, 919-927.

Wagner, B.K., Kitami, T., Gilbert, T.J., Peck, D., Ramanathan, A., Schreiber, S.L., Golub, T.R., and Mootha, V.K. (2008). Large-scale chemical dissection of mitochondrial function. Nat Biotechnol *26*, 343-351.

Wang, G., Chen, H.W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., McCaffery, J.M.*, et al.* (2010a). PNPASE regulates RNA import into mitochondria. Cell *142*, 456-467.

Wang, X., Yan, Q., and Guan, M.X. (2010b). Combination of the loss of cmnm5U34 with the lack of s2U34 modifications of tRNALys, tRNAGlu, and tRNAGln altered mitochondrial biogenesis and respiration. J Mol Biol *395*, 1038-1048.

Yan, Q., and Guan, M.X. (2004). Identification and characterization of mouse TRMU gene encoding the mitochondrial 5-methylaminomethyl-2-thiouridylate-methyltransferase. Biochim Biophys Acta *1676*, 119-126.

Zeharia, A., Shaag, A., Pappo, O., Mager-Heckel, A.M., Saada, A., Beinat, M., Karicheva, O., Mandel, H., Ofek, N., Segel, R.*, et al.* (2009). Acute infantile liver failure due to mutations in the TRMU gene. Am J Hum Genet *85*, 401-407.

Chapter 3
Evaluating the pathogenicity of mitochondrial gene variants using MitoString

Contributors: Ashley R. Wolf, Daniel S. Lieber, Nancy G. Slate, Steven Hershman, Sarah E. Calvo, Katherine B. Sims, Vamsi K. Mootha

Summary

Despite the widespread use of next-generation sequencing to identify novel variants in clinical cases, providing molecular diagnoses to patients with mitochondrial disease is challenging. This challenge derives from two issues: (1) the diagnosis of mitochondrial disease is challenging due to phenotypic and genetic heterogeneity and (2) assigning pathogenicity to variants requires both enrichment in patients and laboratory complementation experiments. In the case presented here, we have a patient with a suspected mitochondrial disorder and no family history, who was found to have variants of unknown functional significance in *SUPV3L1*, the gene encoding a helicase involved in mitochondrial RNA degradation. No mutations in *SUPV3L1* had been previously linked to disease. In the laboratory, we used MitoString, a technique for interrogating mitochondrial RNA transcripts in parallel, to probe patient fibroblasts. We found that the patient fibroblasts did not look appreciably different from control fibroblasts, while genetic disruption of *SUPV3L1* resulted in accumulation of mitochondrial DNA light strand RNA transcripts. Thus, we found no evidence for pathogenicity of the *SUPV3L1* variants in our fibroblast model. Subsequently, other

researchers identified pathogenic variants in a non-mitochondrial gene, N-glycanase 1 (*NGLY1*), which accounted for the patient's features. Our study underlines the challenge of variant interpretation and the power of laboratory techniques, including MitoString, to assess variant pathogenicity in difficult cases.

Background

Mitochondrial disease can present in infancy or adulthood and is characterized by a wide range of presentations affecting multiple organ systems. Common symptoms include lactic acidosis, myopathy, deafness, blindness, neurodegeneration, intestinal dysmotility and peripheral neuropathy (Vafai and Mootha, 2012). Although mitochondrial disease is one of the most abundant inborn errors of metabolism (1:5000 live births), there are no effective treatments and diagnosis is far from straightforward (Sanderson et al., 2006). Clinical, biochemical, and molecular criteria are applied to determine the likelihood of a patient having mitochondrial disease based on their presentation of symptoms and lab results (Morava et al., 2006). However, recent work has found that a number of non-mitochondrial diseases can present with a clinical picture indistinguishable from mitochondrial disease (Lieber et al., 2013).

Next-generation sequencing has proven invaluable for genetic diagnosis, but patients suspected of mitochondrial disease provide a striking example of the limitations of this approach. A recent cohort of 102 patients suspected of mitochondrial disorders established molecular diagnosis in just 22% of cases by MitoExome sequencing, which targets genes encoding mitochondrial-localized proteins (Lieber et al., 2013). In an additional 26% of cases, Lieber et al. is able to prioritize variants of unknown

significance that may cause disease. However, absent additional experimental proof, these variants cannot be deemed pathogenic. To prove the pathogenicity of genetic variants in cases without a history of Mendelian transmission, Chakravarti et al. have proposed a version of Koch's postulates for Complex Human Disease. The first criteria, that candidate gene variants are enriched in patients, is provable by exome-sequencing alone, but the remainder of the criteria require a model system with a disease-relevant human phenotype. Further, they require proof at two levels: first, the mutant allele must be proven responsible for a mutant phenotype in a model system, and second, the model phenotype must be accepted as equivalent to the human disease. For the first level, this model phenotype must be produced by disruption of the candidate disease gene, complemented by a wild-type human allele, and not complemented by the candidate variant (Chakravarti et al., 2013). Thus, molecular experiments are necessary in addition to sequencing to close the circle on diagnostic dilemmas.

In the following case, the clinical presentation was suggestive of mitochondrial disease and the patient underwent targeted MitoExome sequencing (Calvo et al., 2012; Lieber et al., 2013), which identified compound heterozygous variants in the gene *SUPV3L1,* which encodes a mitochondrial-localized protein. SUPV3L1 is a helicase implicated in mitochondrial RNA degradation, and previous research established that a dominant negative form of *SUPV3L1* increased build up of light strand noncoding transcripts in a cellular model (Szczesny et al., 2010). We investigated whether the patient-derived cell line had a similar phenotype to the validated dominant negative, in order to determine if we could use this model to test the candidate disease variants.

Case Presentation

A two-year-old female of European descent was referred to the mitochondrial disease clinic with hypotonia, failure to thrive, light twitching, and developmental delay (language). She had a history of constipation and an enlarged spleen upon exam. Upon further testing, she exhibited problems with her eyes (bilateral strabismus) and ears (right-sided mild hearing loss, left-sided sensorineural loss in the upper frequencies).

The patient had persistent mild elevation of liver function tests and intermittent mildly elevated lactate levels. In liver biopsies, the mitochondrial DNA (mtDNA) copy number was 39% of controls and electron microscopy identified abnormal, enlarged mitochondria with a depletion of cristae. These multi-systemic findings suggested the possibility of a mitochondrial disorder.

Results

The patient was enrolled in the mitochondrial disease registry at Massachusetts General Hospital. We performed MitoExome sequencing (Calvo et al., 2012) on the patient and her parents under the hypothesis that her disease was the result of a mutation in a gene encoding a mitochondrial-localized protein. During MitoExome sequencing, the mtDNA and 1600 nuclear genes encoding mitochondrial proteins and related disease genes were enriched by hybrid capture and sequenced.

In the patient's sequenced nuclear DNA, we identified 36 rare (<2% allele frequency in 1000Genomes), protein-modifying variants. Hypothesizing a recessive model of inheritance, we screened these 36 variants for either homozygous or putatively compound heterozygous variants. We identified no homozygous variants and two

variants in the gene *SUPV3L1,* which were confirmed to be compound heterozygous after phasing with the parental DNA. Both mutations, c.505A>G / p.I169V (<1:8000 allele frequency) and c.2075C>T / p.P692L (1.6% allele frequency), altered residues that were moderately to highly conserved and predicted to be "benign" by Polyphen2 (Table 3-1). *SUPV3L1* encodes a helicase implicated in degradation of mitochondrial RNA. Each mutation altered an amino acid residue in a protein required for appropriate mitochondrial function, so we proceeded to assess the effect of these mutations in a patient-derived fibroblast cell line and two healthy control cell lines.

First, we measured baseline SUPV3L1 protein expression in the patient and control cell lines. We observed equivalent SUPV3L1 expression in each cell line tested (Figure 3-1A). Thus, at least one patient variant generates a full-length, stable SUPV3L1 protein in fibroblasts. This does not confirm that the observed SUPV3L1 is functional, and therefore we proceeded to measure mitochondrial RNA levels.

Previous studies had established a signature of *SUPV3L1* dysfunction, which we leveraged to assess the function of our variants within the fibroblast cellular model. Szczesny et al. identified a *SUPV3L1* dominant negative (*SUPV3L1dn*) that stabilizes light strand noncoding transcripts. Using MitoString, which profiles mitochondrial

Table 3-1. Prioritized variants

Gene symbol	Protein	Chr:Pos	cDNA *	Ref. codon	Var. codon	Allele Freq **
SUPV3L1	p.I169V	10:70947445	c.505A>G	ATA	GTA	<1:8,000
SUPV3L1	p.P692L	10:70968505	c.2075C>T	CCA	CTA	1.6%***

* Transcript NM_003171, "+" strand
** Exome Variant Server (EVS), European Americans (EA)
*** Three p.P692L EA homozygotes in EVS (N=4300)

RNA expression, we found that *SUPV3L1* knockdown elevates light strand noncoding transcripts, while depleting some mitochondrial mRNAs (Chapter 2). Here, we use MitoString to assess whether the patient cell line has any features of *SUPV3L1* dysfunction, which would establish a cellular model for more thoroughly testing the pathogenesis of the candidate variants.

In patient and control cell lines, we stably overexpressed wildtype *SUPV3L1* and *SUPV3L1dn*, as well as *GFP* and *SLIRP* as controls and assessed mitochondrial RNA expression (Figure 3-1B). Our MitoString assay measured two noncoding light strand transcripts in the regions complementary to *COX1* (*antiCOX1*) and *ND5* (*antiND5*). Each of these transcripts was elevated during *SUPV3L1dn* expression, in each cell line tested (Figure 3-2). However, there is no difference between control and patient fibroblasts at baseline, suggesting SUPV3L1 function is not compromised in the patient fibroblasts. Further, overexpression of wild type *SUPV3L1* is indistinguishable from overexpression of the controls *mtGFP* and *SLIRP*.

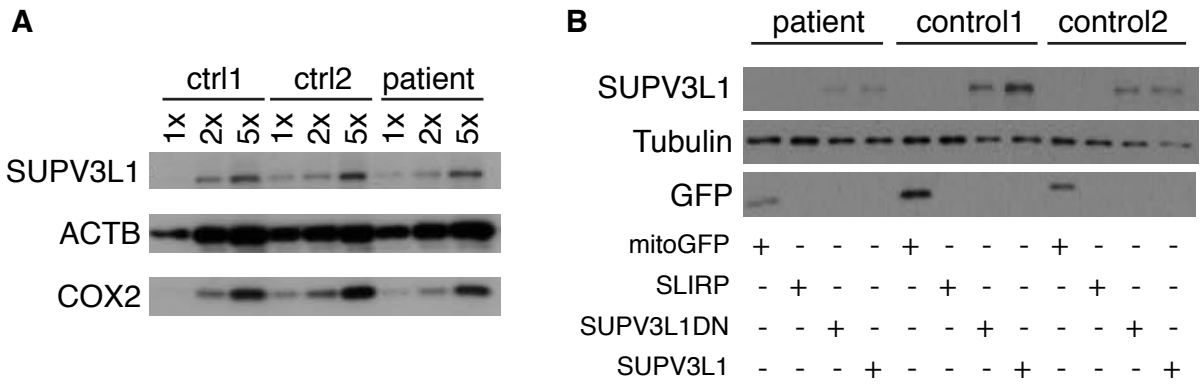


Figure 3-1. Protein expression in fibroblast control and patient cell lines.
(A) Baseline expression of SUPV3L1, ACTB, and mtCOX2 protein in control and patient-derived fibroblast cell lines by Western blot. (B) Overexpression of mitoGFP-FLAG, SUPV3L1-FLAG, and SUPV3L1dn-FLAG in control and patient cell lines.

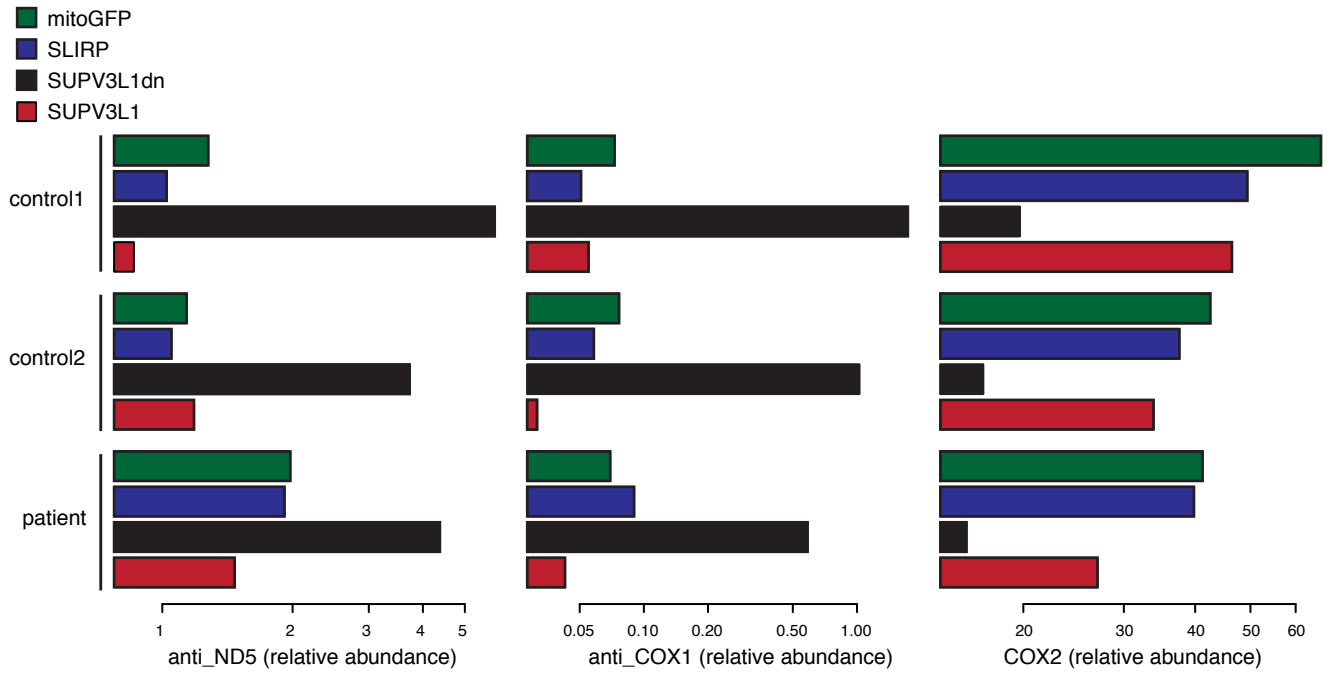


Figure 3-2. mtRNA expression in control and patient cell lines.
Expression of anti_ND5 (A), anti_COX1 (B), COX2 (C), and ND6 (D) transcripts in control and patient cell lines overexpressing FLAG-tagged mtGFP, SLIRP, SUPV3L1dn, and SUPV3L1.

The heavy strand transcript *COX2* was reliably depleted by *SUPV3L1dn* expression, but again, the patient cell line was not rescued in any way by wild type *SUPV3L1*, and in fact overexpression of *SUPV3L1* looked more like the dominant negative than GFP. Taken together, these results did not offer any evidence of a *SUPV3L1*-mediated defect in our patient cell line.

Conclusion

Despite leveraging advanced genetic and experimental techniques for the genetic analysis and variant interpretation, we were unable to assign a molecular diagnosis in this case. However, MitoString proved valuable in assessing a set of putative pathogenic variants in the gene *SUPV3L1*, as previous studies had established

the expected cellular phenotype for *SUPV3L1* disruption. We find that introduction of a SUPV3L1dn protein disrupts degradation of light strand transcripts, while patient fibroblasts were indistinguishable from controls in this respect. Further, introduction of wildtype SUPV3L1 did not have a differential effect on patient versus control cells, further confirming the lack of a measurable defect in our cellular model.

When a gene is well understood in a cellular context, putative pathogenic variants can be easily tested through cellular assays and re-introduction of the wild type and variant construct. In addition to enrichment of variants in affected individuals, this experimental proof is necessary to firmly establish pathogenicity within the model system (Chakravarti et al., 2013). A second level of proof, equivocating the model and the human disease is also necessary, and we do not explore that here. In the case explored here, we found no evidence that *SUPV3L1* function was compromised in our fibroblast model. In fact, subsequent whole-exome sequencing revealed mutations in N-glycanase 1 (*NGLY1*) that are considered causal (Enns et al., 2014). NGLY1 is a cytoplasmic enzyme that degrades misfolded glycoporoteins as a component of the endoplasmic reticulum–associated degradation machinery. This finding highlights that mutations in genes encoding non-mitochondrial proteins can mimic as mitochondrial disorders in a clinical context.

Experimental procedures

MitoExome sequencing and variant detection

MitoExome sequencing was as described previously (Lieber et al., 2013). Specifically, we targeted 3.1 Mb of DNA including the mtDNA and coding exons of 1,598

nuclear- encoded genes: 113 genes known to cause mitochondrial disease, 945 high-confidence mitochondrial genes from MitoCarta (Pagliarini et al., 2008), 327 genes predicted to associate with mitochondrial function, and 213 genes underlying monogenic disorders in the differential diagnosis, i.e., metabolic and neurologic disorders with phenotypic similarity to mitochondrial disease.

DNA was extracted from Epstein-Barr virus–immortalized lymphoblastoid cell lines. Targeted DNA regions were captured and sequenced (paired-end 76 base pair reads, Illumina GA-IIx or Hi-Seq) at the MGH Next-Generation Sequencing Core. Sequence alignment, variant detection, and annotation were performed as previously described (Lieber et al., 2013).

Variant prioritization

We identified variants with allele frequency <2% in 1000 Genomes Project (low coverage release 20110511) (Abecasis et al., 2010) and prioritized the following nuclear variants: i) known pathogenic variants (Human Gene Mutation Database version 2010.3) (Stenson et al., 2009), and ii) variants that were likely deleterious, including nonsense, splice site, coding indel, and missense variants at sites conserved in ≥ 10 of 44 aligned vertebrate species (Calvo et al., 2012). All prioritized variants were manually reviewed to remove likely sequencing artifacts and to phase potential compound heterozygous variants (Lieber et al., 2013).

Cell Culture

Patient and control fibroblasts were obtained from Gregory Enns and grown in Memα media (Life technologies 12571-063) with 17% fetal bovine serum (Sigma Cat.

No. F6178) with 1X penicillin, streptomycin, and glutamine (Life Technologies Cat. No. 10378-016) at 37° C with 5% CO_2.

FLAG-tagged protein expression

plexFLAG and mitoGFP were generated as described in Chapter 2. pDONR223-SUPV3L1 and pDONR223-SLIRP were obtained from a previously described cDNA library (Pagliarini et al., 2008). The SUPV3L1dn G207V (GGC->GTC) mutation was introduced via QuikChange mutagenesis (Agilent) using Primer 1: 5'-GATAATATTTCATTCAGTCCCCACAAACAGTGG and Primer 2: 5'-CCACTGTTTGTGGGGACTGAATGAAATATTATC. All cDNAs were cloned into plexFLAG via LR reaction (Invitrogen). The plexFLAG constructs were used to produce virus and infect HEK-293T cells as previously described (Baughman et al., 2009). Selection was introduced for two weeks with 5 ug/ml blasticidin.

MitoString Assay

Fibroblasts were lysed with RLT buffer (Invitrogen) and frozen at -80C. RLT lysates were hybridized overnight at 65C with a custom MitoString codeset (described in Chapter 2) from NanoString technologies, and processed with the nCounter robotic system per manufacturer's instructions (Geiss et al., 2008).

Immunoblotting

Fibroblasts were lysed with RIPA buffer and protein was loaded on a 12% Novex Tris-Glycine polyacrylamide gel (Life Technologies) and electrophoresed at constant voltage (125V). The gel was transferred to a polyvinylidene difluoride membrane using the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked for one hour at room temperature in tris-buffered saline with 0.1% Tween-20 and 5% BSA

(TBST-BSA). Membranes were incubated with primary antibody: anti-GFP (Abcam AB291) 1:50000, anti-SUPV3L1 (gift of W.Lee, UC-Irvine) 1:2000, anti-TUB (cell signaling #2128) 1:2000 in TBST-BSA overnight at 4°C. Mouse (1:5000) and Rabbit (1:10000) secondary antibody was used for one hour at room temperature. Membranes were developed using WesternLightning Plus-ECL (PerkinElmer).

Acknowledgements

We thank G. Enns and J. Platt for patient fibroblasts. We thank R. Sharma for comments on the manuscript.

Author contributions

A.W. performed all experiments and wrote the manuscript. D.L., S.H., and S.C. contributed to variant detection and prioritization. N.S., K.S., V.M., and S.C. selected the patient cohort and designed the MitoExome sequencing project.

References

Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A., and Consortium, G.P. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Baughman, J.M., Nilsson, R., Gohil, V., Arlow, D.H., Gauhar, Z., and Mootha, V. (2009). A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. PLoS Genetics *5*, e1000590.

Calvo, S.E., Compton, A.G., Hershman, S.G., Lim, S.C., Lieber, D.S., Tucker, E.J., Laskowski, A., Garone, C., Liu, S., Jaffe, D.B.*, et al.* (2012). Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Sci Transl Med *4*, 118ra110.

Chakravarti, A., Clark, A.G., and Mootha, V.K. (2013). Distilling pathophysiology from complex disease genetics. Cell *155*, 21-26.

Enns, G.M., Shashi, V., Bainbridge, M., Gambello, M.J., Zahir, F.R., Bast, T., Crimian, R., Schoch, K., Platt, J., Cox, R.*, et al.* (2014). Mutations in NGLY1 cause an inherited disorder of the endoplasmic reticulum-associated degradation pathway. Genet Med.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T.*, et al.* (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology *26*, 317-325.

Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R., Berry, G.T.*, et al.* (2013). Targeted exome sequencing of suspected mitochondrial disorders. Neurology *80*, 1762-1770.

Morava, E., van den Heuvel, L., Hol, F., de Vries, M.C., Hogeveen, M., Rodenburg, R.J., and Smeitink, J.A. (2006). Mitochondrial disease criteria: diagnostic applications in children. Neurology *67*, 1823-1826.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K.*, et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112-123.

Sanderson, S., Green, A., Preece, M.A., and Burton, H. (2006). The incidence of inherited metabolic disorders in the West Midlands, UK. Arch Dis Child *91*, 896-899.

Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S., and Cooper, D.N. (2009). The Human Gene Mutation Database: 2008 update. Genome Med *1*, 13.

Szczesny, R.J., Borowski, L.S., Brzezniak, L.K., Dmochowska, A., Gewartowski, K., Bartnik, E., and Stepien, P.P. (2010). Human mitochondrial RNA turnover caught in flagranti: involvement of hSuv3p helicase in RNA surveillance. Nucleic Acids Res *38*, 279-298.

Vafai, S.B., and Mootha, V.K. (2012). Mitochondrial disorders as windows into an ancient organelle. Nature *491*, 374-383.

Chapter 4
Characterization of the mitochondrial transcriptome in mouse liver

Contributors: Ashley R. Wolf, Xian Adiconis, Joshua Z. Levin, Vamsi K. Mootha

Summary

Approximately 1,100 nuclear proteins are imported into the mitochondria, through well-characterized import machinery, but the presence and role of mitochondrial-imported RNA is debated. Import of nuclear-encoded *5S* rRNA and the enzyme components *MRP* RNA and *RNase P* RNA have been described. Although the 22 tRNAs encoded by mtDNA have been considered sufficient for mitochondrial translation, import of nuclear tRNAs is also reported. Here we use sequentially purified mitochondria, an approach that has been successfully used to characterize the mitochondrial proteome, to characterize the mitochondrial transcriptome in mouse liver. We find strong enrichment of mitochondrial proteins and mtDNA-encoded transcripts in our mitochondrial samples. Nuclear-encoded *5S* rRNA is persistent in our mitochondrial samples, consistent with previous reports, but we do not observe mitochondrial import of nuclear tRNAs, in contrast with several recent mammalian studies. We discuss the possible reasons for this discrepancy and the challenges of using genome-wide sequencing to identify nuclear-imported transcripts.

Introduction

Mammalian mitochondria contain their own genome (mtDNA), which encodes 2 rRNAs, 22 tRNAs and 11 mRNAs. These transcripts are extremely abundant within mitochondria, and are essential for proper function of the respiratory chain. Roughly 1,100 nuclear encoded proteins are translocated into the mitochondria by well-established protein translocation machinery. However, there is much debate about which RNAs may be imported into the mitochondria, and how exactly they are imported. Various reports have identified nuclear-encoded miRNAs, sRNAs, tRNAs, and rRNA within mammalian mitochondria (Kren et al., 2009; Mercer et al., 2011; Rubio et al., 2008; Yoshionari et al., 1994).

Two mitochondrial-localized enzymes contain nuclear-encoded RNA components. The first, RNase MRP, is the endonuclease responsible for cleavage of the RNA primer required for mtDNA replication (Chang and Clayton, 1987). The *RNase MRP* RNA is dually partitioned to the nucleoli and cytoplasm, however it interacts with a distinct protein to perform its enzymatic activity in each location (Li et al., 1994; Lu et al., 2010). Mitochondrial RNAse P activity cleaves the 5' end of mitochondrial tRNAs. Both a protein-only complex and a protein-RNA complex have been assigned this activity (Holzmann et al., 2008; Puranam and Attardi, 2001). The RNA component of mitochondrial RNase P, like RNase MRP, is identical to the nuclear version, thus must be imported into mitochondria (Puranam and Attardi, 2001). The mitochondrial protein polynucleotide phosphorylase (PNPase) stimulates import of both the *MRP* and *RNase P* RNAs, but the exact mechanism remains unclear (Wang et al., 2010).

Import of nuclear-encoded *5S* rRNA into mammalian mitochondria has been biochemically characterized, confirming that RNA import into the mammalian mitochondria is possible (Magalhaes et al., 1998; Yoshionari et al., 1994). In fact, specific aspects of the RNA sequence have been identified as essential for import (Smirnov et al., 2008) and an import system involving the proteins rhodanese and MRPL18 has been described (Smirnov et al., 2010; Smirnov et al., 2011). Depletion of rhodanese, which decreases mitochondrial-localized *5S* rRNA, also decreases mitochondrial translation (Smirnov et al., 2010; Smirnov et al., 2011). PNPase also augments *5S* rRNA import (Wang et al., 2010).

The 22 tRNAs present in the mammalian mitochondria are sufficient to decode the mtDNA-encoded mRNAs, if wobbling of the third anticodon base is allowed, enabling the tRNAs to match more than one codon (Chomyn and Attardi, 2009). However, this idea has been challenged by the discovery of a nuclear-encoded glutamine tRNA in mammalian mitochondria (Rubio et al., 2008). Mitochondrial tRNA import has been described for Tetrahymena, trypanosomatids, yeast, plants, and marsupials (reviewed in Tarassov et al., 2007), but this was the first mammalian example, and the import pathway is unknown (Alfonzo and Soll, 2009).

A recent study that performed RNA sequencing on mitochondria and mitochondria matrix fractions from human cells identified multiple tRNAs localized to the mitochondria (Mercer et al., 2011). The identified tRNAs are encoded at multiple locations in the nuclear genome, including regions containing nuclear copies of mitochondrial DNA (NUMTs). NUMTs are regions in the nuclear genome that are near-

copies of the mitochondrial genome, inserted into the nuclear genome over evolution. Humans have 286 to 612 NUMTs present in the genome, with the largest spanning 90% of the mtDNA genome (Hazkani-Covo et al., 2010). NUMTs often contain regions that are just a few bases different from the mitochondrial genome, which makes it difficult to distinguish NUMTs from mtDNA using modern sequencing techniques, which can lead to identification of mtDNA “mutations”, which are in fact NUMTs differing from mtDNA by a single base for a given sequencing read (Wallace et al., 1997). In identifying imported tRNAs located within NUMTs, it is possible that this identification occurred using reads that were inadvertently mismapped to the nuclear genome, when they were in fact transcribed from mtDNA (Mercer et al., 2011).

Given the evidence that specific RNAs are imported into the mitochondria, we sought to validate previous findings and identify novel imported transcripts using an unbiased approach. Our approach complements the findings of Mercer et al., which performed RNA sequencing on isolated mitochondria and mitoplasts from human cells. We isolated mitochondria from mouse liver by sequential enrichment, which allows us to observe enrichment of transcripts across four samples of increasing mitochondrial purity. By using two different RNA sequencing approaches, we were able to scan the mitochondrial transcriptome for full-size and tRNA-sized imported RNAs. In contrast to Mercer et al., we find no evidence for enrichment of novel nuclear-encoded transcripts. In particular, despite good coverage of the nuclear tRNAs, we observe no evidence for import into mouse mitochondria.

Results

Sequential mitochondrial enrichment enriches mitochondrial proteins

To identify the transcriptome present within mitochondria, we isolated four samples at different levels of mitochondrial purity from mouse liver (Figure 4-1A). First, whole tissue lysate was sampled directly after homogenization of the liver. Next, crude mitochondria were isolated from the homogenate by differential centrifugation. To degrade nucleotides clinging to the outside of the mitochondria, we incubated the crude mitochondria in Micrococcal Nuclease (MNase). Finally, we inactivated the MNase and

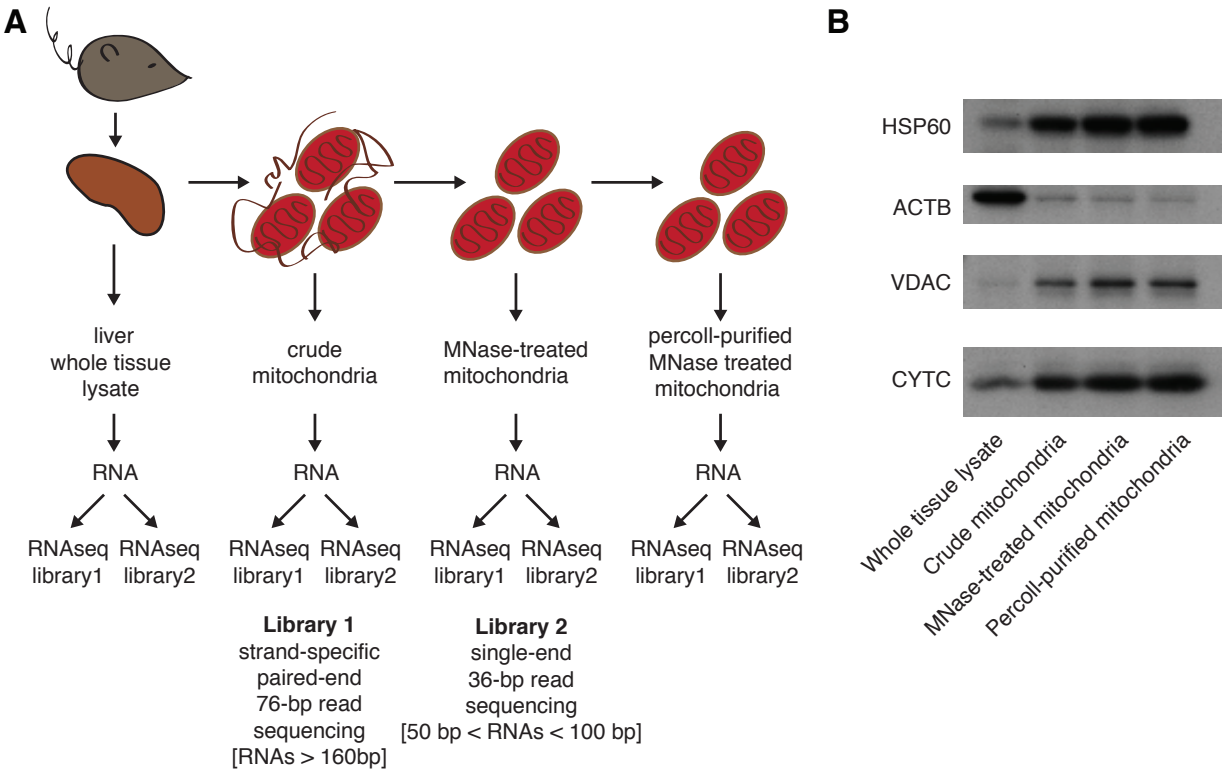
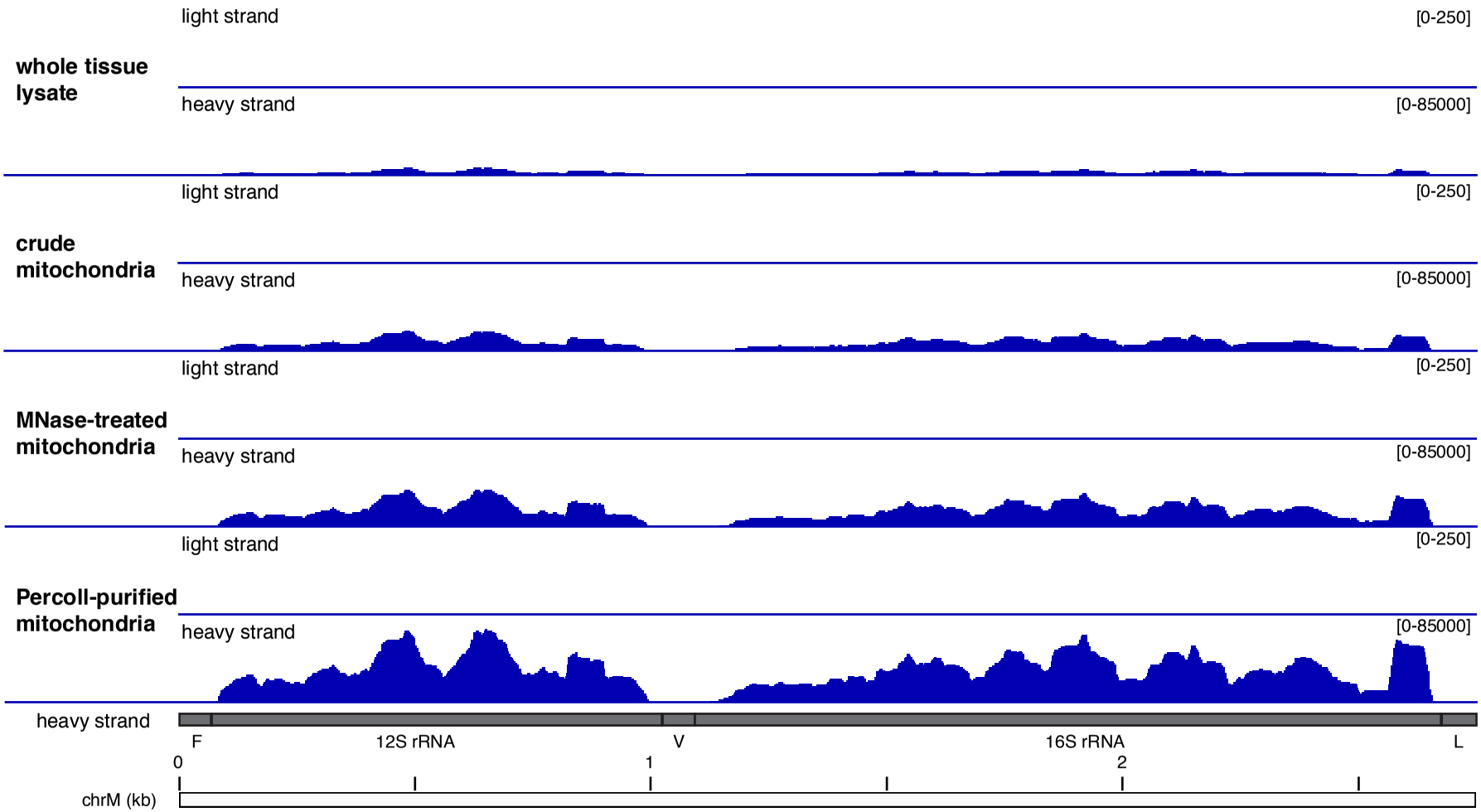


Figure 4-1. Sequential mitochondrial enrichment
(A) Mitochondria were isolated from mouse liver by differential centrifugation (crude mitochondria) then treated with MNase and further purified over a Percoll density gradient. RNA was isolated from each sample and two different RNA sequencing libraries were generated, as noted. (B) Western blot for mitochondrial protein abundance performed with protein from each sample in (A).

Figure 4-2. RNA expression of mitochondrial rRNA transcripts
For each sample, RNA-seq coverage from the large strand-specific library is visualized as generated by igvtools summarizing the Tophat alignment across 5 bp windows. The position of rRNAs and tRNAs (demarcated by one-letter code) encoded on the heavy strand of the mtDNA genome are depicted in grey at the bottom.


light strand
[0-250]
whole tissue lysate
heavy strand
[0-85000]
light strand
[0-250]
crude mitochondria
heavy strand
[0-85000]
light strand
[0-250]
MNase-treated mitochondria
heavy strand
[0-85000]
light strand
[0-250]
Percoll-purified mitochondria
heavy strand
[0-85000]
heavy strand
F
12S rRNA
V
16S rRNA
L
0
1
2
chrM (kb)

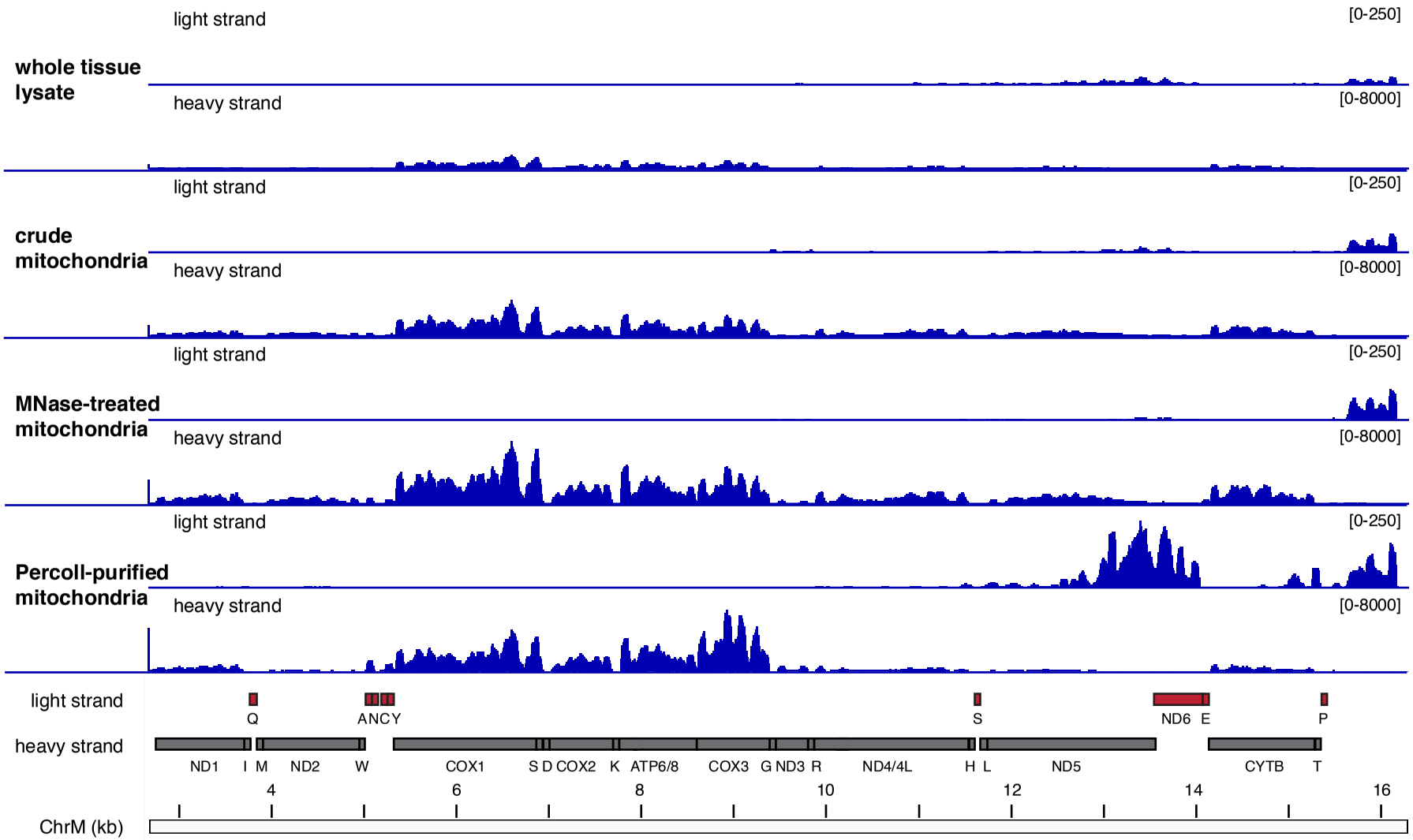
purified the treated mitochondria over a Percoll density gradient. Each step of this procedure enables more removal of non-mitochondrial contaminants.

We isolated protein from each sample and measured mitochondrial-localized proteins alongside actin (ACTB), an abundant cytoplasmic contaminant (Figure 4-1B). We find that the matrix-localized protein HSP60, inner membrane space protein CYTC and outer membrane protein VDAC are all enriched, while ACTB is depleted, as the samples are sequentially enriched for mitochondrial content. This confirms that we have achieved mitochondrial enrichment, and so we proceeded to RNA isolation.

Transcripts from the mitochondrial genome are enriched in large RNA library

We isolated RNA from the four samples and developed two distinct RNA-seq libraries. First, we created a strand-specific RNA-seq library for detecting large RNAs greater than 160 base pairs (bp) and performed paired-end sequencing with 76 bp reads. This library was not poly-A selected, in order to capture diverse transcripts types. In the large library, we clearly see enriched coverage of the mitochondrial rRNAs and mRNAs (Figure 4-2,3). The mitochondrial rRNAs are an order of magnitude more abundant than the most abundant mRNAs (Figure 4-2). Reads for the light strand in this region are not detected. As expected, the long-lived mRNAs *Cox1-3* and *Atp6/8* are more abundant than the other transcripts (Figure 4-3). Although gene-to-gene comparisons may be affected by sequencing efficiency, the difference is striking. As predicted by our size selection for transcripts greater than 160 bp, the tRNA transcripts are not well represented in this library (Figure 4-3).

Figure 4-3. RNA expression across the mtDNA genome
For each sample, RNA-seq coverage from the large strand-specific library is visualized as generated by igvtools summarizing the Tophat alignment across 5 bp windows. The position of mRNAs and tRNAs (demarcated by one-letter code) encoded on the heavy strand (grey) and light strand (red) of the mtDNA genome are depicted at the bottom.


light strand
[0-250]
whole tissue lysate
heavy strand
[0-8000]
light strand
[0-250]
crude mitochondria
heavy strand
[0-8000]
light strand
[0-250]
MNase-treated mitochondria
heavy strand
[0-8000]
light strand
[0-250]
Percoll-purified mitochondria
heavy strand
[0-8000]
light strand
Q
ANCY
S
ND6
E
P
heavy strand
ND1
I
M
ND2
W
COX1
S
D
COX2
K
ATP6/8
COX3
G
ND3
R
ND4/4L
H
L
ND5
CYTB
T
4
6
8
10
12
14
16
ChrM (kb)

As evidenced in Figure 4-3, the light strand is very lightly transcribed, and *Nd6* transcription is only apparent from the final enrichment step. Interestingly, the *Nd6* transcript continues into the sequence complementary to *Nd5*, as previously reported for human mtDNA (Rackham et al., 2011). We also see marked transcription upstream of the first light strand tRNA, between *Nd6* and the light strand promoter, which may include transcripts serving as the replication primer (Figure 4-3).

No tRNA import observed

To test reports of tRNA import into the mitochondria, we selected RNA transcripts from 50-100 bp in length from the whole tissue lysate and each of our mitochondrial enriched samples. We developed a protocol involving synthetic poly-A addition in order to sequence these short transcripts, which may be tRNAs of both mitochondrial and nuclear origin. This tRNA library was sequenced with single-end 36 bp reads.

Although *5S* rRNA at 121 bases is technically above our size threshold, we still have reasonable coverage of this RNA in our tRNA library, and it serves as a control for our study. The import of *5S* rRNA into the mitochondria has been biochemically studied, but a recent sequencing study noted only modest persistence of the transcript, as it is also abundant in the cytosol (Entelis et al., 2001; Mercer et al., 2011). By directly mapping to the *5S* rRNA reference sequence, we find that our crude mitochondrial sample contains 40% of the *5S* rRNA found in the whole tissue lysate sample, consistent with previous reports (Mercer et al., 2011) (Figure 4-4A).

We tested for tRNA import by analyzing our dataset in two distinct ways. We first aligned, allowing one mismatch, to a validated list of tRNAs (Chan and Lowe, 2009) and

the mtDNA-encoded tRNAs. As discussed, with this approach the crude mitochondrial sample contains 40% of the whole tissue lysate, and subsequent enrichments with MNase and Percoll did not decrease the *5S* rRNA abundance, suggesting residence within the mitochondrion (Figure 4-4A). In contrast, nuclear tRNAs *ArgTCT* and *LeuTAA* are markedly depleted with each step of the enrichment, suggesting they are abundant cytoplasmic contaminants (Figure 4-4A). Similarly, all nuclear tRNAs tested are depleted markedly in the mitochondrial-enriched samples (Figure 4-4B). The mitochondrial tRNA *Leu1* is 10-fold enriched, as it exists only within the mitochondrion (Figure 4-4A). When testing all tRNA transcripts, we found that the only tRNAs that were enriched were those that were mitochondrial-encoded (Figure 4-4B).

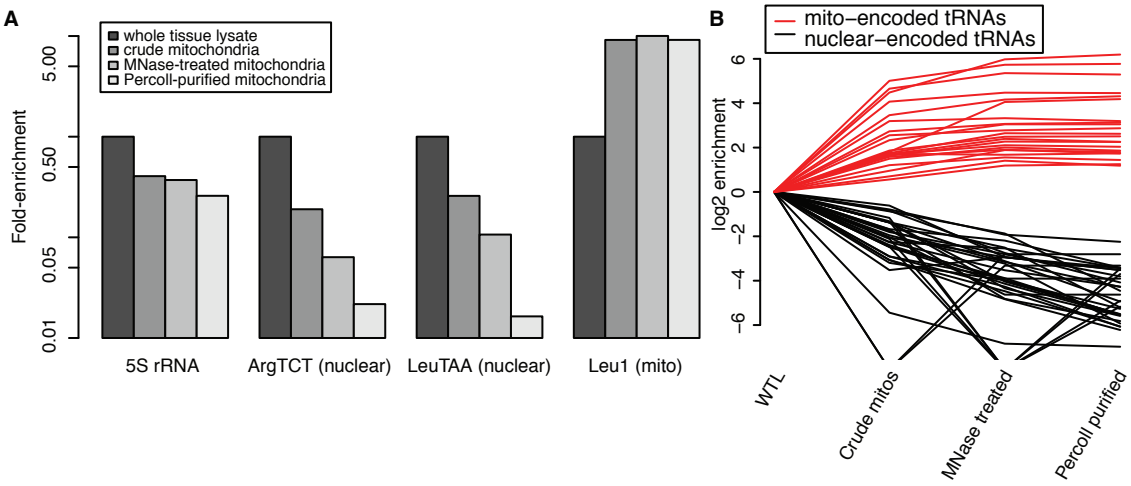


Figure 4-4. Expression of tRNAs and 5S rRNA
(A) Fold-enrichment of transcripts is shown across each sample, with whole tissue lysate normalized to 1. (B) Log2 (fold-enrichment) shown for each mtDNA-encoded tRNA (red) and nuclear-encoded tRNA (black) across all four samples, with whole tissue lysate (WTL) normalized to 0.

As previous reports had found mitochondrial enrichment of nuclear tRNAs in rat tissues and human cells, we verified our findings with a complementary approach. We aligned, allowing just one mismatch, and summed reads across intervals defined as

tRNAs by the RepeatMasker database. Only one annotated tRNA has a non-zero value for all four samples and an increased rpkm value for each mitochondrial sample that exceeds the whole tissue samples. That is *tRNA-Gln-CAA_* on chr12, which contains a 41-nucleotide span that is identical to mtDNA, since it exists within a NUMT. Thus reads from mtDNA can map to this nuclear transcript, likely falsely suggesting it is imported. This region was not annotated as a tRNA by the database used in the first analysis.

No novel imported transcripts observed

We also probed our large RNA-seq library to identify the *RNase MRP* and *RNase P* RNAs and any novel transcripts that might be imported. Unfortunately, the *RNase MRP* and *RNase P* RNAs were not well detected due to their low abundance, which has been previously remarked upon (Mercer et al., 2011). We used the Cufflinks software package to identify transcripts de novo based on the mouse genome sequence, and also used the Trinity software package to identify transcripts de novo without any genome assumed (Grabherr et al., 2011; Trapnell et al., 2010). In both cases, we found no mitochondrial-enriched transcripts. However, as noted, the *5S* rRNA transcript is not enriched with sequential mitochondrial purification steps, just persistent, so we also relaxed our threshold. We found that transcripts mapping to the extremely abundant *LSU-rRNA_Hsa* repeat, which encodes multiple rRNA transcripts, were also persistent. However, there was no specific sequence within this large repeat that was found to be systematically enriched, and these repeats also contain many repetitive elements that might cause mapping issues.

Discussion

The RNAs *5S*, *RNase MRP*, and *RNase P* RNAs have been biochemically proven to be imported into the mammalian mitochondria, but they were not striking in our RNA-seq dataset. The *RNase MRP* and *RNase P* RNAs were of too low abundance to be quantitated, and the *5S* rRNA was merely persistent, not necessarily enriched. Due to the high concentration of *5S* rRNA in the cytoplasm, this is perhaps expected, as mitochondrial abundance of *5S* may be lower than cytosolic abundance of *5S*. This quantitation is also consistent with what was found in human mitochondria with a similar approach (Mercer et al., 2011). However, the difficulty in picking up these known imported transcripts highlights the challenge of identifying new imported transcripts by unbiased genomic approaches. Noncoding RNAs of low abundance like the *RNase MRP* and *RNase P* RNAs may exist, but we may have not sequenced deep enough to detect them.

Although Mercer et al. find a number of tRNAs enriched in their mitoplast preparation when compared to the mitochondria, we do not find any evidence for mitochondrial import of nuclear tRNAs. Both biological and technical reasons may contribute to this discrepancy. Biologically speaking, Mercer et al. used mitochondria isolated from human cells, while we used mitochondria purified from mouse liver. It is possible that humans have developed import of tRNAs in a way distinct from mouse, or that this import is cell-type specific.

Technically, Mercer et al. used mitoplasts, a mitochondrial matrix preparation which removes the outer membrane, and may have acquired different information with

this method. We do however see similar *5S* rRNA persistence, and enrichment of the mtDNA encoded tRNAs with our method (Figure 4-4A), suggesting that we have reasonable power to detect transcripts. Mercer et al. find that *tRNA-LeuTAA* and *tRNA-GlnTTG* are enriched. We do measure each of these, although our *tRNA-GlnTTG* is of low abundance, which may have affected our detection. When looking at the human tRNA list they used to compile measurements (Chan and Lowe, 2009), we noticed that two of the seven sequences used to compile counts for *tRNA-LeuTAA* fell in human NUMTs, as well as one of the eleven sequences encoding *tRNA-GlnTTG*. None of the mouse sequences attributed to *tRNA-LeuTAA* and *tRNA-GlnTTG* have significant homology to mtDNA. This may mean that either regions that encode human NUMTs are more likely to be imported into the mitochondria, or that sequencing and mapping errors cause alignment to the nuclear genome when they are actually transcribed from mtDNA.

Biochemical characterization like that undertaken to validate the import of *5S* rRNA is necessary to determine the validity of tRNA import, as genomic reports have provided conflicting results. The reason for these discrepancies require further investigation. However, our dataset and approaches like it may also be valuable for categorizing reads as NUMT vs. mtDNA, as we expect NUMTs to be transcribed and depleted as mitochondrial purity increases, while mtDNA originating transcripts will be enriched. In our study, the four sequential stages of enrichment provide incredible value, clearly demonstrating cytoplasmic contaminants, and removing variation associated with looking at just two samples. It also further emphasizes how carefully one must consider the complications of aligning sequencing data given the presence of NUMTs.

Experimental Procedures

Mitochondrial Isolation

One three-month-old C57BL/6 male mouse was euthanized with CO_2 and the liver was quickly rinsed in ice-cold Buffer A (220 mM mannitol, 70 mM sucrose, 5mM HEPES, 1 mM EGTA, pH 7.4 + cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche)). The liver was minced and rinsed in Buffer A containing 1 mg/ml BSA, and homogenized with two strokes of a dounce homogenizer. Whole tissue lysate sample was set aside at this point. Crude mitochondria were isolated by differential centrifugation at 1000g for 10 min at 4^{0}C, followed by centrifugation of the supernatant at 8000g for 10 min at 4^{0}C. A portion of the pellet was saved as the crude mitochondrial sample. The remainder was resuspended in 300 μl Buffer A containing no EGTA and supplemented with 1mM $CaCl_2$, BSA 100 μg/ml, and 4000 gel units Micrococcal Nuclease (NEB M0247S). The remainder was separated on a percoll gradient spun at 20000 rpm for 45 min. in an SW60 rotor. Percoll-purified mitochondria were removed from the 26%-56% interface. All mitochondrial samples were spun for 1.5 min at 10000g and the pellet was lysed with 600 ul lysis/binding buffer (MirVana RNA isolation kit). RNA was isolated according to manufacturers instructions.

DNase treatment

RNA was treated with 2U TURBO DNase (Ambion) in a 50ul reaction at 37^{0}C for 30 min., then an additional 2U was added and incubated for another 30 min. at 37^{0}C. DNase was inactivated with .1 volumes DNase Inactivation Reagent (Ambion) and incubated five min. at room temperature. After centrifugation at 10000g for 1.5 min, the

RNA was transferred into a fresh tube and ethanol precipitated with .1 volumes 3M NaOAc, 1 μl glycogen and 2.5 volumes cold 100% EtOH. The sample was placed at -80^{0}C for at least 0.5 hr, then centrifuged at full speed (>13,000 x g) at 4^{0}C for 20 min. The pellet was twice washed with 500 μl 70% ethanol and centrifuged at full speed (>13,000 x g) at 4^{0}C for 10 min. The pellet was air dried and resuspended in water.

Large RNA-seq library preparation

RNAs larger than 160 bp were isolated on a 10% TBE urea gel. Strand-specific RNA-seq library preparation was as described (Yassour et al., 2010). Library was sequenced with 76 bp paired end reads on an Illumina GAII instrument.

tRNA-sized RNA-seq library preparation

RNAs from 50-100 bp in size were isolated on a 10% TBE urea gel. We dissolved 200 ng of each RNA sample in 25 μl H_2O and denatured at 75^{0}C for 4 min. then quickly chilled on ice. We added a 25 μl poly (A) tailing reaction mix that contained 25 U Poly(A) polymerase (New England Biolabs), 50 nmol ATP, 25 U of SUPERaseIn RNase Inhibitor (Ambion), 1x Poly(A) Polymerase Reaction Buffer, and incubated at 37^{0}C for 15 min., adding 312.5 nmol EDTA to stop the reaction. We cleaned up the reaction using Phenol:Chloroform:IsoAmyl Alcohol (PCIA) extraction and followed by ethanol precipitation. We denatured poly(A)-tailed RNA together with 50 pmol of 5'-GGCATTCCTGCTGAACCGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTVN-3' in 10.4 μl H_2O at 75^{0}C for 4 min., then quickly chilled on ice. We added 9.6 μl first-strand cDNA synthesis mix that contained 20 U Superscript III (Invitrogen), 200 nmol DTT, 40 nmol dNTP (New England Biolabs), 20 U SUPERaseIn RNase Inhibitor and 1x first-

strand cDNA synthesis buffer, incubated immediately at 55^{0}C for 15 min. We removed RNA by adding 10 μl 1M NaOH and heating at 70^{0}C for 30 min. We then added 10 μl 1M HCl to neutralize the reaction. We cleaned up the reaction using PCIA extraction and followed by ethanol precipitation with only 0.067x volume of 3M NaOAC. We removed the leftover reverse transcripion oligonucleotides by gel separation on a 10% TBE urea gel. We denatured the first-strand cDNA together with 1 nmol tagged random primer 5'- CTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNN-3' in 35 μl H_2O at 75^{0}C for 4 min. then quickly chilled on ice. We added 65 μl second-strand cDNA synthesis mix that contains 20 U Large (Klenow) fragment (New England Biolabs), 50 nmol dNTP and 1x NEB Buffer 2, incubated at room temperature for 10 min. followed by 37^{0}C for 30 min. We cleaned up this reaction by using 1.8x AMPure XP SPRI beads (Beckman Coulter Genomics). We used 25% double stranded cDNA in the final 50 μl PCR reaction that contained 1x Phusion® High-Fidelity PCR Master Mix with GC Buffer (New England Biolabs), 50 μmol Betaine (USB), and 12.5 pmol each of forward primer 5'- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCT-3' and reverse primer
5'- CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGC-3'. We performed the PCR reaction at 98^{0}C for 30 seconds, 5 cycles of 10 seconds at 98^{0}C, 30 seconds at 50^{0}C and 30 seconds at 72^{0}C, followed by 7 cycles of 10 seconds at 98^{0}C, 30 seconds at 65^{0}C and 30 seconds at 72^{0}C, then 5 min. at 72^{0}C. We gel size selected a PCR product ranging from 160 to 240 bp for sequencing. The library was sequenced on Illumina GAII with 36b PE reads.

Sequence alignment

Unless otherwise specified the sequencing libraries were aligned with Tophat, a gapped align tool for RNA-seq (Trapnell et al., 2009). For visualization of coverage, igvtools was used to compute coverage using a 5 bp window (Thorvaldsdóttir et al., 2013).

Quantitation of transcripts with Cufflinks

We used the Cufflinks command to quantitate transcripts without annotation within each library based on the mm9 genome, then used the cuffcompare and cuffdiff commands to compare abundance of transcripts across samples (Trapnell et al., 2010).

Quantitation of transcripts with Trinity

We used the Trinity package to quantitate transcripts without any genome reference in our 76 bp dataset as described (Grabherr et al., 2011; Haas et al., 2013).

Quantitation of targeted transcripts from the tRNA-sized library

We obtained tRNA annotations for the mouse genome (mm9) from the Genomic tRNA Database (http://lowelab.ucsc.edu/GtRNAdb/) (Chan and Lowe, 2009) and combined this list with sequences of mtDNA-encoded tRNAs extracted from the mouse mitochondrial chromosome Genbank file (NC_005089). We ran Bowtie against this file using the --best --a --strata -m 1 flags to require only one mismatch and only report the best match or matches in the same strata. This was also repeated for the *5S* rRNA sequence (NR_030686.1). For the second method, we used the Tophat alignments described and the Scripture score function (Guttman et al., 2010) to count the number of reads mapping to the repeatmasker annotated regions (http://repeatmasker.org).

Quantitation of RNase MRP *and* RNase P *RNA in large RNA-seq library*

We ran Bowtie against this a file containing the *RNase MRP* RNA (NR_001460.1) *RNase P* RNA (NR_002142.2) sequences using the --best --a --strata -m 1 flags to require one mismatch and only report the best match or matches in the same strata.

Author Contributions

A.W. performed mitochondrial enrichment and RNA-seq data analysis. J.L. and X.A. designed and performed RNA-seq library preparation. A.W. and V.M. designed the project.

Acknowledgements

We thank N.M. Cabili for advice on the RNA-seq software packages and F. Perocchi for advice on mitochondrial isolation.

References

Alfonzo, J., and Soll, D. (2009). Mitochondrial tRNA import--the challenge to understand has just begun. Biological Chemistry *390*, 717-722.

Chan, P.P., and Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res *37*, D93-97.

Chang, D.D., and Clayton, D.A. (1987). A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. Science *235*, 1178-1184.

Chomyn, A., and Attardi, G. (2009). Mitochondrial Gene Products. Current topics in bioenergetics *15*, 295-329.

Entelis, N.S., Kolesnikova, O.A., Dogan, S., Martin, R.P., and Tarassov, I.A. (2001). 5 S rRNA and tRNA import into human mitochondria. Comparison of in vitro requirements. The Journal of biological chemistry *276*, 45642-45653.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C.*, et al.* (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol *28*, 503-510.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M.*, et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc *8*, 1494-1512.

Hazkani-Covo, E., Zeller, R.M., and Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet *6*, e1000834.

Holzmann, J., Frank, P., Löffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008). RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell *135*, 462-474.

Kren, B.T., Wong, P.Y., Sarver, A., Zhang, X., Zeng, Y., and Steer, C.J. (2009). MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. RNA biology *6*, 65-72.

Li, K., Smagula, C.S., Parsons, W.J., Richardson, J.A., Gonzalez, M., Hagler, H.K., and Williams, R.S. (1994). Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis. The Journal of cell biology *124*, 871-882.

Lu, Q., Wierzbicki, S., Krasilnikov, A.S., and Schmitt, M.E. (2010). Comparison of mitochondrial and nucleolar RNase MRP reveals identical RNA components with distinct enzymatic activities and protein components. RNA *16*, 529-537.

Magalhaes, P.J., Andreu, A.L., and Schon, E.A. (1998). Evidence for the presence of 5S rRNA in mammalian mitochondria. Molecular biology of the cell *9*, 2375-2382.

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

Puranam, R.S., and Attardi, G. (2001). The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. Mol Cell Biol *21*, 548-561.

Rackham, O., Shearwood, A.M., Mercer, T.R., Davies, S.M., Mattick, J.S., and Filipovska, A. (2011). Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA *17*, 2085-2093.

Rubio, M.A., Rinehart, J.J., Krett, B., Duvezin-Caubet, S., Reichert, A.S., S√ ∂ ll, D., and Alfonzo, J. (2008). Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. Proc Natl Acad Sci USA *105*, 9186-9191.

Smirnov, A., Comte, C., Mager-Heckel, A.M., Addis, V., Krasheninnikov, I.A., Martin, R.P., Entelis, N., and Tarassov, I. (2010). Mitochondrial enzyme rhodanese is essential for 5 S ribosomal RNA import into human mitochondria. J Biol Chem *285*, 30792-30803.

Smirnov, A., Entelis, N., Martin, R.P., and Tarassov, I. (2011). Biological significance of 5S rRNA import into human mitochondria: role of ribosomal protein MRP-L18. Genes Dev *25*, 1289-1305.

Smirnov, A., Tarassov, I., Mager-Heckel, A.M., Letzelter, M., Martin, R.P., Krasheninnikov, I.A., and Entelis, N. (2008). Two distinct structural elements of 5S rRNA are needed for its import into human mitochondria. RNA *14*, 749-759.

Tarassov, I., Kamenski, P., Kolesnikova, O., Karicheva, O., Martin, R.P., Krasheninnikov, I.A., and Entelis, N. (2007). Import of nuclear DNA-encoded RNAs into mitochondria and mitochondrial translation. Cell cycle (Georgetown, Tex) *6*, 2473-2477.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Wallace, D.C., Stugard, C., Murdock, D., Schurr, T., and Brown, M.D. (1997). Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. Proc Natl Acad Sci U S A *94*, 14900-14905.

Wang, G., Chen, H.W., Oktay, Y., Zhang, J., Allen, E.L., Smith, G.M., Fan, K.C., Hong, J.S., French, S.W., McCaffery, J.M.*, et al.* (2010). PNPASE regulates RNA import into mitochondria. Cell *142*, 456-467.

Yassour, M., Pfiffner, J., Levin, J.Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D.A., Friedman, N., and Regev, A. (2010). Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol *11*, R87.

Yoshionari, S., Koike, T., Yokogawa, T., Nishikawa, K., Ueda, T., Miura, K., and Watanabe, K. (1994). Existence of nuclear-encoded 5S-rRNA in bovine mitochondria. FEBS Lett *338*, 137-142.

Chapter 5
Conclusion

In this dissertation, I explore the intersection of genomics, biochemistry, and medicine to expand our knowledge of mitochondrial RNA biology. In Chapter 2, I discuss the application of MitoString, a tool for mitochondrial RNA profiling, to broad-scale RNAi knockdown of mitochondrial proteins. This leads to the identification of the previously uncharacterized protein, FASTKD4, as a regulator of mitochondrial RNAs, which I validate in-depth. In Chapter 3, I apply MitoString to a fibroblast model of suspected mitochondrial disease, in which MitoExome sequencing had prioritized variants of unknown significance. This case highlights the value of an appropriate model to provide insight in unknown diseases. In Chapter 4, I use an unbiased approach to characterize RNAs residing within the mitochondria. Collectively, these studies expand our knowledge of mitochondrial RNA biology and pose questions for future research.

As illustrated by our case of suspected mitochondrial disease in Chapter 3, an understanding of the function of mitochondrial proteins is essential to understanding pathogenesis of mitochondrial disorders. In Chapter 2 we knock down over 100 mitochondrial-localized proteins with predicted RNA-binding domains and survey the effect on mitochondrial RNA abundance and processing. This atlas is a powerful resource for scientists investigating the function of these proteins in health and disease. Further, I have investigated the role of FASTKD4, a new regulator of mitochondrial RNAs. Reconciling that depletion of FASTKD4 affects only a subset of mRNAs while binding all of them, likely involves the identification of a new player in mitochondrial RNA

metabolism, which future studies should investigate. How the many different regulators of mitochondrial RNA, especially LRPPRC/SLIRP, interact with FASTKD4 is also an open question. Future studies may determine whether FASTKD4 associates with the LRPPRC/SLIRP complex, mitochondrial ribosomes, or acts independently on RNA transcripts. How FASTKD4 interacts with the degradation machinery is still unclear, as is the exact mechanism and specificity of the degradation machinery itself, including SUPV3L1.

Understanding the effect of specific variants on disease pathogenesis is currently one of the major challenges of human genetics. Chakravarti et al. suggest proof is necessary at two levels to indict a particular variant in a complex genetic case: the variant must be found necessary for the phenotype in a tested model system and the model system phenotype must be considered a valid stand-in for the disease itself. We sought the first level of proof with fibroblasts derived from the human patient. As we suspected variants in the gene *SUPV3L1*, and disruption of this gene is known to cause elevation in noncoding light strand transcripts, we tested this phenotype with our MitoString assay. We found no phenotype suggestive of a *SUPV3L1* defect in the patient fibroblasts, thus our model provided no evidence that the mutations were pathogenic. It is possible that our model was not relevant to the disease, but patient fibroblasts have been considered a useful model to validate pathogenic mutations in other mitochondrial patient cases (Calvo et al., 2010). Further, others subsequently found mutations in a non-mitochondrial gene (Enns et al., 2014). Our study provides an

example of determining the functional significance of variants of unknown function based on Koch's postulates for Complex Human Disease (Chakravarti et al., 2013).

Both our investigation of RNA processing proteins in Chapter 2 and our suspected mitochondrial disease patient in Chapter 3 highlight the power and limitations of using MitoCarta for prioritized screening. In Chapter 2, we identified a novel protein involved in mitochondrial RNA regulation, but we also missed a chance to understand some proteins that were not in MitoCarta at the start of the study, but were subsequently found to be mitochondrial-localized and involved in RNA processing. However, focusing on a larger pool of candidates would not have been experimentally feasible with MitoString, and would have resulted in less interesting data for a larger pool of candidates. In our case, with our smaller set of candidates, we obtained rich data on mitochondrial RNA transcript abundance for each knockdown. In the patient case, we obtained the MitoExome of the individual on the hypothesis that the patient suffered from mitochondrial disease. In the end, this was not the case, thus our MitoExome sequencing did not predict the pathogenic variant. For many other cases, this approach was successful (Lieber et al., 2013). However, this particular case highlights the complexity of diagnosing mitochondrial disorders, as non-mitochondrial disorders may present as mitochondrial. Future exome-sequencing studies on mitochondrial patients would be wise to cover the full genome, as plummeting sequencing costs make that more feasible.

Our survey of mitochondrial RNA transcripts in Chapter 4 highlights both the value and limitations of unbiased genome-wide approaches. We performed RNA

sequencing on mouse liver homogenate and sequentially enriched mitochondrial samples to identify whether any nuclear-encoded RNAs were imported into mitochondria. We found that even the one thoroughly characterized case, *5S* rRNA, was difficult to identify in our dataset without specifically looking for it. *5S* rRNA is extremely abundant within the cell, so mitochondrial fractions contained 40% of the *5S* rRNA found in the whole tissue lysate, yet in vitro studies have biochemically characterized its import into the mitochondrion. For *5S* rRNA we do note persistence across our three purified mitochondrial samples that is absent for measured nuclear tRNAs, which continue to be depleted with each enrichment step. However, absent prior information, I am not certain that our genome-wide study or a similar study in human cells would have identified *5S* rRNA as imported. Further, the lowly expressed *MRP* and *RNase P* RNAs were not well detected in our dataset, although deeper sequencing may solve this issue. Yet, for the import case, careful biochemical characterization may be more definitive than genome-wide studies.

Biochemical studies have also identified *tRNA-Gln* as imported into the mitochondria, but absent the level of proof presented in the *5S* rRNA case, there may be reason to be skeptical (Rubio et al., 2008). We find no evidence for import of nuclear tRNAs, and note that there are annotated tRNAs within NUMTs in the human nuclear genome that may confound studies on this point. However, as there are numerous biological and technical discrepancies between our study and the others, there are many possible reasons for our differing findings, as detailed in Chapter 4.

Overall, this dissertation uses a diverse set of approaches to characterize aspects of mitochondrial RNA biology. I have introduced a new regulator of mitochondrial RNAs, FASTKD4, and further characterized established regulators with MitoString. I also applied MitoString to fibroblasts derived from a suspected mitochondrial disease patient, providing an example of the power of molecular approaches to provide guidance in difficult genetic cases. Further, I have described an unbiased approach to characterize the mitochondrial-localized transcriptome of mouse liver, which may be in conflict with previous findings in humans and rat (Mercer et al., 2011; Rubio et al., 2008). Together, these findings contribute to our understanding of the essential mitochondrial RNAs that produce energy-generating OXPHOS subunits and raise questions to be answered by future studies.

References

Calvo, S.E., Tucker, E.J., Compton, A.G., Kirby, D.M., Crawford, G., Burtt, N.P., Rivas, M., Guiducci, C., Bruno, D.L., Goldberger, O.A.*, et al.* (2010). High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. Nat Genet *42*, 851-858.

Chakravarti, A., Clark, A.G., and Mootha, V.K. (2013). Distilling pathophysiology from complex disease genetics. Cell *155*, 21-26.

Enns, G.M., Shashi, V., Bainbridge, M., Gambello, M.J., Zahir, F.R., Bast, T., Crimian, R., Schoch, K., Platt, J., Cox, R.*, et al.* (2014). Mutations in NGLY1 cause an inherited disorder of the endoplasmic reticulum-associated degradation pathway. Genet Med.

Lieber, D.S., Calvo, S.E., Shanahan, K., Slate, N.G., Liu, S., Hershman, S.G., Gold, N.B., Chapman, B.A., Thorburn, D.R., Berry, G.T.*, et al.* (2013). Targeted exome sequencing of suspected mitochondrial disorders. Neurology *80*, 1762-1770.

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

Rubio, M.A., Rinehart, J.J., Krett, B., Duvezin-Caubet, S., Reichert, A.S., Soll, D., and Alfonzo, J. (2008). Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. Proc Natl Acad Sci USA *105*, 9186-9191.

Appendix A
Supplemental Information for Chapter 2

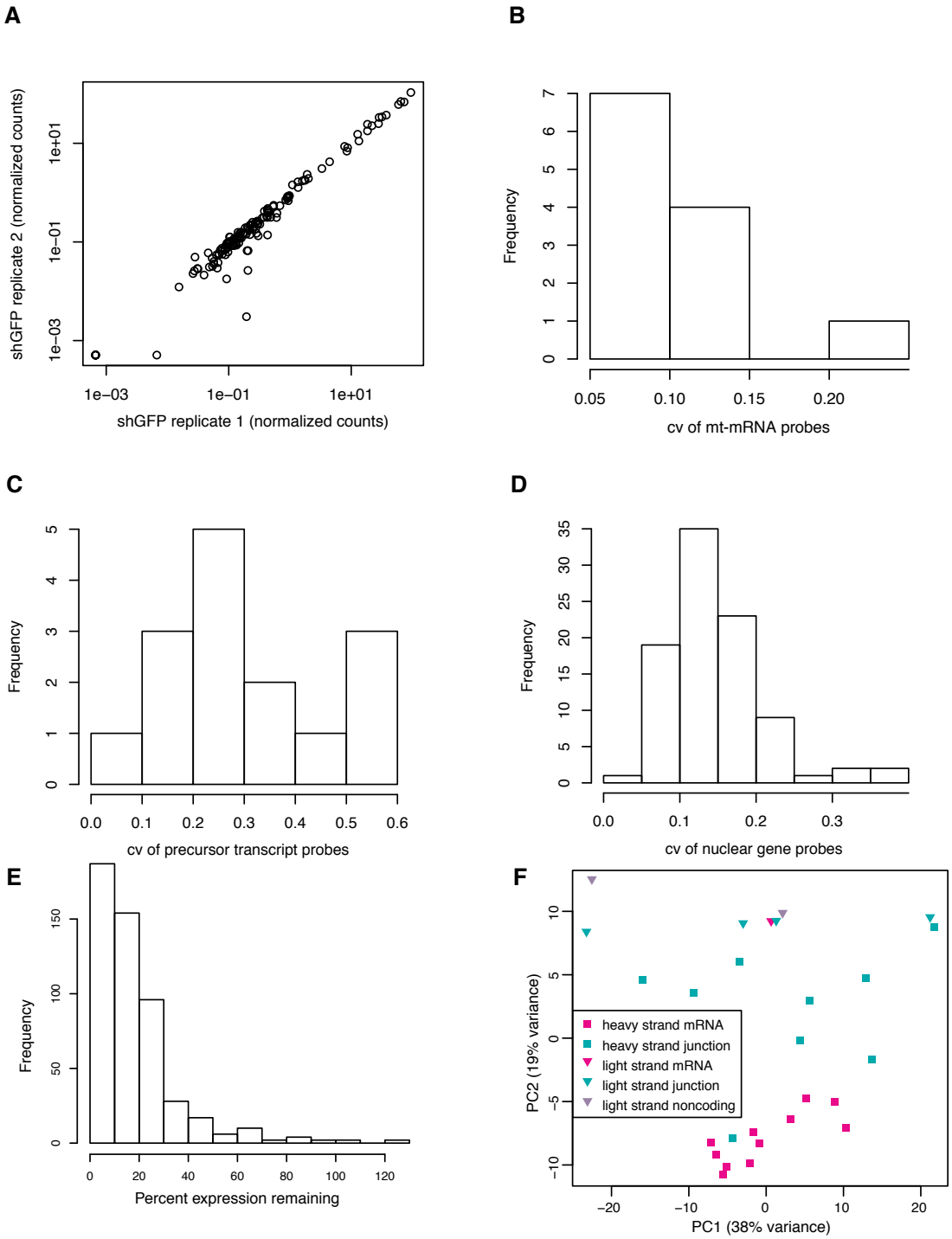


Figure 2-S1: Reproducibility of the MitoString assay, Related to Figure 2-1
(A) Normalized nCounter counts for all probes are plotted for one *shGFP* replicate vs. another. Histogram of CV values for individual probes targeting mtDNA-encoded mRNAs (B) and precursor (C). (D) Histogram of CV values for individual probes targeting nuclear genes. (E) Histogram of percent expression remaining of a given silenced gene within the dataset. (F) Principal component analysis of expression data matrix displayed in Figure 2-2. PC1 and PC2 account for 38% and 19% of the variance respectively and PC2 separates the probes based on the displayed categories (light or heavy strand, junction or mRNA).

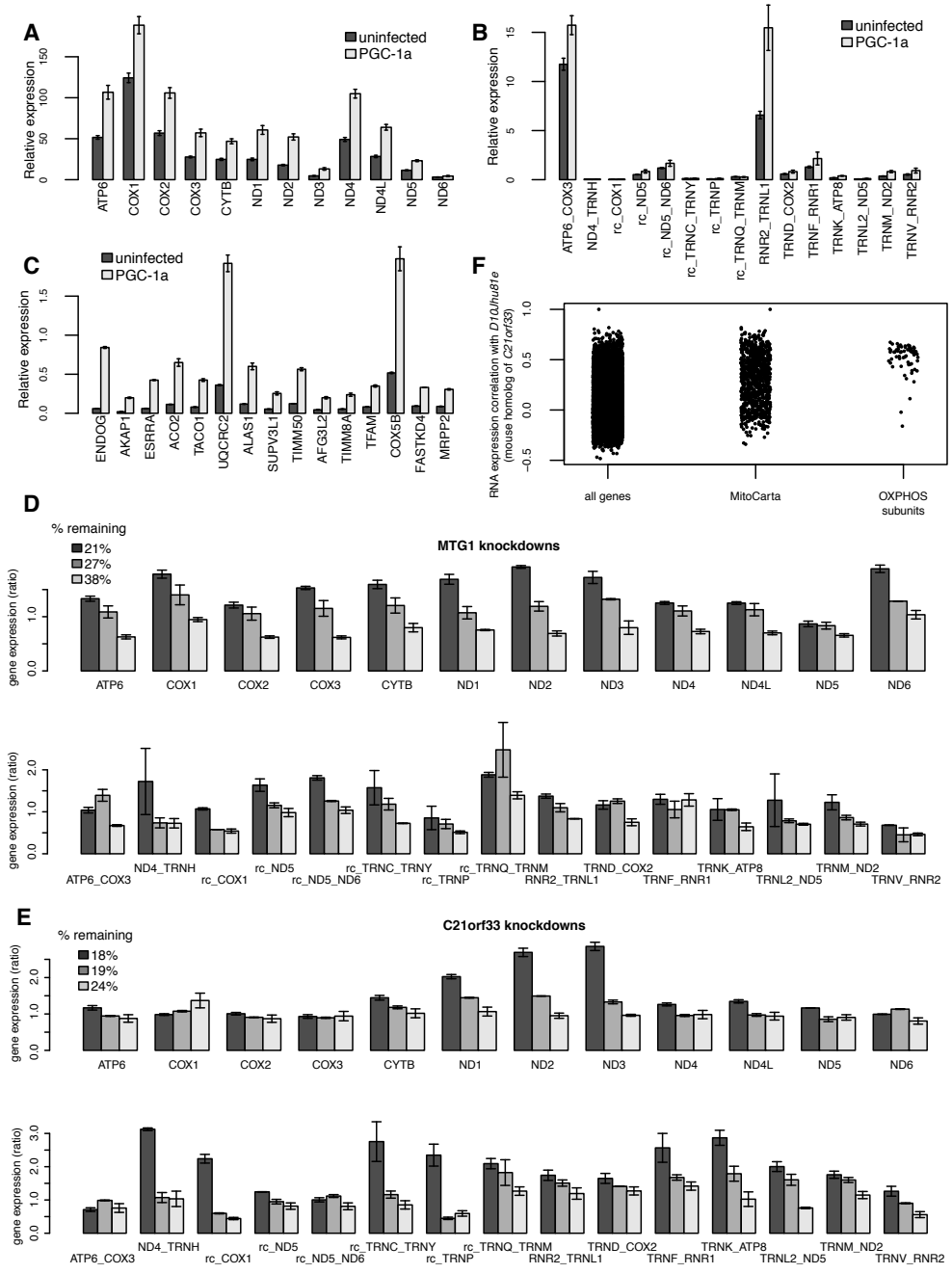


Figure 2-S2: Perturbations that increase mtRNA abundance, Related to Figure 2-2
MitoString measurements of RNA expression in uninfected cells and *PGC1alpha* overexpressing cells for mt-mRNAs (A), mt-RNA junction and non-coding probes (B) and for nuclear mitochondrial-localized genes (C). (D) MitoString quantitation for mtDNA probes under *MTG1* silencing by three distinct hairpins (% expression remaining of *MTG1* is denoted) E) MitoString quantitation for mtDNA probes under *C21orf33* silencing. (F) The Pearson correlation score of the mouse homolog of human *C21orf33* (*D10Jhu81e*) with the expression of all genes, genes in MitoCarta, and OXPHOS subunits across samples in the GNF Mouse GeneAtlas v3. Here, barplots represent the mean of duplicates, with the error bars representing the range.

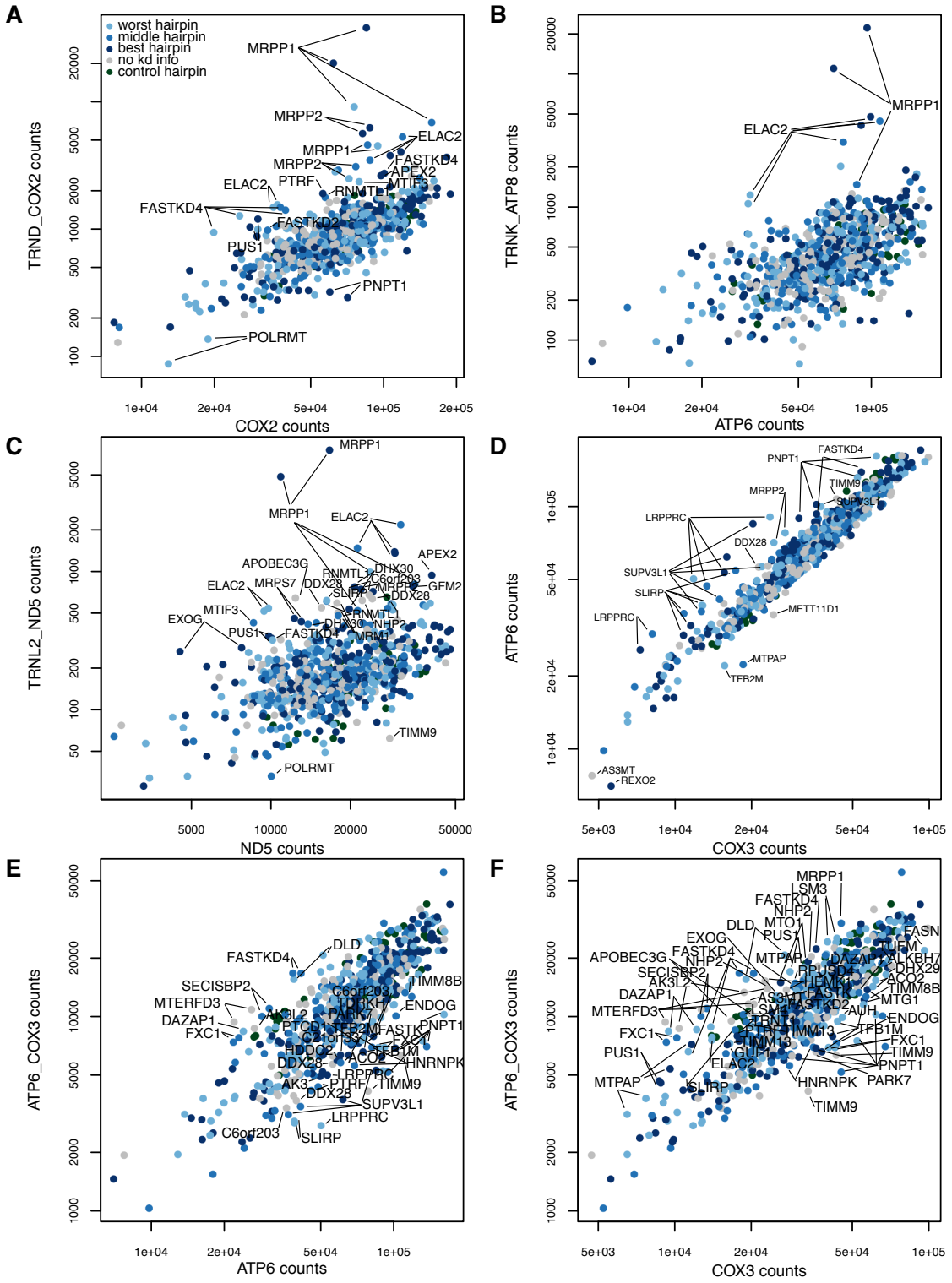


Figure 2-S3: Measuring RNase Z activity at measured tRNA_mRNA junctions and assessing the *ATP6/8_COX3* junction, Related to Figure 2-3

(A-F) The probe count for each knockdown hairpin was plotted for each set of neighboring probes. Genes that disproportionally affect one probe are found offset from the distribution and labeled (Supplemental experimental procedures).

Supplemental data tables
(Excel spreadsheets can be found in the accompanying file or online in supplemental information for Wolf and Mootha, 2014 Cell Reports)

Table 2-S1: shRNA hairpins used, Related to Figure 2-2 (see attached Excel sheet)

Table 2-S2: Knockdown efficiency, Related to Figure 2-2 (see attached Excel sheet)

Table 2-S3: Raw matrix of nCounter counts for shRNA knockdown samples, Related to Figure 2-2 (See attached Excel sheet)
Samples identified by TRCNXXXXXXXXXX, identified in Table 2-S1.

Table 2-S4: Processed nCounter data, Related to Figure 2-2 (see attached Excel sheet)
Raw data was normalized by the geometric mean of endogenous controls and divided by the mean of *shGFP* replicates (TRCN0000072179). Duplicates were averaged and the log2 was taken of all samples.

Table 2-S5: Mitochondrial disease genes targeted in the current study, Related to Figure 2-2

Gene Symbol	Disease (OMIM ID)	Proposed Pathogenesis	Previous result	Current study	References
AFG3L2	spinocerebellar ataxia type 28 (610246), spastic ataxia-5 (604581)	translation defect	unchanged (northern blot, mice)	slight decrease some transcripts	Almajan et al., 2012
ELAC2	Combined oxidative phosphorylation deficiency 17 (615440)	defect in RNA processing	precursor stabilized	precursor stabilized	Haack et al., 2013
FASTKD2	cytochrome *c* oxidase deficiency (220110)	apoptosis defect	–	decrease heavy strand mt-mRNAs	Ghezzi et al., 2008
GFM1	combined OXPHOS deficiency (609060)	translation defect	-	negligible effect	Smits et al., 2011
LRPPRC	Leigh syndrome (220111)	decreased mt-mRNA levels	decreased mt-mRNA levels (qPCR, northern blot)	decrease all heavy strand RNAs	Sasarman et al., 2010; Gohil et al., 2010
MTO1	hypertrophic cardiomyopathy and lactic acidosis (614702)	translation defect	decreased mt-mRNA levels (northern blot, yeast)	slight decrease heavy strand mt-mRNAs	Ghezzi et al., 2012; Wang et al., 2009
MTPAP	spastic ataxia (613672)	polyadenylation defect	conflicting results (qPCR)	decrease some mt-mRNAs	Crosby et al., 2010; Nagao et al., 2008; Piechota et al., 2006

Table 2-S5 (continued)

Gene Symbol	Disease (OMIM ID)	Proposed Pathogenesis	Previous result	Current study	References
PNPT1	combined OXPHOS deficiency (614932); hereditary hearing loss (614934)	translation defect	transcripts distinctly regulated (northern blot)	*CYTB* and *ND2* increased, some junctions decreased	Vedrenne et al., 2012 ; von Ameln et al., 2012; Borowski et al., 2013
PUS1	MLASA (600462)	translation defect	-	strong 1-hairpin depletion of *ATP6/8*,*COX3*, and *CYTB*; junction probes increased	Fernandez-Vizarra et al., 2007; Patton et al., 2005
TACO1	Leigh syndrome (256000)	translation defect	-	slight light strand depletion	Weraarpachai et al., 2009
TIMM8A	deafness–dystonia (Mohr-Tranebjaerg) syndrome (304700)	protein translocation	-	slight mt-mRNA depletion	Aguirre et al., 2006
TRMU	acute reversible infantile liver failure (613070)	translation defect	-	negligible effect	Zeharia et al., 2009
TUFM	combined OXPHOS deficiency (610678)	translation defect	-	*ND5* depletion	Valente et al., 2007

Table 2-S6: Antibodies, Related to Experimental Procedures

Source	Catalogue	Type	Antigen	Western Dilution
Abcam	ab99317	rabbit polyclonal	FASTKD4	1:10000
Cell Signaling	2128	rabbit polyclonal	Tubulin	1:2000
Abcam	ab291	mouse	GFP	1:50000
BD Bioscience	611980	mouse monoclonal	OXAL1	1:2500
Mitosciences	MS405	mouse monoclonal	COX2	1:10000
Mitosciences	MSA04	mouse monoclonal	Cyclophilin D	1:10000
Mitosciences	MSA06	mouse monoclonal	Cytochrome C	1:10000
Santa Cruz	SC-373752	mouse	FASTKD4	1:1000
Sigma	SAB2700419	rabbit	LRPPRC	1:1000
Abcam	ab114993	rabbit	TFAM	1:1000

Table 2-S7: Primers, Related to Experimental Procedures

Name	Description	Sequence
FASTKD4_725_F	Quickchange to generate FASTKD4 resistant to hairpin TRCN0000157725	CCGCTGGCGCATGCGCAAACTGAAGTATAAGCACCTGGC
FASTKD4_725_R	Quickchange to generate FASTKD4 resistant to hairpin TRCN0000157725	GCCAGGTGCTTATACTTCAGTTTGCGCATGCGCCAGCGG
U6	For FASTKD4ko generation	GCCTCTAGAGGTACCTGAGGGCCTATTTCCCATGATTCC
sgRNA	For FASTKD4ko generation	ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTAATGAACCGCCTGGAAGACGGTGTTTCGTCCTTTCCACAAG
CRISPR_L	For FASTKD4ko generation	CAAGATCGTGCTGTGATGGG
CRISPR_R	For FASTKD4ko generation	GGTTACAGACACCAAGCGTC

Supplemental experimental procedures

shRNA selection

The three best hairpins per gene were selected for screening based on qPCR knockdown validation data from the Broad RNAi platform, where available, or computational predictions using RuleSet8 (The RNAi Consortium).

Screen

15000 WI-38 cells/well were seeded in 96-well format and infected after 24-hours with 10 ul of virus prepared by the Broad RNAi consortium in media containing 8ug/ml polybrene. Plates were spun at 800g for 15 minutes. 24 hours, 72 hours, and 96 hours post-infection the media was refreshed with media containing 2ug/ml puromycin. 6 days post-infection the cells were lysed with RLT buffer (Invitrogen) and β-Mercaptoethanol (100:1) and frozen. RLT lysates were hybridized overnight at 65C with a custom MitoString codeset from NanoString technologies, and processed with the nCounter

robotic system per manufacturer's instructions (Geiss et al., 2008). nCounter counts are presented as either raw counts or normalized counts. Normalized counts are raw counts divided by the geometric mean of endogenous controls (ABCF1, ACTB, CLTC, HPRT1, TUBB) after background subtraction, presented as log2(normalized counts/shGFP), where shGFP is the mean of all infections with hairpin TRCN0000072179. Samples that were either too dense or too sparse were discarded from further analysis. Lowly expressing probes were also discarded. In raw count x-y scatterplots, points were labeled if they were categorized into the smaller of two linear regression groups in a mixture model explaining all data (Leisch, 2004).

FLAG-tagged protein expression

The V5 tag of the plex-TRC-983 Gateway destination vector (Broad RNAi Consortium) was replaced with a FLAG tag via PCR (Accuprime), restriction digestion (BstZ171, NheI, New England Biolabs), and ligation (Quick Ligase, New England Biolabs) to create plexFLAG. Mitochondrial-targeted eGFP was amplified via PCR to append BP sites and introduced into pDONR221 via Gateway BP reaction (Invitrogen). pDONR223-FASTKD4 was obtained from a previously described cDNA library (Pagliarini et al., 2008) and the RNAi site and RAP domain mutagenized via QuikChange mutagenesis (Agilent) using the primers described in Table S7. FASTKD4mut-FLAG contains the following amino acid changes: Y616A, L617A, K618A, K620A (NP_004740). All cDNAs were cloned into plexFLAG via LR reaction (Invitrogen).

Co-expression analysis

Using the GNF Mouse GeneAtlas v3 (Lattin et al., 2008), which quantifies the RNA expression of over 17,000 genes across 91 C57BL/6 mouse tissues or cell types using Affymetrix MOE430_2 microarrays, we calculated the Pearson correlation coefficient for each gene with *D10Jhu81e*, the mouse homolog of *c21orf33* across all samples. We plotted this Pearson coefficient for all genes, MitoCarta genes (Pagliarini et al., 2008), and OXPHOS genes (Baughman et al., 2009).

Supplemental References

Almajan, E.R., Richter, R., Paeger, L., Martinelli, P., Barth, E., Decker, T., Larsson, N.G., Kloppenburg, P., Langer, T., and Rugarli, E.I. (2012). AFG3L2 supports mitochondrial protein synthesis and Purkinje cell survival. J Clin Invest *122*, 4048-4058.

Crosby, A.H., Patel, H., Chioza, B.A., Proukakis, C., Gurtz, K., Patton, M.A., Sharifi, R., Harlalka, G., Simpson, M.A., Dick, K.*, et al.* (2010). Defective mitochondrial mRNA maturation is associated with spastic ataxia. Am J Hum Genet *87*, 655-660.

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). J Med Genet *44*, 173-180.

Ghezzi, D., Baruffini, E., Haack, T.B., Invernizzi, F., Melchionda, L., Dallabona, C., Strom, T.M., Parini, R., Burlina, A.B., Meitinger, T.*, et al.* (2012). Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. Am J Hum Genet *90*, 1079-1087.

Haack, T.B., Kopajtich, R., Freisinger, P., Wieland, T., Rorbach, J., Nicholls, T.J., Baruffini, E., Walther, A., Danhauser, K., Zimmermann, F.A.*, et al.* (2013). ELAC2 mutations cause a mitochondrial RNA processing defect associated with hypertrophic cardiomyopathy. Am J Hum Genet *93*, 211-223.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. J Stat Softw *11*.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K.*, et al.* (2008). A mitochondrial protein compendium elucidates complex I disease biology. Cell *134*, 112-123.

Smits, P., Antonicka, H., van Hasselt, P.M., Weraarpachai, W., Haller, W., Schreurs, M., Venselaar, H., Rodenburg, R.J., Smeitink, J.A., and van den Heuvel, L.P. (2011). Mutation in subdomain G' of mitochondrial elongation factor G1 is associated with combined OXPHOS deficiency in fibroblasts but not in muscle. Eur J Hum Genet *19*, 275-279.

Valente, L., Tiranti, V., Marsano, R.M., Malfatti, E., Fernandez-Vizarra, E., Donnini, C., Mereghetti, P., De Gioia, L., Burlina, A., Castellan, C*., et al.* (2007). Infantile encephalopathy and defective mitochondrial DNA translation in patients with mutations of mitochondrial elongation factors EFG1 and EFTu. Am J Hum Genet *80*, 44-58.

Wang, D.D., Shu, Z., Lieser, S.A., Chen, P.L., and Lee, W.H. (2009). Human mitochondrial SUV3 and polynucleotide phosphorylase form a 330-kDa heteropentamer to cooperatively degrade double-stranded RNA with a 3'-to-5' directionality. J Biol Chem *284*, 20812-20821.

Weraarpachai, W., Antonicka, H., Sasarman, F., Seeger, J., Schrank, B., Kolesar, J.E., Lochmüller, H., Chevrette, M., Kaufman, B.A., Horvath, R*., et al.* (2009). Mutation in TACO1, encoding a translational activator of COX I, results in cytochrome c oxidase deficiency and late-onset Leigh syndrome. Nat Genet *41*, 833-837.