



Computational and Experimental Approaches For Evaluating the Genetic Basis of Mitochondrial Disorders

Citation

Lieber, Daniel Solomon. 2013. Computational and Experimental Approaches For Evaluating the Genetic Basis of Mitochondrial Disorders. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11158264>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Computational and Experimental Approaches
For Evaluating the Genetic Basis of Mitochondrial Disorders

A dissertation presented

by

Daniel Solomon Lieber

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 2013

© 2013 - Daniel Solomon Lieber

All rights reserved.

Computational and Experimental Approaches For Evaluating the Genetic Basis of Mitochondrial Disorders

Abstract

Mitochondria are responsible for some of the cell's most fundamental biological pathways and metabolic processes, including aerobic ATP production by the mitochondrial respiratory chain. In humans, mitochondrial dysfunction can lead to severe disorders of energy metabolism, which are collectively referred to as mitochondrial disorders and affect approximately 1:5,000 individuals. These disorders are clinically heterogeneous and can affect multiple organ systems, often within a single individual. Symptoms can include myopathy, exercise intolerance, hearing loss, blindness, stroke, seizures, diabetes, and GI dysmotility. Mutations in over 150 genes in the mitochondrial DNA (mtDNA) and nuclear genome are known to cause mitochondrial diseases and an additional ~1,000 nuclear-encoded mitochondrial proteins have the potential to underlie mitochondrial disorders but have not yet been linked to human disease. As a result, determining a molecular diagnosis for patients with suspected mitochondrial disorders remains a challenge.

To improve our understanding of the genetic basis of mitochondrial disorders, we have developed a "Mito-Exome" sequencing approach that targets the mtDNA and the exons of nearly 1600 nuclear-encoded genes implicated in mitochondrial function, mitochondrial disease, or monogenic disorders with phenotypic overlap. We have applied this method to a diverse set of 102 patients referred to Massachusetts General Hospital with clinical and/or biochemical suspicion of mitochondrial disease. In addition to identifying pathogenic mutations in genes previously known to cause mitochondrial disease (e.g. *NDUFV1*, *POLG2*), our study has implicated several new or unexpected genes, including *WFS1*, *HSD17B4*, *DPYD*, and *ATP5A1*. However, the genetic basis of disease remains unclear for the majority of our patients. Our study demonstrates the strengths and limitations of next-generation sequencing approaches for the molecular diagnosis of suspected mitochondrial disorders.

To infer the function of uncharacterized mitochondrial proteins implicated in our genetic studies, we have also developed a computational genomic approach utilizing protein homology and bacterial gene neighborhoods. This approach has implicated the uncharacterized protein C6orf57 in a role in succinate dehydrogenase, also known as Complex II of the mitochondrial respiratory chain. The computational and genomic approaches described in this dissertation seek to improve our understanding of mitochondrial biology and its relationship to human health and disease.

Table of Contents

Chapter I: Introduction	1
Chapter II: Targeted exome sequencing of suspected mitochondrial disorders	15
Chapter III: Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease.....	36
Chapter IV: Targeted exome sequencing reveals HSD17B4-deficiency in a male with cerebellar ataxia, peripheral neuropathy, azoospermia, and hearing loss.	50
Chapter V: Operon analysis implicates C6orf57 in Complex II biology.....	61
Chapter VI: Implications and Future Directions	80
Appendix A	88

Acknowledgements

First, I would like to thank my PhD adviser, Vamsi Mootha, for his mentorship and support. His creativity and passion for science are truly inspiring.

Second, I'd like to thank my past teachers and mentors, without whom I never would have reached this point: Ellen Sidransky, Judith Blake, Tzachi Pilpel, Amy Caudy, Saeed Tavazoie, and David Botstein. I would especially like to thank my high school biology teacher Kimberly Agzigian, who introduced me to the field of genetics and encouraged me to pursue my interests. I would also like to thank all the professors and teaching assistants from Princeton's Integrated Science course. I thank the members of the Mootha lab for providing me with a supportive research environment. I would especially like to thank Sarah Calvo for her mentorship and guidance over the past few years.

Third, I'd like to thank my sources of funding and support. Much of this work was supported by a grant from the National Science Foundation Graduate Research Fellowship Program. I am also grateful to the Systems Biology department at Harvard University and to its administrators, especially Samantha Reed.

Fourth, I would like to thank the members of my Dissertation Advisory Committee (DAC) for their time, advice, and support: Michael Springer, George Church, and Gary Ruvkun, I'd also like to thank my defense committee and thesis readers: Michael Springer, Joel Hirschorn, Mark Korson, and Christopher A. Walsh. I would also like to thank Fritz Roth and Nir Hacohen for the opportunity to rotate in their laboratories.

Fifth, I would like to thank my Systems Biology PhD classmates: Ashley Wolf, Max Staller, Moran Cabili, Steve Hershman, Niall Mangan, Eric Kelsic, and John Bachman. I'd also like to thank all my classmates from Princeton's Integrated Science, who motivate me with their enthusiasm for scientific discovery.

Finally, I'd like to thank my family: my mom (Annette), dad (Marc) and sisters (Rachel, Talia, and Maya). I'd also like to thank my grandparents, Gerald and Ursula Bamberger, who have made so much in my life possible. Lastly, I'd like to thank my fiancée Rachel Gould, whose love and support have enriched my life and helped me towards the completion of my PhD.

CHAPTER I:

Introduction

The Human Genome Project produced a draft of the entire human genome in 2001 and was officially completed in 2003. At a cost of roughly \$3 billion and a length of 13 years, this project altered the course of scientific discovery and laid the foundations for a new future in medicine.¹ Today, the advent of new sequencing technologies has enabled the sequencing of whole genomes for less than \$10,000 – and this cost continues to drop. These new sequencing capabilities have transformed scientific research and have begun to make genomic medicine a reality.

For my dissertation, I have focused on a subset of the human genome pertaining to the mitochondrion, the organelle known to many as the “powerhouse of the cell”. As my adviser, Dr. Vamsi Mootha, determined years ago, mitochondria are, in many ways, the perfect organelle, especially as a focus for scientific research and medical discovery. The mitochondrion is a contained system within the cell, can be physically isolated, and much is known about its biology and the biology of its bacterial ancestors. The study of mammalian mitochondria has the potential to significantly improve our understanding of human health and disease, as mitochondrial dysfunction

has been associated with nearly every disease state, including both rare diseases and common ones, including diabetes, cancer, and Parkinson's disease.²

Through my dissertation research, I have sought to gain insight into the genetic basis of mitochondrial dysfunction and into the biological roles of uncharacterized mitochondrial proteins. This introduction reviews the basic biology of the mitochondrion, discusses the role of mitochondria in human health and disease, and outlines the subsequent chapters.

Mitochondria: the powerhouse of the cell

Mitochondria are energy-producing organelles found in most eukaryotic species. In mammalian cells, mitochondria are responsible for producing over 80% of cellular energy, in the form of adenosine triphosphate (ATP). It is widely accepted that mitochondria derived from a bacterial ancestor whose close symbiosis with a host cell eventually led to the complete integration of the two organisms.³ Over the course of evolution, the genes encoded by the mitochondrial genome were either lost or transferred to eukaryotic host genome, such that today only 13 of the 1,500 estimated human mitochondrial proteins are encoded by the mitochondrial DNA (mtDNA). The closest existing relative to the ancestor of mitochondrial appears to be *Rickettsia prowazekii*, an intracellular alpha-proteobacteria.⁴ Approximately 60% of mammalian mitochondrial proteins have close bacterial homologues,⁵ enabling researchers to infer the function of many mitochondrial proteins from their bacterial counterparts.

The mitochondrial proteome consists of approximately 1,000-1,500 proteins, which are encoded by both the mitochondrial and nuclear genomes. The mitochondrial genome, or mtDNA, is maternally inherited and encodes 13 proteins in addition to the 22 transfer RNAs (tRNAs) and 2 ribosomal RNAs (rRNAs) required for their translation.⁶ The remaining proteins are encoded by the nuclear genome and inherited, as other nuclear-encoded genes, in an autosomal, X-linked, or Y-

linked manner. Although mtDNA is essential to mitochondrial function, the proteins needed to replicate, transcribe, and translate the mtDNA are all encoded in the nuclear genome. Through an integrative Bayesian approach, members of the Mootha laboratory have created MitoCarta, a compendium of 1013 human mitochondrial proteins estimated to be over 85% complete and containing roughly 10% false positives.⁵ This effort has systematically catalogued the pathways and metabolic processes that occur within the mitochondria and has identified numerous proteins of unknown function that operate within the organelle.

The mitochondrion's primary role, arguably, is the creation of energy in the form of ATP through the mitochondrial respiratory chain.⁷ Through a process called oxidative phosphorylation (OXPHOS), the mitochondrial respiratory chain couples electron transport across four complexes (Complexes I-IV) to ATP creation by ATP synthase (Complex V). Complex I, also known as NADH dehydrogenase, mediates transfer of electrons from NADH to coenzyme Q (CoQ), also known as ubiquinone. Complex II, also known as succinate dehydrogenase, mediates transfer of electrons from succinate to CoQ without pumping protons. Complex III, also known as ubiquinol-cytochrome *c* reductase, transfers electrons from reduced CoQ to cytochrome *c*. Complex IV, also known as cytochrome *c* oxidase, transfers electrons from cytochrome *c* to oxygen, forming water. Complexes I, III, and IV couple electron transport to the pumping of protons out of the mitochondrial matrix, forming a proton gradient. This gradient is then dissipated by Complex V, also known as ATP synthase, a molecular motor that couples the flow of protons back into the mitochondrial matrix to the creation of ATP.

The mitochondrion is also host to numerous pathways and metabolic reactions, including the tricarboxylic acid (TCA) cycle, fatty acid catabolism, branched chain amino acid catabolism, and parts of heme biosynthesis. Although many of these pathways ultimately feed into the respiratory chain, other metabolic reactions may be independent of the primary role of mitochondria in ATP

synthesis. For an overview of the mitochondrion and mitochondrial research, see the book “Power, Sex, Suicide: Mitochondria and the Meaning of Life” by Nick Lane.⁸

Mitochondrial disorders

Given the role of the mitochondrion in cellular metabolism, it is not surprising that mitochondrial dysfunction can lead to devastating disorders of energy metabolism in humans. Mitochondrial respiratory chain diseases, collectively the most common inborn error of metabolism, are a large group of rare disorders that are notorious for their phenotypic and genetic heterogeneity.⁹⁻¹¹ The defining feature of these disorders is a biochemical defect in oxidative phosphorylation (OXPHOS). These disorders affect an estimated 1:5,000 births,¹² and can present in infancy or adulthood with varying levels of phenotypic severity. Multiple organ systems can be affected, with features that can include myopathy, seizures, lactic acidosis, sensorineural deafness, optic atrophy, diabetes mellitus, liver failure, seizures, and ataxia.^{9,10} Mitochondrial disorders can be caused by mutations in the nuclear or mitochondrial genomes and over 100 genetic loci have been identified to date.^{13,14}

Although these disorders may be inherited in a maternal, recessive, X-linked, or dominant fashion, many are singleton cases with no obvious family history of the disorder.¹⁵⁻¹⁷ Due to high pleiotropy and locus heterogeneity, traditional sequential genetic testing – including mitochondrial DNA (mtDNA) point mutation analysis and Sanger sequencing of nuclear-encoded disease genes – results in molecular diagnoses for only a fraction of patients.¹⁸ The limitations in traditional genetic testing have thus motivated the recent development of new next-generation sequencing (NGS)-based tools¹⁹⁻²³ and commercially available tests that simultaneously evaluate panels of genes previously shown to cause mitochondrial disorders.

In addition to causing inborn errors of metabolism, mitochondrial dysfunction has been directly linked to cancer. Germline mutations in subunits and assembly factors of succinate dehydrogenase, also known as Complex II of the mitochondrial respiratory chain, are associated with hereditary paraganglioma and pheochromocytoma, which are rare neuroendocrine tumors.²⁴ Additionally, germline mutations of fumarate hydratase are responsible for hereditary leiomyomatosis and renal-cell cancer (HLRCC).²⁵ Although tumor syndromes of a mitochondrial origin are discussed in Chapter V, this dissertation is primarily focused on mitochondrial disorders of energy metabolism that manifest as neurological or multi-system disorders.

Diagnosis of Mitochondrial Disease

Despite all we know, diagnosing a mitochondrial disease remains challenging due to the clinical and genetic heterogeneity of the disease.²⁶ The current diagnostic algorithms integrate biochemical, histological, clinical, and metabolic features to assess the likelihood of the disease. For example, elevated lactate and alanine are markers for mitochondrial dysfunction. A number of biochemical tests can be performed from skin, muscle, or liver biopsy, including testing of specific enzymatic reactions (e.g. pyruvate dehydrogenase, respiratory chain Complexes I-IV). Histological findings indicative of a mitochondrial disorder include ragged-red-fibers and abnormal mitochondrial morphology. Depletion of or deletions within mitochondrial DNA are also evidence of mitochondrial dysfunction, although mtDNA deletions can naturally accumulate with age.²⁷

There are two major diagnostic algorithms used by neurologists to diagnose patients suspected of a mitochondrial disorder, specifically respiratory chain (RC) disorders. The Bernier criteria,²⁸ which are applicable to adults and children, rely heavily on biochemical testing of the RC enzymes in muscle or liver biopsy. The algorithm lists a number clinical, histological, metabolic, enzymological, or genetic features that satisfy “major” or “minor” criteria. Depending on the

number of major and minor criteria, the patient is then classified as “unlikely”, “possible”, “probable”, or “definite”. The Morava criteria,¹¹ designed for the diagnosis of pediatric patients, rely more heavily on clinical observation (Figure 1.1). The algorithm consists of tallying up the number of muscular, central nervous system, and multi-systemic clinical features along with the number of metabolic/imaging or histologic abnormalities suggestive of mitochondrial disease, according to a pre-specified formula. The patient is given a score between 1-12 designating the likelihood of a mitochondrial disorder, with 1 translating to “unlikely”, 2 to 4 as “possible”, 5 to 7 as “probable”, and 8 to 12 as “definite”.

With the exception of Coenzyme Q10 (CoQ10), there are currently no FDA-approved treatments for mitochondrial diseases,²⁹ and therefore medical management is limited to treatment of symptoms (e.g. seizures, diabetes) as well as to the “mito-cocktail”, a combination of cofactors and vitamins, including Coenzyme Q, known to play a role in mitochondrial function. Still, the diagnosis of a mitochondrial disease can be important, in that it ends an often time-consuming and expensive diagnostic odyssey, informs medical management, and enables patients to seek the proper patient advocacy and support groups and to enroll in clinical trials.

Genetic diagnosis of mitochondrial diseases, like all monogenic disorders, has traditionally been on a gene-by-gene basis through the use of Sanger sequencing or variant genotyping (e.g. restriction fragment length polymorphism [RFLP] analysis). However, recent technologies (“next-generation” or “massively parallel” sequencing) have enabled the sequencing of many DNA fragments simultaneously at a dramatically lower cost per base compared to Sanger sequencing. These technologies have revolutionized the molecular diagnosis of genetic disorders.

I. Clinical signs and symptoms, 1 point/symptom (max 4 points)			II. Metabolic/imaging studies (max 4 points)	III. Morphology (max 4 points)
A. Muscular presentation (max. 2 points)	B. CNS presentation (max. 2 points)	C. Multisystem disease (max. 3 points)	Elevated lactate*	Ragged red/blue fibers**
Ophthalmoplegia*	Developmental delay	Hematology	Elevated L/P ratio	Cox-negative fibers**
Facies myopathica	Loss of skills	GI tract	Elevated alanine*	Reduced COX staining**
Exercise intolerance	Stroke-like episode	Endocrine/growth	Elevated CSF lactate	Reduced SDH staining
Muscle weakness	Migraine	Heart	Elevated CSF protein	SDH positive blood vessels*
Rhabdomyolysis	Seizure	Kidney	Elevated CSF alanine*	Abnormal mitochondria/EM*
Abnormal EMG	Myoclonus	Vision	Urinary TA excretion*	
	Cortical blindness	Hearing	Ethylmalonic aciduria	
	Pyramidal signs	Neuropathy	Stroke-like picture/MRI	
	Extrapyramidal signs	Recurrent/familial	Leigh syndrome/MRI*	
	Brainstem involvement		Elevated lactate/MRS	

* = 2 points GI, gastrointestinal; L/P, lactate/pyruvate; COX, cytochrome c oxidase; SDH, succinate dehydrogenase; EM, electron microscopy; EMG, electromyography; TA, tricarboxylic acid

** = 4 points

Score:
0 to 1: mitochondrial disorder unlikely
2 to 4: possible mitochondrial disorder
5 to 7: probable mitochondrial disorder
8 to 12: definite mitochondrial disorder

Figure 1.1. Morava criteria for the clinical diagnosis of mitochondrial disease in children

Adopted from Morava et al. (2006).¹¹

Next-generation sequencing

Although there are a number of next-generation sequencing (NGS) technologies that are currently available, this dissertation primarily focuses on the technology developed by Solexa, which was acquired by Illumina in 2007. For an overview and history of the various next-generation sequencing technologies, see “The \$1,000 Genome” by Kevin Davies.³⁰

The principle behind Illumina’s technology is massively parallel sequencing-by-synthesis.³¹ DNA fragments to be sequenced are immobilized on glass slides. Each DNA fragment is clonally amplified, such that there are millions of spatially separated colonies of PCR products. Sequencing occurs base-by-base through the addition of four fluorescently-labeled, reversibly terminated nucleotides. After each round of synthesis the clusters are excited by a laser, emitting a color (captured by a camera) that identifies the newly added base. The fluorescent label and blocking group are then removed, allowing for the addition of the next base and this process is repeated for N cycles, reading N base pairs from both ends of the PCR fragments, with N typically ranging from 76-150. In this way, millions of DNA fragments are sequenced in parallel.

Next-generation sequencing technologies are often used downstream of methods that facilitate targeted capture of particular regions of genomic DNA. Most commonly, this targeted capture is used to target the exome, the 1% of the human genome that encodes exons of protein-coding genes. The process of targeted exome capture followed by NGS is therefore referred to as “exome sequencing”. More generally, targeted DNA capture followed NGS is referred to as “targeted sequencing”. Although targeted sequencing is highly efficient, the approach impairs the detection of certain types of variants, including structural variants, which are better detected through whole-genome sequencing.

Since 2009, targeted sequencing has been used to identify the genetic basis of disease in patients suspected of rare, monogenic disorders.³² The first two examples of the successful

application of this technology were from the labs of Richard Lifton and Jay Shendure.^{33,34} In fact, the Shendure study implicated the gene *DHODH*, which encodes dihydroorotate dehydrogenase, a mitochondrial enzyme involved in de novo pyrimidine biosynthesis. Since these landmark papers, exome sequencing has become more commonplace; in 2012 there were 739 papers published with “exome” in the title or abstract, an average of two per day.

Next-generation sequencing of mitochondrial disorders

These sequencing technologies have greatly expanded our understanding of the genetic basis of mitochondrial diseases. Since 2011, there have been over 40 papers describing cases in which exome sequencing led to a genetic diagnosis of a patient with mitochondrial disease. Although some papers describe cases that expand the phenotypic spectrum of a known genetic disorder,³⁵ many of these papers identify causal mutations in genes not previously associated with human disease.³⁶⁻³⁸ In a number of cases (*AGK*, *FOXRED1*), genetic studies led to fundamentally new knowledge of protein function, in addition to having diagnostic implications.^{22,39} The combination of previously established mitochondrial disease genes with ones newly discovered through high throughput sequencing technologies has led to a total of over 150 genes (35 mtDNA, 128 nDNA) associated with mitochondrial disorders (Figure 1.2). This list of genes and their associated phenotypes is critical for genetic diagnosis of suspected mitochondrial disorders.

Although exome sequencing is a powerful method, it does not target the mtDNA and was initially more expensive than targeting a smaller, custom set of genomic regions. For these reasons, our laboratory developed a targeted “MitoExome” sequencing approach for the molecular diagnosis of mitochondrial disease.³⁹ This approach targeted the entire 16kb mtDNA, as well as the exons of ~1000 MitoCarta proteins. Later iterations included an additional ~400 proteins with predicted links to mitochondrial biology as well as ~200 genes known to cause metabolic or neurological disorders

with phenotypic overlap to mitochondrial disease. Sample preparation, sequencing and data processing were performed at Massachusetts General Hospital (MGH) or Broad Institute.

In work prior to this dissertation, our lab benchmarked the approach in a set of 42 severe, infantile cases with biochemically-proven mitochondrial disease in whom traditional approaches had failed to provide a molecular diagnosis.³⁹ MitoExome sequencing enabled new molecular diagnoses in 24% of cases and identified an additional 29% with recessive mutations in candidate genes not previously linked to disease. Pathogenicity of three of these candidate genes has since been independently validated,^{23, 37, 40} suggesting that the diagnostic success rate will rise once the pathogenicity of additional genes is established. Extrapolating to all such cases with or without prior molecular diagnoses, we estimated that MitoExome sequencing would enable diagnosis in 47% of cases and prioritize candidates in a further 20%.³⁹

It remained unclear, however, how such a diagnostic approach would fare when applied more broadly to patients ascertained on the basis of clinical presentation. To answer this question, members of our laboratory performed MitoExome sequencing in 102 patients referred to the metabolic disorders registry at Massachusetts General Hospital. Patients were suspected of mitochondrial disorders based on clinical, biochemical, and/or molecular findings, and had clinical presentations ranging from mild to severe, with varying age of onset. The results of this study are detailed in Chapter II.

For two of these patients, we were surprised to identify presumably causal genetic defects in non-mitochondrial proteins. The first patient, described in Chapter III, was an adult male who presented with diabetes mellitus, diffuse brain atrophy, autonomic neuropathy, optic nerve atrophy, a severe amnesic syndrome, and heteroplasmic mtDNA deletions. The second patient, described in Chapter IV, was an adult male with cerebellar ataxia, peripheral neuropathy, mild hearing loss, azoospermia, and mild reductions in mitochondrial respiratory chain activity.

mtDNA (35)		nuDNA (126)							
<i>ATP6</i>	tRNAGln	<i>AARS2</i>	<i>COQ2</i>	<i>FXN</i>	<i>MRPS22</i>	<i>NDUFS1</i>	<i>PUS1</i>	<i>ST3GAL5</i>	
<i>ATP8</i>	tRNAGlu	<i>ABCB7</i>	<i>COQ6</i>	<i>GARS</i>	<i>MTFMT</i>	<i>NDUFS2</i>	<i>RARS2</i>	<i>SUCLA2</i>	
<i>COX1</i>	tRNAGly	<i>ACAD9</i>	<i>COQ9</i>	<i>GCDH</i>	<i>MTO1</i>	<i>NDUFS3</i>	<i>RMND1</i>	<i>SUCLG1</i>	
<i>COX2</i>	tRNAHis	<i>ACO2</i>	<i>COX10</i>	<i>GFER</i>	<i>MTPAP</i>	<i>NDUFS4</i>	<i>RRM2B</i>	<i>SURF1</i>	
<i>CYTB</i>	tRNAIle	<i>ADCK3</i>	<i>COX14</i>	<i>GFM1</i>	<i>NDUFA1</i>	<i>NDUFS6</i>	<i>SARS2</i>	<i>TACO1</i>	
<i>ND1</i>	tRNALeu1	<i>AFG3L2</i>	<i>COX15</i>	<i>GLRX5</i>	<i>NDUFA10</i>	<i>NDUFS7</i>	<i>SCO1</i>	<i>TAZ</i>	
<i>ND2</i>	tRNALeu2	<i>AGK</i>	<i>COX4I2</i>	<i>HARS2</i>	<i>NDUFA11</i>	<i>NDUFS8</i>	<i>SCO2</i>	<i>TIMM8A</i>	
<i>ND3</i>	tRNALys	<i>AIFM1</i>	<i>COX6B1</i>	<i>HCCS</i>	<i>NDUFA12</i>	<i>NDUFV1</i>	<i>SDHA</i>	<i>TK2</i>	
<i>ND4</i>	tRNAMet	<i>ANT1</i>	<i>DARS2</i>	<i>HSPD1</i>	<i>NDUFA2</i>	<i>NDUFV2</i>	<i>SDHAF1</i>	<i>TMEM70</i>	
<i>ND4L</i>	tRNAPhe	<i>ATP5E</i>	<i>DGUOK</i>	<i>ISCU</i>	<i>NDUFA9</i>	<i>NFU1</i>	<i>SDHAF2</i>	<i>TRMU</i>	
<i>ND5</i>	tRNAPro	<i>ATPAF2</i>	<i>DNA2</i>	<i>LARS2</i>	<i>NDUFAF1</i>	<i>NUBPL</i>	<i>SDHB</i>	<i>TSFM</i>	
<i>ND6</i>	tRNASer1	<i>BCS1L</i>	<i>DNAJC19</i>	<i>LRPPRC</i>	<i>NDUFAF2</i>	<i>OPA1</i>	<i>SDHC</i>	<i>TTC19</i>	
12S rRNA	tRNASer2	<i>BOLA3</i>	<i>DNM1L</i>	<i>MARS2</i>	<i>NDUFAF3</i>	<i>PDK3</i>	<i>SDHD</i>	<i>TUFM</i>	
tRNAAla	tRNAThr	<i>C10orf2</i>	<i>EARS2</i>	<i>MFN2</i>	<i>NDUFAF4</i>	<i>PDSS1</i>	<i>SERAC1</i>	<i>TYMP</i>	
tRNAArg	tRNATrp	<i>C12orf65</i>	<i>ETHE1</i>	<i>MPV17</i>	<i>NDUFAF5</i>	<i>PDSS2</i>	<i>SLC19A3</i>	<i>UQCRB</i>	
tRNAAsn	tRNATyr	<i>C19orf12</i>	<i>FARS2</i>	<i>MRPL3</i>	<i>NDUFAF6</i>	<i>PNPT1</i>	<i>SLC25A1</i>	<i>UQCRC2</i>	
tRNAAsp	tRNAVal	<i>CLPP</i>	<i>FASTKD2</i>	<i>MRPL44</i>	<i>NDUFB3</i>	<i>POLG</i>	<i>SLC25A3</i>	<i>UQCRCQ</i>	
tRNACys		<i>COA5</i>	<i>FOXRED1</i>	<i>MRPS16</i>	<i>NDUFB9</i>	<i>POLG2</i>	<i>SPG7</i>	<i>YARS2</i>	

Figure 1.2. Mitochondrial and nuclear genes implicated in mitochondrial disorders.

Gene list compiled by S. Vafai and S. Calvo. Abbreviations: mtDNA, mitochondrial DNA; nuDNA, nuclear DNA

While Chapters I-IV describe efforts to pinpoint the genetic basis of disease in patients and to discover new disease genes, my research has also focused on developing methods to elucidate the function of uncharacterized mitochondrial proteins. Such methods have important implications for basic science as well as for clinical genetic studies, by helping to identify candidate genes related to a particular pathway or by informing the interpretation of variants in uncharacterized genes.

Chapter V describes an approach using bacterial operons and protein homology to infer the function of uncharacterized mitochondrial proteins. This approach has identified a putative functional relationship between the uncharacterized protein C6orf57 and Complex II of the mitochondrial respiratory chain.

Finally, Chapter VI describes the implications of my dissertation work and future directions. It is my hope that the research presented here facilitates the genetic diagnosis of rare disorders and accelerates future investigations of mitochondrial function in human health and disease.

References

1. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003 Apr 11;300(5617):286-90.
2. Wallace DC. Bioenergetic origins of complexity and disease. *Cold Spring Harbor symposia on quantitative biology*. 2011;76:1-16.
3. Margulis L. Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth: Yale University Press New Haven; 1970.
4. Andersson SG, Zomorodipour A, Andersson JO, et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*. 1998 Nov 12;396(6707):133-40.
5. Pagliarini DJ, Calvo SE, Chang B, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell*. 2008 Jul 11;134(1):112-23.
6. Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. *Nature*. 1981 Apr 9;290(5806):457-65.
7. Vafai SB, Mootha VK. Mitochondrial disorders as windows into an ancient organelle. *Nature*. 2012 Nov 15;491(7424):374-83.
8. Lane N, ebrary Inc. Power, sex, suicide mitochondria and the meaning of life. Oxford ; New York: Oxford University Press;; 2005. Available from: <http://site.ebrary.com/lib/princeton/Doc?id=10180687>.
9. DiMauro S, Hirano M, Schon EA. Mitochondrial Medicine. New York: Informa Healthcare; 2006.
10. Chinnery PF. Mitochondrial Disorders Overview 2010. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20301403.
11. Morava E, van den Heuvel L, Hol F, et al. Mitochondrial disease criteria: diagnostic applications in children. *Neurology*. 2006 Nov 28;67(10):1823-6.
12. Skladal D, Halliday J, Thorburn DR. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain : a journal of neurology*. 2003 Aug;126(Pt 8):1905-12.
13. Tucker EJ, Compton AG, Thorburn DR. Recent advances in the genetics of mitochondrial encephalopathies. *Curr Neurol Neurosci Rep*. 2010 Jul;10(4):277-85.
14. Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. *N Engl J Med*. 2012 Mar 22;366(12):1132-41.
15. Dimauro S, Davidzon G. Mitochondrial DNA and disease. *Ann Med*. 2005;37(3):222-32.
16. Chinnery PF, Turnbull DM. Epidemiology and treatment of mitochondrial disorders. *American journal of medical genetics*. 2001 Spring;106(1):94-101.
17. Kirby DM, Crawford M, Cleary MA, Dahl HH, Dennett X, Thorburn DR. Respiratory chain complex I deficiency: an underdiagnosed energy generation disorder. *Neurology*. 1999 Apr 12;52(6):1255-64.
18. Finsterer J, Jarius C, Eichberger H. Phenotype variability in 130 adult patients with respiratory chain disorders. *J Inherit Metab Dis*. 2001 Oct;24(5):560-76.
19. Vasta V, Ng S, Turner E, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med*. 2009;1(10):100.
20. Tucker EJ, Hershman SG, Kohrer C, et al. Mutations in MTFMT underlie a human disorder of formylation causing impaired mitochondrial translation. *Cell Metab*. 2011 Sep 7;14(3):428-34.
21. Vasta V, Merritt JL, 2nd, Saneto RP, Hahn SH. Next-generation sequencing for mitochondrial diseases reveals wide diagnostic spectrum. *Pediatr Int*. 2012 Apr 12.

22. Calvo SE, Tucker EJ, Compton AG, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nature genetics*. 2010 Oct;42(10):851-8.
23. Haack TB, Haberberger B, Frisch EM, et al. Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *Journal of medical genetics*. 2012 Apr;49(4):277-83.
24. Cascon A, Inglada-Perez L, Comino-Mendez I, et al. Genetics of pheochromocytoma and paraganglioma in Spanish pediatric patients. *Endocrine-related cancer*. 2013 Feb 12.
25. Tomlinson IP, Alam NA, Rowan AJ, et al. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat Genet*. 2002 Apr;30(4):406-10.
26. Haas RH, Parikh S, Falk MJ, et al. The in-depth evaluation of suspected mitochondrial disease. *Molecular genetics and metabolism*. 2008 May;94(1):16-37.
27. Baumer A, Zhang C, Linnane AW, Nagley P. Age-related human mtDNA deletions: a heterogeneous set of deletions arising at a single pair of directly repeated sequences. *American journal of human genetics*. 1994 Apr;54(4):618-30.
28. Bernier FP, Boneh A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology*. 2002 Nov 12;59(9):1406-11.
29. Hurko O. Drug Development for Rare Mitochondrial Disorders. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*. 2013 Feb 21.
30. Davies K. *The \$1,000 genome: the revolution in DNA sequencing and the new era of personalized medicine*: Free Press; 2010.
31. Hudson ME. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular ecology resources*. 2008 Jan;8(1):3-17.
32. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews Genetics*. 2011 Nov;12(11):745-55.
33. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009 Nov 10;106(45):19096-101.
34. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*. 2010 Jan;42(1):30-5.
35. Lieber DS, Vafai SB, Horton LC, et al. Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease. *BMC medical genetics*. 2012;13:3.
36. Meimaridou E, Kowalczyk J, Guasti L, et al. Mutations in NNT encoding nicotinamide nucleotide transhydrogenase cause familial glucocorticoid deficiency. *Nature genetics*. 2012 Jul;44(7):740-2.
37. Mayr JA, Haack TB, Graf E, et al. Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. *American journal of human genetics*. 2012 Feb 10;90(2):314-20.
38. Janer A, Antonicka H, Lalonde E, et al. An RMND1 Mutation causes encephalopathy associated with multiple oxidative phosphorylation complex deficiencies and a mitochondrial translation defect. *American journal of human genetics*. 2012 Oct 5;91(4):737-43.
39. Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science translational medicine*. 2012 Jan 25;4(118):118ra10.
40. Steenweg ME, Ghezzi D, Haack T, et al. Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. *Brain : a journal of neurology*. 2012 May;135(Pt 5):1387-94.

CHAPTER II:

Targeted exome sequencing of suspected mitochondrial disorders

Contributors: Daniel S. Lieber, Sarah E. Calvo, Kristy Shanahan, Nancy G. Slate, Shangtao Liu, Steven G. Hershman, Nina B. Gold, Brad A. Chapman, David R. Thorburn, Gerard T. Berry, Jeremy D. Schmahmann, Mark L. Borowsky, David M. Mueller, Katherine B. Sims, Vamsi K. Mootha

This chapter originally appeared as Lieber et al., “Targeted exome sequencing of suspected mitochondrial disorders”. *Neurology* (2013)

Abstract

Objective: To evaluate the utility of targeted exome sequencing for the molecular diagnosis of mitochondrial disorders, which exhibit marked phenotypic and genetic heterogeneity.

Methods: We considered a diverse set of 102 patients with suspected mitochondrial disorders based on clinical, biochemical, and/or molecular findings, and whose disease ranged from mild to severe, with varying age of onset. We sequenced the mitochondrial genome (mtDNA) and the exons of 1598 nuclear-encoded genes implicated in mitochondrial biology, mitochondrial disease, or monogenic disorders with phenotypic overlap. We prioritized variants likely to underlie disease and established molecular diagnoses in accordance with current clinical genetic guidelines.

Results: Targeted exome sequencing yielded molecular diagnoses in 22% of cases, including 17/18 (94%) with prior molecular diagnoses and 5/84 (6%) without. The five new diagnoses implicated two genes associated with canonical mitochondrial disorders (*NDUFV1*, *POLG2*), and three genes known to underlie other neurological disorders (*DPYD*, *KARS*, *WFS1*), underscoring the phenotypic and biochemical overlap with other inborn errors. We prioritized variants in an additional 26 patients, including recessive, X-linked, and mtDNA variants that were enriched two-fold over background and await further support of pathogenicity. In one case, we modeled patient mutations in yeast to provide evidence that recessive mutations in *ATP5A1* can underlie combined respiratory chain deficiency.

Conclusion: The results demonstrate that targeted exome sequencing is an effective alternative to the sequential testing of mtDNA and individual nuclear genes as part of the investigation of mitochondrial disease. Our study underscores the ongoing challenge of variant interpretation in the clinical setting.

Introduction

Mitochondrial disorders are a heterogeneous collection of rare disorders caused by mitochondrial dysfunction.¹ Multiple organ systems can be affected, with features that can include myopathy, encephalopathy, seizures, lactic acidosis, sensorineural deafness, optic atrophy, diabetes mellitus, liver failure, and ataxia.² Mitochondrial disorders can be caused by mutations in the nuclear or mitochondrial genomes and over 100 genetic loci have been identified to date.^{3,4} While these disorders may be inherited in a maternal, recessive, X-linked, or dominant fashion, many cases have no obvious family history of the disorder.⁵⁻⁷ Due to phenotypic and locus heterogeneity, traditional sequential genetic tests result in molecular diagnoses for only a fraction of patients.⁸ Limitations in traditional genetic testing have motivated our group and others to develop “next-generation sequencing” (NGS) approaches for molecular diagnosis.⁹⁻¹²

Previously, we developed a targeted “MitoExome” sequencing approach for the molecular diagnosis of mitochondrial disease and benchmarked its performance in a cohort of 42 severe, infantile cases with biochemically-proven mitochondrial respiratory chain disease.¹³ MitoExome sequencing enabled new molecular diagnoses in 24% of cases and, in an additional 29% of cases, identified recessive candidate disease genes, three of which have since been independently validated.¹⁴⁻¹⁶ Here, we evaluate a similar approach in a collection of 102 patients with suspicion of mitochondrial disease, spanning a broad range of age and phenotypic severity.

Subjects and Methods

Subjects

We considered patients that were referred to the MGH Mitochondrial Disorders Clinic for evaluation of possible mitochondrial disease and enrolled in the Partners DNA and Tissue Repository for Molecular Studies in Disorders of Energy Metabolism during 2004-2011. Of the 159

unrelated cases for whom high quality DNA was available, we selected the 102 subjects with the highest clinical suspicion of mitochondrial disease, based on chart review and review of biochemical testing results. The selected cases included as positive controls all 18 patients with prior molecular diagnoses acquired through traditional genetic testing (13 mtDNA and 5 nuclear mutations), eleven of which were made prior to referral to MGH. The remaining 84 patients had no prior molecular diagnosis, despite varying degrees of genetic testing in 74 cases. Of the 102 selected cases, 75% had histological studies, 63% had blood metabolite studies, 33% had neuroimaging studies, and 68% had enzyme analysis of the electron transport chain (ETC) from muscle biopsy. Muscle biopsies were not typically performed in cases with prior molecular diagnoses.

Standard Protocol Approvals, Registrations, and Patient Consents

Study protocols were approved by the Partners Human Research Committee. All samples were obtained with informed consent for sequence analysis and data deposition into public databases.

MitoExome Sequencing and Variant Detection

We expanded our previously reported MitoExome sequencing protocol to include additional genes of interest (Appendix A-1). Specifically, we targeted 3.1Mb of DNA including the mtDNA and coding exons of 1598 nuclear-encoded genes (Table A-1): 113 genes known to cause mitochondrial disease, 945 high-confidence mitochondrial genes from MitoCarta,¹⁷ 327 genes predicted to associate with mitochondrial function, and 213 genes underlying monogenic disorders in the differential diagnosis (DDx), i.e. metabolic and neurological disorders with phenotypic similarity to mitochondrial disease. DNA was extracted from whole blood or Epstein-Barr virus immortalized lymphoblastoid cell lines (LBLs). Targeted DNA regions were captured¹⁸ and

sequenced (paired-end 76bp reads, Illumina GA-IIx or Hi-Seq)^{18,19} at the MGH Next-Generation Sequencing Core. Sequence alignment, variant detection and annotation were performed as described in Appendix A-1. mtDNA variants with $\geq 1\%$ heteroplasmy based on read count were identified, and confirmatory qPCR assays were performed at m.3243, m.8344, and m.8993 (Baylor College of Medicine, test 3006).

Variant Prioritization

We prioritized the following mtDNA variants: (i) confirmed pathogenic variants in MITOMAP,²⁰ with allele frequency (AF) $< 0.5\%$ in mtDB²¹ (ii) mtDNA deletions/rearrangements detected through *de novo* mtDNA assembly,¹³ and (iii) loss-of-function (i.e. nonsense or frameshift) mutations with AF $< 0.3\%$ in mtDB²¹ and present in $< 10\%$ of 102 cases. Variants with heteroplasmy $< 20\%$ were confirmed in an independent DNA sample or were considered prioritized variants of unknown significance.

We prioritized the following nuclear variants: (i) known pathogenic variants (Human Gene Mutation Database²² version 2010.3) with AF $< 1\%$ in 1000 Genomes Project²³ (low coverage release 20110511), and (ii) variants with AF $< 0.5\%$ that were likely deleterious, including nonsense, splice site, coding indel, and missense variants at sites conserved in ≥ 10 of 44 aligned vertebrate species.¹³ For established autosomal recessive disease genes, we required presence of homozygous or two heterozygous variants. For X-linked recessive disease genes, we required two variants in females or a hemizygous variant in males. Nuclear variants in candidate genes not previously linked to disease were prioritized as above but considered to be prioritized variants of unknown significance. For dominant disease genes (autosomal or X-linked), we required AF $< 0.1\%$. All prioritized variants were manually reviewed to remove likely sequencing artifacts and to phase potential compound heterozygous variants (Appendix A-1). Prioritized variants are listed in Table A-2.

Estimating the Background Frequency of Variants

mtDNA was analyzed in 284 healthy individuals of European descent with whole-genome data from the 1000 Genomes Project. Nuclear DNA was analyzed in 371 healthy individuals of European descent with whole-exome data from the NIMH Control Sample repository. Background frequency comparisons were limited to sequence regions well covered in all individuals, as previously described.¹³ Enrichment was calculated as the percent of cases versus controls containing at least one qualifying variant, with significance assessed by a one-tailed binomial exact test.

Modeling the *ATP5A1* Mutation in Yeast

The *S. cerevisiae atp1* null mutant was created in an *ade2* background and transformed with the pRS305 plasmid expressing wild-type or mutant alleles of *ATP1* (Appendix A-1). Mitochondrial membrane potential, phosphate/oxygen (P/O) ratio, respiratory control rate (RCR), and ATPase activity were measured as described.²⁴ *Petite* frequency (percent white colonies, where lack of red pigment in *ade2* strains indicates mtDNA loss/deletion) was assessed in single haploid colonies grown, plated, and incubated in YPD, as previously described.²⁴ Loss of mtDNA in *petite* colonies was confirmed via mating with a ρ^0 strain.

Data Availability

Sequence and phenotype data are available in dbGaP, accession phs000339.

Results

MitoExome Sequencing in 102 patients

MitoExome sequencing was applied to 102 MGH patients suspected of a mitochondrial disorder. The cohort represented a broad spectrum of phenotypic severity ranging from lethal,

infantile to mild, adult-onset disease (Figure 2.1). Likelihood of mitochondrial disease was assessed as “definite”, “probable” or “possible” based on the Bernier and Morava criteria.^{25,26} 51% of patients had a biochemical ETC defect in muscle, defined as $\leq 30\%$ of normal activity of any complex after normalization to citrate synthase activity. 11% had positive family history, defined as presence of a first-degree relative with a highly similar phenotype. Age at diagnosis, defined as the date of initial biochemical or genetic testing for a mitochondrial disorder, ranged from 0 to 64, with median 27 years. 67% of patients were female and 87% were of European descent.

The approach yielded extremely high sequence coverage with mean 226X at targeted nuclear sites and 12,680X at mtDNA sites (Table A-3). 95% of targeted bases were covered with at least 10 reads, enabling confident detection of variants at the majority of sites. The deep coverage of mtDNA enabled sensitive detection of known heteroplasmic variants, including 5/6 variants present at $<10\%$ heteroplasmy (Appendix A-2).

We reasoned that high confidence “molecular diagnoses” could only be established within genes proven to cause disease consistent with the patient’s phenotype. We therefore first examined variants in established disease loci, adopting guidelines from the American College of Medical Genetics (ACMG) to make new molecular diagnoses.²⁷ Prioritized variants not consistent with previously reported phenotypes, or prioritized variants in genes not previously implicated with disease, were designated “prioritized variants of unknown significance” (pVUS).

Diagnoses in mtDNA

Within mtDNA, we prioritized previously reported pathogenic variants, large deletions or rearrangements, and rare loss-of-function variants (Methods). Other rare missense and tRNA variants were not prioritized, as their collective frequency could not be distinguished from that observed in healthy individuals (Figure A-1).

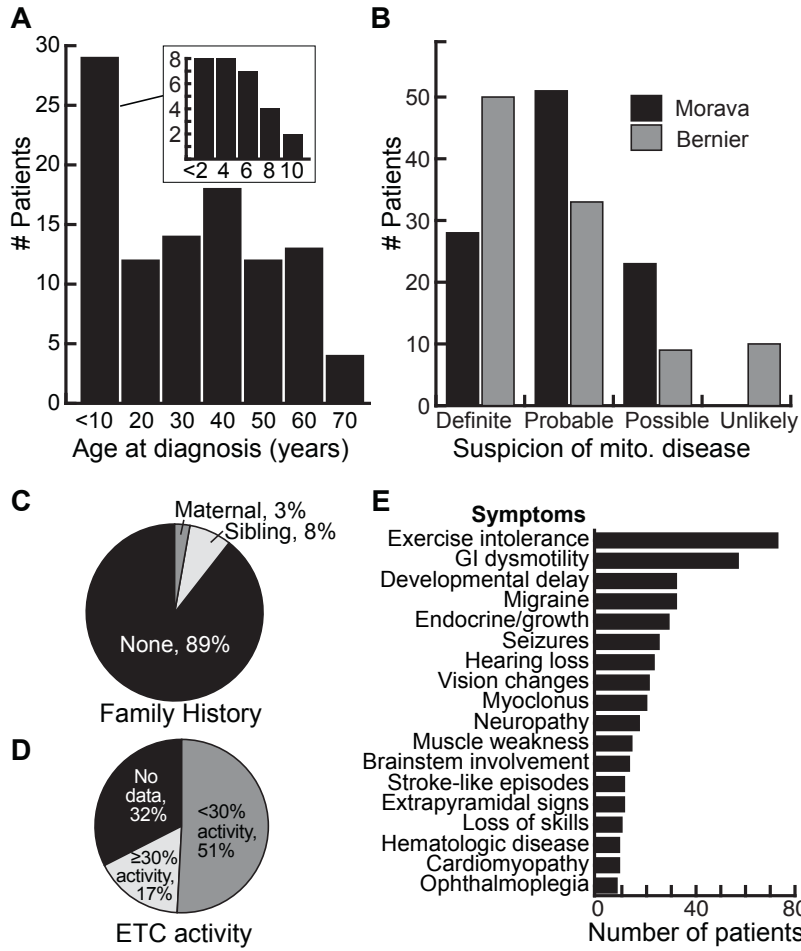


Figure 2.1. Characteristics of 102 MGH Patients Suspected of Mitochondrial Disease

Our diagnostic procedure correctly recovered molecular diagnoses for 12/13 patients with previously identified pathogenic mtDNA mutations (Table A-4). The exception was a patient with m.3243A>G present at 3% heteroplasmy based on qPCR in the sequenced LBL sample, underscoring the importance of tissue selection in studies of mitochondrial disease.²⁸

No confident mtDNA diagnoses were established in the remaining individuals. However, we identified six patients with pVUS in mtDNA (Appendix A-2, Table A-5), representing a 2.5-fold excess over the background rate in healthy controls ($p < 0.05$).

Diagnoses in Previously Established Nuclear Disease Genes

We next prioritized nuclear variants within 113 genes known to underlie mitochondrial disease and 213 genes underlying monogenic disorders in the differential diagnosis. Our diagnostic procedure recovered the prior molecular diagnoses for 5/5 cases with pathogenic nuclear mutations (Table A-6), and established new molecular diagnoses in five patients (Table 2.1).

Two of the five new molecular diagnoses implicated genes underlying primary mitochondrial disorders: a heterozygous *POLG2* mutation known to be pathogenic^{29,30} was detected in an individual with PEO, and compound heterozygous mutations in *NDUFV1*³¹ were detected in an individual with complex I deficiency in whom prior *NDUFV1* sequencing had detected only one variant.

The other three new molecular diagnoses, described below, implicated autosomal recessive variants in genes known to cause monogenic disorders in the differential diagnosis: dihydropyrimidine dehydrogenase (DPD) deficiency (*DPYD*),³² Wolfram syndrome (*WFS1*),³³ and a Charcot-Marie-Tooth-like axonal neuropathy (*KARS*).³⁴ Importantly, the phenotypes of these patients were not clearly discernible even in retrospect from the rest of the cohort, underscoring the phenotypic and biochemical overlap between canonical mitochondrial disease and other monogenic disorders.

Table 2.1. Individuals with New Molecular Diagnoses from MitoExome Sequencing.

Abbreviations: PEO, progressive external ophthalmoplegia; IUGR, Intrauterine growth restriction; COX, cytochrome oxidase; sib, affected sibling; ↓↓, <20% activity, ↓, 20-30% activity relative to citrate synthase; -, not determined; EM, electron microscopy; LM, light microscopy; cmpd. het., compound heterozygous; hom., homozygous; na, not applicable; * gene not previously linked to disease

ID	Sex	Age of Onset (Death)	Clinical Features	Family History	ETC Biochemistry					mtDNA/Histology Defects	Gene	Type	Mutation [Reference]	ACMG Category	Disease Assoc. with Gene (OMIM)
					I	I+III	II	III	IV						
1004	F	1 yr	Stroke, exercise intolerance, dyspnea, leukodystrophy, headache	1 sib (Leigh's disease) d. 13 mo	↓↓	↓↓	-		↓	EM: subsarcolemmal mito. aggregates, curvilinear cristae	<i>NDUFV1</i>	cmpd. het.	p.P113L p.S158PfsX15	3 2	Complex I deficiency (#252010)
1019	F	44 yr	Bi-lateral ptosis, PEO, ataxia, fatigue, dysphagia	3 sibs (PEO)	-	-	-	-	-	LM: ragged red fibers, COX neg. fibers	<i>POLG2</i>	het.	p.R369G [29]	1	PEO, autosomal dominant 4 (#610131)
1073	M	33 yr	Diabetes mellitus type 1, memory loss, cerebellar atrophy, autonomic dysfunction, mild hearing loss	None						Heteroplasmic mtDNA deletions	<i>WFS1</i>	hom.	p.R558C [38]	1	Wolfram syndrome (#222300)
1097	F	4 yr	Toe walking, tremors, myoclonic jerks, falls, extrapyramidal spasticity, dystonia, fatigue	None		↓↓				EM: accumulation of pleomorphic, subsarcolemmal mitochondria	<i>DPYD</i>	hom.	IVS14+1G>A [32]	1	Dihydropyrimidine dehydrogenase deficiency (#274270)
1098	M	6 mo (3 yr)	Hypotonia, global developmental delay, hearing loss, strabismus, ophthalmoplegia, dystonia	None					-	Increased mtDNA: 187% (muscle)	<i>KARS</i>	cmpd. het.	p.T587M p.P228L	3 3	Charcot-Marie-Tooth disease (#613641)
1020	F	Newborn (3 mo)	Failure to thrive, microcephaly, encephalopathy, IUGR, hypotonia, bacteremia, pulmonary hypertension, heart failure	1 sib (hypotonia, microcephaly, seizures) d. 15 mo		↓↓			↓	mtDNA depletion: 42% (muscle) 39% (liver)	<i>ATP5A1*</i>	hom.	p.Y321C	3	na

Patient 1097, of Indian descent, developed multiple neurological symptoms starting at age four, including toe walking, tremors, and dystonia. Muscle tissue showed reduced complex I+III activity and electron microscopy showed accumulation of pleomorphic subsarcolemmal mitochondria, some with complex cristae. Sequencing revealed a homozygous splice mutation in *DPYD* previously associated with recessive DPD deficiency and 5-fluorouracil (5-FU) toxicity.^{32, 35} The diagnosis of DPD deficiency was confirmed by subsequent urine tests that showed highly elevated uracil (395 mmol/mol creatinine; reference range <28) and thymine (300 mmol/mol creatinine; reference range <6).

Patient 1073, of Ashkenazi ancestry, presented with type 1 diabetes at age 33 and developed autonomic dysfunction, cerebellar atrophy, and severe memory loss. Muscle biopsy showed heteroplasmic mtDNA deletions. MitoExome sequencing revealed a homozygous missense mutation in *WFS1* underlying Wolfram syndrome, which was not detected by initial clinical gene sequencing, as previously described.⁴

Patient 1098, of European descent, presented at six months of age with developmental delay, hypotonia, ophthalmoplegia, and other neurological manifestations (Table 2.1), and died by age four. EEG and MRI were unremarkable, however brain stem auditory evoked potential study at 10 months was markedly abnormal, suggesting severe peripheral conduction defects in the auditory pathways bilaterally. Metabolic studies showed elevated plasma alanine (657 umol/L; reference range 142-421) and CSF lactate (2.4 mmol/L; reference range 0.5-2.2); increased mtDNA levels were found in muscle (187% of controls). Sequencing revealed compound heterozygous missense mutations in *KARS*, which encodes both the cytoplasmic and mitochondrial lysyl-tRNA synthetase (Figure A-2). Although formal experimental proof is needed, given the predicted severity of the mutations at highly conserved residues and the previous report of recessive *KARS* mutations in

patients with overlapping symptoms, the observed mutations are likely to underlie our patient's phenotype, as discussed in Appendix A-2.

In addition to five new molecular diagnoses, 11 patients harbored pVUS in established nuclear disease loci (Appendix A-2): seven in recessive-acting genes (four-fold enrichment over background frequency, $p < 0.05$), and four in dominant-acting genes (no enrichment over background) (Methods).

Together, analysis of over 300 established disease loci in the mitochondrial and nuclear genome recovered the diagnoses of 17/18 patients with prior molecular diagnoses and led to new molecular diagnoses in 5/84 previously unsolved cases (Figure 2.2).

Parental sequencing

To assess whether *de novo* mutations could be a major contributor to sporadic disease, we performed MitoExome sequencing on parental DNA from five probands without family history of disease. No confirmed *de novo* mutations were identified.

Prioritized Recessive and X-linked Variants in Genes Not Previously Linked to Disease

We next explored whether disease-causing mutations may lie in 1283 nuclear-encoded candidate genes associated with the mitochondrion but not previously linked to disease.

If candidate genes truly underlie disease, we might expect an increased number of predicted deleterious variants in patients compared to healthy individuals (Figure A-3). We observed modest two-fold enrichment for likely recessive variants ($p < 0.05$) and seven-fold enrichment for hemizygous X-linked variants in male patients ($p < 0.01$). No significant enrichment was observed for prioritized heterozygous variants, X-linked variants in females, or synonymous variants.

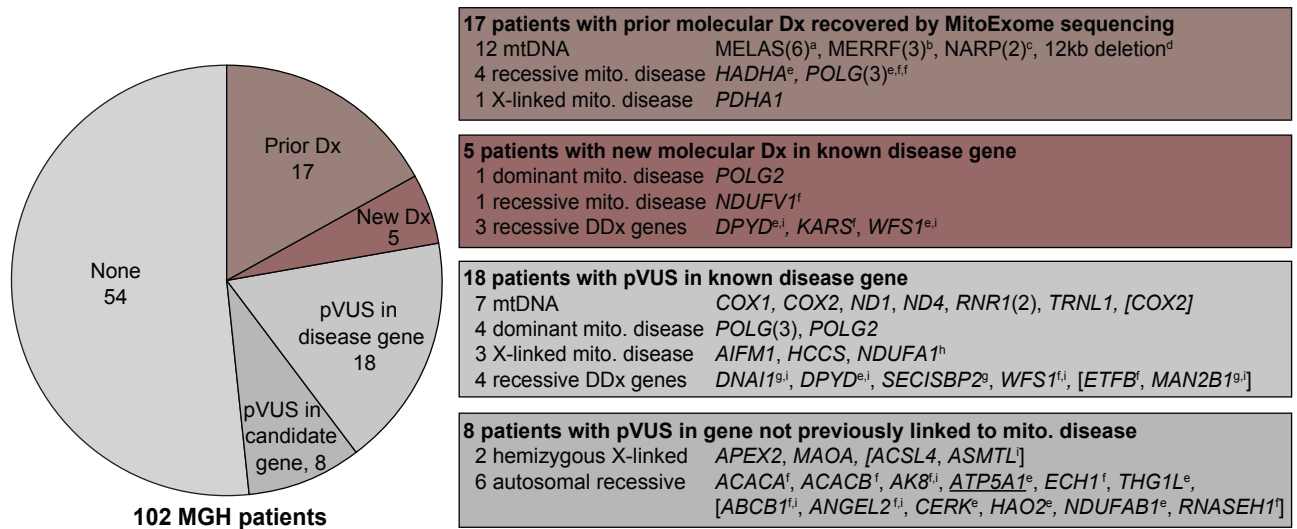


Figure 2.2. MitoExome Sequencing in 102 Patients with Suspected Mitochondrial Disease

Number of patients with molecular diagnoses or prioritized variants of unknown significance (pVUS), with the prioritized loci listed at right. Patients who harbored multiple variants were annotated based on their highest priority variant, with all additional pVUS listed in brackets.

Parentheses indicate number of patients. Underline indicates experimental support of pathogenicity.

^a m.3243A>G; ^b m.8344A>G; ^c m.8993T>G/C; ^d m.3643_15569del; ^e homozygous; ^f compound heterozygous (phased); ^g potential compound heterozygous (unphased); ^h heterozygous in female; ⁱ not in MitoCarta

Altogether, 17 candidate genes not previously linked to mitochondrial disease harbored autosomal recessive or hemizygous X-linked variants (Figure 2.2, Table A-5). We emphasize that due to the modest enrichment over background, only a subset is expected to be causal. Further genetic and experimental studies are needed to establish pathogenicity.

Pathogenicity of *ATP5A1* Mutations in a Yeast Model

In one such case, we investigated prioritized variants detected in an infantile patient with combined oxidative phosphorylation (OXPHOS) deficiency and mtDNA depletion (Figure 2.3). Patient 1020, born to consanguineous first-cousin parents of Moroccan descent, presented at birth with microcephaly, pulmonary hypertension and heart failure requiring extracorporeal membrane oxygenation, hypotonia, hyperalaninemia, and bacteremia. She died at the age of three months. Muscle tissue showed decreased ETC activity (22% complex I+III and 30% complex IV activity) and mtDNA depletion (42% of controls in muscle, 39% in liver). The patient had one healthy sister and one sister who died at 15 months with a similar clinical presentation and combined OXPHOS deficiency in muscle, consistent with a recessive disorder.

MitoExome sequencing revealed a homozygous *ATP5A1*:c.962A>G (p.Y321C) variant, which was not present in 1000 Genomes nor EVS, and introduced a missense mutation at a residue evolutionary conserved to bacteria. The variant was homozygous in the affected sister and heterozygous in their mother. DNA from the father and unaffected sister were unavailable. The only other homozygous variant prioritized in the proband (*HAO2*:p.D259E) did not segregate with disease in the family.

Since cultured patient fibroblasts did not survive lentiviral transduction required for human complementation studies, we modeled the *ATP5A1*:p.Y321C mutation in yeast. This mutation lies within a highly conserved region of the protein associated with mitochondrial genomic integrity (*mg1*)

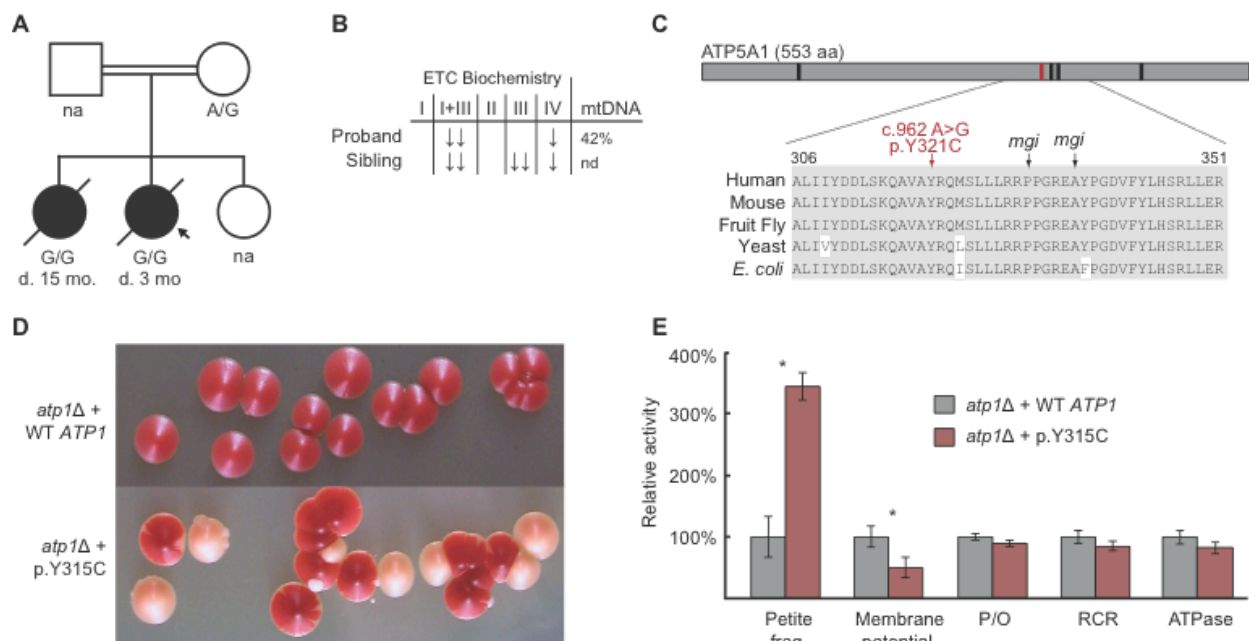


Figure 2.3. Modeling *ATP5A1* Mutations in Yeast

(A) Family pedigree and genotype at *ATP5A1*:c.962A>G. (B) ETC biochemistry and mtDNA quantitation in muscle. (C) Schematic of ATP5A1 protein, with sequence alignment shown in inset. Vertical bars indicate patient mutation (red) and yeast mitochondrial genome integrity (*mgi*) mutations (black). (D) Yeast *atp1Δ* deletion strains show normal (red) and *petite* (white) colonies. (E) Mitochondrial physiology of yeast deletion strains report mean values +/- standard deviation across three replicates. * p<0.01, one-tailed t-test; Abbreviations: na, not available; nd, not determined; d., deceased; ↓↓, <20%, ↓, 20-30% activity relative to citrate synthase.

in yeast; nearby mutations result in decreased coupling of the ATP synthase and lead to a high rate of mtDNA loss (Figure A-4).²⁴ We expressed the analogous yeast variant (Δ *ATP1*:p.Y315C) in an *atp1* knockout strain and observed a three-fold increase in *petite* frequency, reflective of the rate of mtDNA loss, and a two-fold decrease of mitochondrial membrane potential ($p < 0.01$, t-test) compared to expression of wild-type *ATP1*, consistent with reported *mgj* mutants.²⁴

Collectively, the familial segregation, severe amino acid change of a highly conserved residue, and increased rate of mtDNA loss in a yeast model consistent with the patient phenotype strongly support the pathogenicity of this *ATP5A1* mutation.

Predictors of Diagnostic Efficacy

We investigated whether any patient features correlated with the rate of diagnostic success, including family history, age-of-onset, and ETC deficiency. We observed significant enrichment of molecular diagnoses only in cases with clear family history (Table A-7).

Cohort Comparison

Lastly, we compared the rate of molecular diagnosis from this study and our previous sequencing study of 42 infants with severe, biochemically-proven OXPHOS disease.¹³ We limited the analysis to cases without prior molecular diagnoses, and only focused on the subset of 1,034 genes that were analyzed in both studies.¹³ Compared with the infantile cohort, the 84 MGH patients suspected of mitochondrial disorders contained markedly fewer new diagnoses and prioritized recessive variants in mitochondrial proteins not yet linked to disease (Figure 2.4). Similar differences were observed when considering all patients, including those with prior molecular diagnoses (Figure A-5).

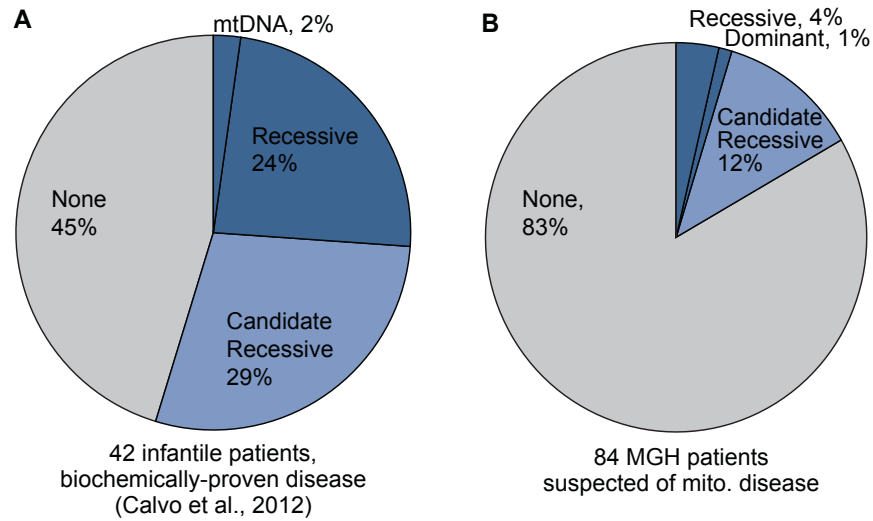


Figure 2.4. New Molecular Diagnoses and Prioritized Recessive Genes

Comparison of MitoExome analysis from our (A) previous study¹³ and (B) current study, showing the fraction of patients with firm molecular diagnoses (dark blue) and autosomal recessive candidate genes (light blue). Comparison is restricted to the subset of patients who lacked molecular diagnosis prior to MitoExome sequencing and to the 1,034 genes analyzed in both studies (thereby excluding 212 DDx genes from the current study).

Discussion

We assessed the efficacy of targeted NGS for molecular diagnosis in 102 patients suspected of a mitochondrial disorder, providing new molecular insights into the genetic basis of mitochondrial disease and lessons for incorporating NGS in the clinic. Unlike our previous study of biochemically-proven infantile cases, the current study includes many patients with late-onset presentations and uncertain clinical diagnoses. MitoExome sequencing recovered 94% of prior molecular diagnoses and yielded molecular diagnoses for 7% of patients refractory to traditional genetic testing, less than a third the diagnostic success rate (24%) observed in our previous study.¹³ Two major possibilities exist to explain the lower yield. First, milder cases may be due not to severe, recessive, coding mutations, but perhaps due to variants with incomplete penetrance, regulatory variants, or genetic interaction of weaker alleles. Second, a subset of our MGH patients may not actually have primary mitochondrial disease, despite clinical and biochemical features used in standard diagnostic algorithms.

It is notable that two proteins implicated in disease localize outside of mitochondria, which may help delineate pathways of cross-talk between mitochondria and other cellular compartments. Mutations in WFS1, a protein localized to the endoplasmic reticulum, have been previously linked with mtDNA deletions.³⁶ DPYD, a cytosolic enzyme in pyrimidine metabolism also responsible for catabolism of the chemotherapeutic drug 5-FU, has not been linked to mitochondrial function, but defects in two adjacent metabolic enzymes cause mtDNA depletion syndromes: TYMP (OMIM #603041) and TK2 (OMIM #609560). Further studies are needed to explore whether mitochondrial defects are typical in DPD deficiency.

Our study identified 26 patients with prioritized variants of unknown significance in known or candidate disease genes. In one case, we provide evidence that recessive mutations in ATP5A1, the alpha subunit of the F₁-F_o ATP synthase, can underlie combined OXPHOS deficiency. This is

the second nuclear-encoded complex V subunit linked to disease.³⁷ Our study suggests that mutations in this complex may have a secondary impact on mtDNA homeostasis and the respiratory chain (Appendix A-2).

A number of lessons have emerged from our study that may be useful as NGS is introduced into the clinic. First, targeted exome sequencing is a cost-effective and time-efficient alternative to traditional, sequential testing of the mtDNA and individual nuclear genes. For a subset of patients, a firm molecular diagnosis can be established quickly in a minimally invasive manner. Prospective studies will be required to determine the efficacy of such technology as a first-line diagnostic test. Second, our study underscores the phenotypic overlap between mitochondrial disorders and other genetic syndromes, and as costs drop, argues for whole exome or whole genome sequencing. Third, when possible we advocate familial sequencing to facilitate phasing of haplotypes, detection of *de novo* variants, and filtering of candidate variants. Fourth, our study underscores the need for guidelines for interpretation of NGS data for mitochondrial disorders. Larger databases of healthy and morbid genomes will aid in future interpretation, and to this end, our data has been deposited into dbGaP. Finally, DNA sequence must not be interpreted alone, but rather in the context of the broader clinical picture. Principled methods combining genomic, biochemical, and clinical data will be required to usher NGS into the clinical arena.

Acknowledgements

We thank Winnie Xu, Rosemary Barone, Mary Anne Anderson and Tammy Gillis for technical assistance; Khalid Shakir, David Jaffe, Mark DePristo, and the Broad GATK team for assistance with data analysis; Irina Anselm for clinical assistance; John Walker, David Holtzman, Elaine Lim, and Scott Vafai for discussion; Mark Daly, Benjamin Neale, the NIMH Schizophrenia Genetics Initiative, and the NIMH Control Sample repository for control exome data. Finally we'd

like to thank the patients and their families for their participation. This work was supported by an NSF Graduate Research Fellowship Program grant (DSL), a grant from NIH R01GM66223 (DMM), and American Recovery and Reinvestment Act (ARRA) funds through grant number RC2HG005556 (VKM) from the National Human Genome Research Institute, National Institutes of Health.

Author contributions

Study design: SEC, VKM; Sequence data generation: SL, BAC, MLB; Sequence data analysis: DSL, SGH, NBG; Variant confirmation: DSL; Patient registry and clinical assessment: NGS, DRT, GTB, JDS, KBS; Biological experiments: KS, DMM; Manuscript preparation: DSL, SEC, VKM.

References

1. DiMauro S, Hirano M, Schon EA. *Mitochondrial Medicine*. New York: Informa Healthcare; 2006.
2. Chinnery PF. Mitochondrial Disorders Overview. In: Pagon RA, Bird TD, Dolan CR, Stephens K, Adam MP, editors. *GeneReviews*. Seattle (WA)1993.
3. Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. *N Engl J Med*. 2012 Mar 22;366(12):1132-41.
4. Lieber DS, Vafai SB, Horton LC, et al. Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease. *BMC Med Genet*. 2012 Jan 6;13(1):3.
5. Dimauro S, Davidzon G. Mitochondrial DNA and disease. *Ann Med*. 2005;37(3):222-32.
6. Chinnery PF, Turnbull DM. Epidemiology and treatment of mitochondrial disorders. *American journal of medical genetics*. 2001 Spring;106(1):94-101.
7. Kirby DM, Crawford M, Cleary MA, Dahl HH, Dennett X, Thorburn DR. Respiratory chain complex I deficiency: an underdiagnosed energy generation disorder. *Neurology*. 1999 Apr 12;52(6):1255-64.
8. Finsterer J, Jarius C, Eichberger H. Phenotype variability in 130 adult patients with respiratory chain disorders. *J Inher Metab Dis*. 2001 Oct;24(5):560-76.
9. Vasta V, Ng S, Turner E, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med*. 2009;1(10):100.
10. Tucker EJ, Hershman SG, Kohrer C, et al. Mutations in MTFMT underlie a human disorder of formylation causing impaired mitochondrial translation. *Cell Metab*. 2011 Sep 7;14(3):428-34.
11. Vasta V, Merritt JL, 2nd, Saneto RP, Hahn SH. Next-generation sequencing for mitochondrial diseases reveals wide diagnostic spectrum. *Pediatr Int*. 2012 Apr 12.
12. Calvo SE, Tucker EJ, Compton AG, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nature genetics*. 2010 Oct;42(10):851-8.
13. Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science translational medicine*. 2012 Jan 25;4(118):118ra10.
14. Mayr JA, Haack TB, Graf E, et al. Lack of the mitochondrial protein acylglycerol kinase causes Sengers syndrome. *American journal of human genetics*. 2012 Feb 10;90(2):314-20.
15. Haack TB, Haberberger B, Frisch EM, et al. Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *Journal of medical genetics*. 2012 Apr;49(4):277-83.
16. Steenweg ME, Ghezzi D, Haack T, et al. Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. *Brain : a journal of neurology*. 2012 May;135(Pt 5):1387-94.
17. Pagliarini DJ, Calvo SE, Chang B, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell*. 2008 Jul 11;134(1):112-23.
18. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*. 2009 Feb;27(2):182-9.
19. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008 Nov 6;456(7218):53-9.
20. Ruiz-Pesini E, Lott MT, Procaccio V, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res*. 2007 Jan;35(Database issue):D823-8.
21. Ingman M, Gyllensten U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res*. 2006 Jan 1;34(Database issue):D749-51.
22. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: 2008 update. *Genome Med*. 2009;1(1):13.

23. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. Oct 28;467(7319):1061-73.
24. Wang Y, Singh U, Mueller DM. Mitochondrial genome integrity mutations uncouple the yeast *Saccharomyces cerevisiae* ATP synthase. *The Journal of biological chemistry*. 2007 Mar 16;282(11):8228-36.
25. Morava E, van den Heuvel L, Hol F, et al. Mitochondrial disease criteria: diagnostic applications in children. *Neurology*. 2006 Nov 28;67(10):1823-6.
26. Bernier FP, Bonch A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology*. 2002 Nov 12;59(9):1406-11.
27. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008 Apr;10(4):294-300.
28. Shanske S, Pancrudo J, Kaufmann P, et al. Varying loads of the mitochondrial DNA A3243G mutation in different tissues: implications for diagnosis. *American journal of medical genetics Part A*. 2004 Oct 1;130A(2):134-7.
29. Craig K, Young MJ, Blakely EL, et al. A p.R369G POLG2 mutation associated with adPEO and multiple mtDNA deletions causes decreased affinity between polymerase gamma subunits. *Mitochondrion*. 2012 Mar;12(2):313-9.
30. Young MJ, Longley MJ, Li FY, Kasiviswanathan R, Wong LJ, Copeland WC. Biochemical analysis of human POLG2 variants associated with mitochondrial disease. *Human molecular genetics*. 2011 Aug 1;20(15):3052-66.
31. Schuelke M, Smeitink J, Mariman E, et al. Mutant NDUFV1 subunit of mitochondrial complex I causes leukodystrophy and myoclonic epilepsy. *Nature genetics*. 1999 Mar;21(3):260-1.
32. Al-Sanna'a NA, Van Kuilenburg AB, Atrak TM, Abdul-Jabbar MA, Van Gennip AH. Dihydropyrimidine dehydrogenase deficiency presenting at birth. *J Inherit Metab Dis*. 2005;28(5):793-6.
33. Barrett TG, Bunday SE, Macleod AF. Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome. *Lancet*. 1995 Dec 2;346(8988):1458-63.
34. McLaughlin HM, Sakaguchi R, Liu C, et al. Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. *American journal of human genetics*. 2010 Oct 8;87(4):560-6.
35. Maring JG, van Kuilenburg AB, Haasjes J, et al. Reduced 5-FU clearance in a patient with low DPD activity due to heterozygosity for a mutant allele of the DPYD gene. *British journal of cancer*. 2002 Apr 8;86(7):1028-33.
36. Rotig A, Cormier V, Chatelain P, et al. Deletion of mitochondrial DNA in a case of early-onset diabetes mellitus, optic atrophy, and deafness (Wolfram syndrome, MIM 222300). *J Clin Invest*. 1993 Mar;91(3):1095-8.
37. Mayr JA, Havlickova V, Zimmermann F, et al. Mitochondrial ATP synthase deficiency due to a mutation in the ATP5E gene for the F1 epsilon subunit. *Human molecular genetics*. 2010 Sep 1;19(17):3430-9.
38. Cano A, Rouzier C, Monnot S, et al. Identification of novel mutations in WFS1 and genotype-phenotype correlation in Wolfram syndrome. *American journal of medical genetics Part A*. 2007 Jul 15;143A(14):1605-12.

CHAPTER III:

Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease

Contributors: Daniel S. Lieber*, Scott B. Vafai*, Laura C. Horton, Nancy G. Slate, Shangtao Liu, Mark L. Borowsky, Sarah E. Calvo, Jeremy D. Schmahmann, Vamsi K. Mootha

** These authors contributed equally*

This chapter originally appeared as Lieber et al., “Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease.” *BMC Medical Genetics* (2012)

Abstract

Background: Mitochondrial diseases comprise a diverse set of clinical disorders that affect multiple organ systems with varying severity and age of onset. Due to their clinical and genetic heterogeneity, these diseases are difficult to diagnose. We have developed a targeted exome sequencing approach to improve our ability to properly diagnose mitochondrial diseases and apply it here to an individual patient. Our method targets mitochondrial DNA (mtDNA) and the exons of 1,600 nuclear genes involved in mitochondrial biology or Mendelian disorders with multi-system phenotypes, thereby allowing for simultaneous evaluation of multiple disease loci.

Case Presentation: Targeted exome sequencing was performed on a patient initially suspected to have a mitochondrial disorder. The patient presented with diabetes mellitus, diffuse brain atrophy, autonomic neuropathy, optic nerve atrophy, and a severe amnesic syndrome. Further work-up revealed multiple heteroplasmic mtDNA deletions as well as profound thiamine deficiency without a clear nutritional cause. Targeted exome sequencing revealed a homozygous c.1672C>T (p.R558C) missense mutation in exon 8 of *WFS1* that has previously been reported in a patient with Wolfram syndrome.

Conclusion: This case demonstrates how clinical application of next-generation sequencing technology can enhance the diagnosis of patients suspected to have rare genetic disorders. Furthermore, the finding of unexplained thiamine deficiency in a patient with Wolfram syndrome suggests a potential link between *WFS1* biology and thiamine metabolism that has implications for the clinical management of Wolfram syndrome patients.

Introduction

Mitochondrial diseases are a clinically heterogeneous set of disorders that are caused by defects in mitochondria, the organelles responsible for producing most of the cellular ATP in humans ¹. Multiple organ-systems are typically affected in these disorders, with neurologic and myopathic features being the most prominent ². Common clinical features of mitochondrial disease include skeletal myopathy accompanied by exercise intolerance, cardiomyopathy, sensorimotor peripheral polyneuropathy, sensorineural deafness, optic atrophy, diabetes mellitus, seizures, and ataxia ³. These disorders can be caused by mutations in the nuclear or mitochondrial genomes, and over 100 loci have been identified to date ⁴.

The clinical and genetic heterogeneity of mitochondrial diseases as well as the technical difficulty of assessing mitochondrial function create a significant diagnostic challenge ⁵. One major source of cost and time delays in the clinical evaluation of patients with suspected mitochondrial disease is the use of sequential genetic tests to evaluate individual disorders; as such, a systematic genetic test could accelerate the diagnostic process and reduce overall cost. In the present report, we describe a suspected case of mitochondrial disease whose diagnosis of Wolfram syndrome was ultimately revealed through targeted exome sequencing.

Case Presentation

In January 2008, a 61-year-old man with a history notable for diabetes mellitus (DM), autonomic neuropathy, diffuse brain atrophy, optic nerve atrophy (OA), and profound amnesia was referred to us to establish neurologic care. The patient carried a diagnosis of multiple system atrophy- cerebellar type (MSAc), principally because of severe cerebellar and brainstem atrophy on MRI.

The patient's early history was remarkable only for childhood bedwetting and urinary urgency as a young adult. He was otherwise well during this time and was a talented athlete who completed college and practiced as an accountant. In his early 20s, he developed bladder dysfunction of unclear etiology requiring intermittent straight catheterization, as well as erectile dysfunction.

At age 33, he was diagnosed with DM, presumed to be type 1, and began treatment with insulin therapy. Although there is no biochemical data available from the time of his original diagnosis, recent testing demonstrated a random C-peptide level of 0.6 ng/mL (reference range 0.9 to 4.3 ng/mL) at a time when his blood glucose was 83 mg/dL. He takes an average of 24 units of insulin per day, and has had good glycemic control with hemoglobin A1c measurements ranging between 6.5 and 7.2% over the last several years. He has had no evidence of retinopathy, or other microvascular or macrovascular complications. He had polyuria and polydipsia at the time of his initial DM diagnosis, but these symptoms resolved once he initiated insulin therapy.

The patient began dressing in strange colors in his 30s, and color blindness was ultimately diagnosed in his 40s. At age 53, the patient presented for a routine screening ophthalmology exam and was discovered to have bilateral OA with preserved vision. Brain MRI at that time revealed severe atrophy of the cerebellar hemispheres and vermis, pons, and middle cerebellar peduncles as well as moderate cerebral atrophy; a more recent study at age 61 showed these findings as well as more severe cerebral atrophy (Figure 3.1). Despite these radiographic findings, the patient and his wife reported no gait instability or upper extremity incoordination.

During his late 50s, the patient's neurologic status deteriorated. Formal neuropsychological evaluation revealed profound anterograde amnesia, with additional impairments in cognitive flexibility, executive function, naming, and high order visual processing skills. Attention span, mental tracking, verbal abstract reasoning, complex auditory instructions, and visual spatial functions were

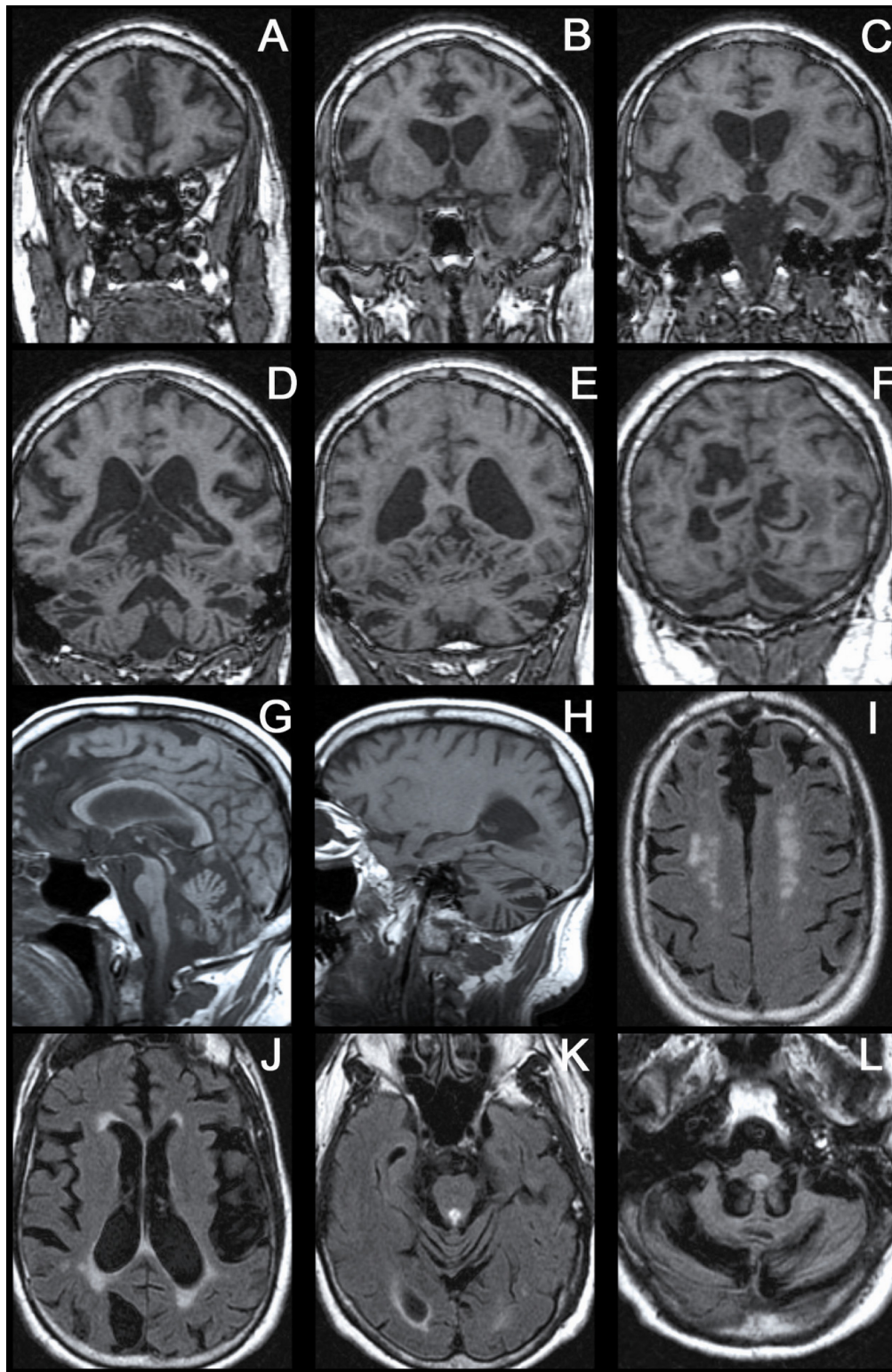


Figure 3.1. MRI at age 61 demonstrating severe atrophy of the cerebral hemispheres, cerebellum, and brainstem. A-F: 3-D SPGR coronal MRI; G,H: T1 sagittal MRI; I-L: FLAIR axial MRI, showing white matter signal hyperintensities.

preserved. From a psychiatric perspective, he developed symptoms of depression, which responded to treatment with sertraline.

In parallel with the decline in his memory, the patient also developed progressive autonomic neuropathy, with gastroparesis and severe postural hypotension. The autonomic dysfunction exceeded what might be expected from his diabetes mellitus, given his good glycemic control and the absence of other diabetic complications. His bladder dysfunction worsened and he required suprapubic catheter placement at the age of 61. Due to his multiple functional deficits, the patient became unable to work and is now completely reliant upon his wife for care.

Regarding his family history, the patient was born to Ashkenazi Jewish parents and there was no parental consanguinity. His mother died from melanoma, and his father died from multiple strokes and a myocardial infarction. He has two adult daughters, one of whom has attention deficit hyperactivity disorder (ADHD) and Tourette syndrome, while the other suffers from chronic urinary tract infections. His maternal grandmother had type 2 DM, and a maternal first cousin had type 1 DM. No other close relatives have suffered from endocrine, psychiatric, or neurologic disease.

On physical examination, he appeared generally medically well. He weighed 79 kilograms and was 178 cm tall, yielding a body mass index of 25. Postural hypotension was evident with systolic blood pressure falling from 150 to 95 after one minute of standing, though asymptomatic. Funduscopic examination revealed optic atrophy bilaterally with no sign of diabetic retinopathy. Visual acuity was 20/40 in each eye. Pupil responses to light and accommodation were normal. Eye movements were normal with the exception of saccadic intrusion into horizontal smooth pursuit. Clinical examination revealed high tone hearing loss bilaterally. Audiometry demonstrated moderate sensorineural hearing loss in the high frequencies on the left, and mild sensorineural hearing loss in the mid-frequencies on the right sloping to a severe loss in the high frequencies (Figure 3.2). Word

recognition was excellent in both ears; 98% on the right and 96 % on the left. Muscle tone in the extremities was normal, bulk was intact, and strength was full. There was no evidence of dysmetria with finger-to-nose and heel-to-shin testing, and gait was slow but stable. His affect was flat and he was passive throughout the interview, speaking only when spoken to. He was not oriented to time or place. He could repeat four words, but could not learn them despite multiple attempts. He was unable to provide information concerning major current political or national news. He could, however, recall sizable fragments of remote memory from his college years.

The absence of the cerebellar motor syndrome and the presence of a profound amnesic syndrome on examination called the patient's diagnosis of MSAc into question ⁶, and we undertook re-evaluation of his case to explore alternate diagnoses. His laboratory work-up revealed an undetectable thiamine level, a surprising finding given his normal diet and the absence of alcohol abuse. We ascribed his amnesic disorder to presumed long-standing thiamine deficiency, but repletion produced minimal clinical impact. The involvement of multiple systems suggested the possibility of a mitochondrial disorder. Genetic testing for OPA1, MELAS, MERFF, LHON and NARP were negative, however analysis of mitochondrial DNA (mtDNA) from a muscle biopsy sample by both Southern blotting and PCR analysis revealed multiple heteroplasmic deletions. Biochemical testing revealed a minor defect in complex I of the electron transport chain. COX and SDH staining of the muscle biopsy specimen were unremarkable, and the mitochondria appeared grossly normal on electron microscopic examination. Occasional central vacuoles and tubular aggregates were seen in the myocytes, which were felt to be consistent with a mild non-specific myopathy.

Given the diagnostic uncertainty and concern for a mitochondrial disorder, the patient was enrolled in the mitochondrial disease registry at Massachusetts General Hospital. As part of this program, a sample of the patient's DNA from whole blood underwent targeted exome

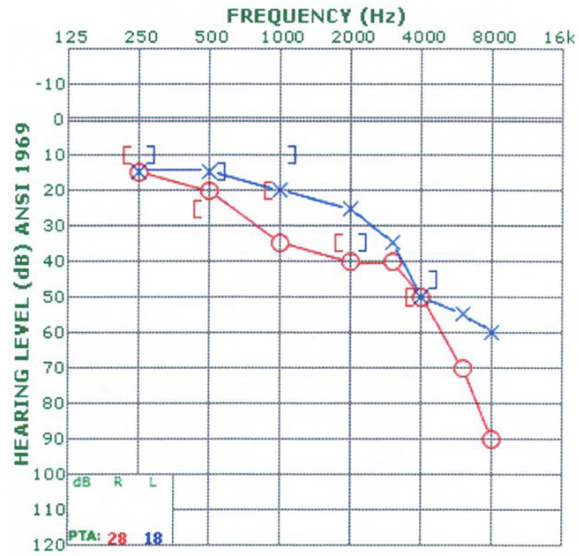


Figure 3.2. Results of audiometric testing at age 61 conforming to ANSI 1969 standards.

ANSI=American National Standards Institute; dB=decibel; Hz=hertz; PTA=pure tone average.

("MitoExome") sequencing. Mitochondrial DNA and the exons of 1,600 nuclear genes either encoding mitochondrial proteins or implicated in Mendelian disorders with multi-system phenotypes were targeted using hybrid selection ⁷. Amplified targets were sequenced on the Illumina GAIIx platform. Rare, protein-modifying variants found to be homozygous or potentially compound heterozygous were prioritized (Figure 3.3), revealing an X-linked functional polymorphism c.937G>T (p.D313Y) in *GLA* that is not considered pathogenic ⁸ and a homozygous c.1672C>T (p.R558C) missense mutation in exon 8 of *WFS1* that has previously been reported in a patient with Wolfram syndrome ⁹. No heteroplasmic mtDNA deletions were detected in whole blood. The patient's *WFS1* mutation was verified through Sanger sequencing in a CLIA-certified laboratory, though not without complications; the initial report came back negative and only after requested follow-up was the homozygous mutation detected, thereby confirming the diagnosis of Wolfram syndrome.

Discussion and Conclusions

Wolfram syndrome (OMIM #222300) is an autosomal recessive disorder also known as DIDMOAD for its typical clinical features of Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy and Deafness.¹⁰ The minimum criteria for the diagnosis of Wolfram syndrome are the presence of diabetes mellitus and optic atrophy, both of which usually occur by the age of fifteen.¹¹ The disorder is also associated with diverse neurologic and psychiatric symptoms, including cognitive impairment.¹² Though with genetic data in hand the diagnosis of Wolfram syndrome now seems clear for the patient presented here, his late age of onset and the prominence of his amnesia are quite atypical for this syndrome and made the diagnosis seem less likely. Furthermore, the clinical presentation, mtDNA deletions, and biochemical data were highly suggestive of a mitochondrial disorder.

Targeted bases	3.13 Mb
Total Reads	11.4 million
Aligned high quality bases	1.52 Gb
Bases covered at >10x	94.9%
Mean target coverage	272x
Mean mtDNA coverage	4,557x

<u>Category</u>	<u># Variants (% Total)</u>
Total	1,573 (100%)
Rare	55 (3.5%)
Protein-modifying	30 (2%)
Homozygous or two heterozygous	1 (.1%), <i>WFS1</i>




Figure 3.3. Results of MitoExome sequencing. Rare, protein-modifying variants were prioritized, revealing a homozygous c.1672C>T (p.R558C) missense mutation in exon 8 of *WFS1* that has previously been reported in a patient with Wolfram syndrome.⁹

Given the multiple organ system dysfunction characteristic of Wolfram syndrome, it was initially thought by a number of clinicians to be a *bona fide* mitochondrial disorder¹³. Further supporting this hypothesis was the observation of mtDNA deletions in some patients with Wolfram syndrome¹⁴⁻¹⁶. However, in 1998 two groups identified *WFS1*, which encodes an 890 amino acid trans-membrane protein localized to the endoplasmic reticulum, as the gene responsible for the majority of Wolfram syndrome cases, thereby arguing against a mitochondrial etiology^{17, 18}. Mechanistic studies of *WFS1* have revealed an important role in dampening the ER stress response in pancreatic beta cells and neurons, and loss of the gene has been shown to result in an exaggerated stress response and accelerated cell death¹⁹. However, given the shared clinical features between Wolfram syndrome and many mitochondrial disorders, as well as the presence of mtDNA deletions in some patients, many unanswered questions remain about the biology of *WFS1* and its impact on mitochondrial function and mtDNA maintenance.

Mutations in residue 558 of *WFS1*, as identified in the patient described here, have been previously reported. Our patient's p.R558C mutation was first reported as a heterozygous variant in a cohort of psychiatric patients, though it is unclear whether the variant contributed to disease pathogenesis in that population²⁰. Our patient's p.R558C mutation, as well as a p.R558H mutation, have been reported in multiple patients with Wolfram syndrome^{9, 21, 22}. Exome databases suggest that the allele frequency of *WFS1* c.1672C>T (p.R558C) is quite rare; specifically, the variant is not found in the 1,000 Genomes Project²³ nor in over 2,500 publicly available European and African American exomes²⁴.

Chausseot *et al.* specifically discuss the phenotype of a Wolfram case with the same p.R558C mutation identified in our patient in compound heterozygosity with a p.F354del frameshift mutation¹². Like our patient, Chausseot *et al.*'s case had significant cerebellar atrophy by MRI, neurogenic bladder symptoms, dementia, DM, OA, as well as the absence of diabetes insipidus.

However, unlike our patient, the compound heterozygote case had earlier onset of DM, OA, and neurologic symptoms (at ages 9, 10, and 27 respectively), lacked hearing impairment, and developed cerebellar ataxia and nystagmus as part of her neurologic presentation. Of note, the profound amnesic syndrome seen in our patient was not reported for the compound heterozygote case, and is atypical for Wolfram syndrome patients in general. These cases demonstrate that phenotypic heterogeneity is present even among patients with similar mutations, likely attributable to different genetic and environmental backgrounds.

Some *WFS1* mutations have been suggested to function in an autosomal dominant manner yielding symptoms such as optic atrophy and hearing impairment in a heterozygous state^{25,26}. The patient has one daughter who struggled with Tourette syndrome and ADHD, but the family history is otherwise unremarkable, with no close family members suffering from symptoms associated with Wolfram syndrome. No familial DNA was available to determine the carrier status of his daughter. It is unlikely that his daughter's symptoms are related to *WFS1*, and we suspect that the p.R558C variant likely acts in a recessive manner.

One of the most striking aspects of the case presented here is the profound thiamine deficiency, and its devastating impact on the patient's memory. The symptom complex of diabetes mellitus, deafness, and optic atrophy has been found in association with defects in thiamine metabolism previously in the context of thiamine-responsive megaloblastic anemia syndrome, or Rogers syndrome (OMIM # 249270)²⁷⁻²⁹. Rogers syndrome, a recessive condition caused by mutations in the thiamine transporter *SLC19A2*, is characterized by thiamine-responsive megaloblastic anemia, sensorineural deafness, and DM; its phenotypic similarity to Wolfram syndrome suggests the possibility that disruption of thiamine metabolism may contribute to the pathophysiology of Wolfram syndrome. The coding exons of *SLC19A2* were sequenced in our patient and no rare mutations were detected, although the presence of non-coding variants cannot

be excluded. Given that there was no dietary reason for our patient to develop thiamine deficiency, his case suggests that *WFS1* may affect thiamine transport or utilization. Furthermore, given that there are no therapies available for the treatment of Wolfram syndrome, assessment of thiamine sufficiency and possibly empiric thiamine supplementation should be considered for these patients.

The patient presented here had a diagnostic journey that is not unusual for disorders involving dysfunction across multiple organ systems, in that several years elapsed and numerous costly genetic tests were performed without a molecular diagnosis. The application of MitoExome sequencing in this case illustrates how clinical application of next-generation sequencing technology can allow for rapid, simultaneous evaluation for multiple genetic disorders with a single test thereby accelerating the diagnostic process. Furthermore, this case demonstrates how careful study of an individual patient with a rare disorder can yield potentially new insight into gene function and disease pathogenesis.

Consent

Written informed consent was obtained from the patient and his wife for publication of this case report.

List of abbreviations

(mtDNA): mitochondrial DNA; (MRI): Magnetic Resonance Imaging; (MSAc): multiple system atrophy, cerebellar-type; (DM): diabetes mellitus; (DI): diabetes insipidus; (OA): optic atrophy; attention deficit hyperactivity disorder (ADHD)

Competing Interests

The authors declare that they have no competing interests.

Author contributions

SBV & JDS looked after the patient. DSL, NGS, SL, MLB, & SEC generated and analyzed the sequence data. DSL, SBV, LCH, JDS, & VKM wrote the report. All authors read and approved the final manuscript

Acknowledgements and Funding

This research was supported by NIH/NHGRI grant RC2HG005556 (VKM), NIH/NIDDK grant T32 DK007028-37 (SBV), an NSF Graduate Research Fellowship Program grant (DSL), and the MINDlink and Birmingham Foundations (JDS). The authors would like to thank Molly Plovovich and Dr. Joseph Avruch for comments on the manuscript.

References

1. Calvo SE, Mootha VK. The mitochondrial proteome and human disease. *Annu Rev Genomics Hum Genet.* 2010 Sep 22;11:25-44.
2. Morava E, van den Heuvel L, Hol F, et al. Mitochondrial disease criteria: diagnostic applications in children. *Neurology.* 2006 Nov 28;67(10):1823-6.
3. Chinnery PF. Mitochondrial Disorders Overview 2010. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20301403.
4. Tucker EJ, Compton AG, Thorburn DR. Recent advances in the genetics of mitochondrial encephalopathies. *Curr Neurol Neurosci Rep.* Jul;10(4):277-85.
5. Bernier FP, Bonch A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology.* 2002 Nov 12;59(9):1406-11.
6. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology.* 2008 Aug 26;71(9):670-6.
7. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* 2009 Feb;27(2):182-9.
8. Froissart R, Guffon N, Vanier MT, Desnick RJ, Maire I. Fabry disease: D313Y is an alpha-galactosidase A sequence variant that causes pseudodeficient activity in plasma. *Mol Genet Metab.* 2003 Nov;80(3):307-14.
9. Cano A, Rouzier C, Monnot S, et al. Identification of novel mutations in WFS1 and genotype-phenotype correlation in Wolfram syndrome. *American journal of medical genetics Part A.* 2007 Jul 15;143A(14):1605-12.

10. Tranebjaerg L, Barrett T, Rendtorff ND. WFS1-Related Disorders 1993. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20301750.
11. Rigoli L, Lombardo F, Di Bella C. Wolfram syndrome and WFS1 gene. *Clin Genet*. 2011 Feb;79(2):103-17.
12. Chaussenot A, Bannwarth S, Rouzier C, et al. Neurologic features and genotype-phenotype correlation in Wolfram syndrome. *Ann Neurol*. 2011 Mar;69(3):501-8.
13. Bu X, Rotter JI. Wolfram syndrome: a mitochondrial-mediated disorder? *Lancet*. 1993 Sep 4;342(8871):598-600.
14. Rotig A, Cormier V, Chatelain P, et al. Deletion of mitochondrial DNA in a case of early-onset diabetes mellitus, optic atrophy, and deafness (Wolfram syndrome, MIM 222300). *J Clin Invest*. 1993 Mar;91(3):1095-8.
15. Barrientos A, Casademont J, Saiz A, et al. Autosomal recessive Wolfram syndrome associated with an 8.5-kb mtDNA single deletion. *Am J Hum Genet*. 1996 May;58(5):963-70.
16. Barrientos A, Volpini V, Casademont J, et al. A nuclear defect in the 4p16 region predisposes to multiple mitochondrial DNA deletions in families with Wolfram syndrome. *J Clin Invest*. 1996 Apr 1;97(7):1570-6.
17. Strom TM, Hortnagel K, Hofmann S, et al. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum Mol Genet*. 1998 Dec;7(13):2021-8.
18. Inoue H, Tanizawa Y, Wasson J, et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet*. 1998 Oct;20(2):143-8.
19. Fonseca SG, Ishigaki S, Osowski CM, et al. Wolfram syndrome 1 gene negatively regulates ER stress signaling in rodent and human cells. *J Clin Invest*. 2010 Mar 1;120(3):744-55.
20. Torres R, Leroy E, Hu X, et al. Mutation screening of the Wolfram syndrome gene in psychiatric patients. *Mol Psychiatry*. 2001 Jan;6(1):39-43.
21. Colosimo A, Guida V, Rigoli L, et al. Molecular detection of novel WFS1 mutations in patients with Wolfram syndrome by a DHPLC-based assay. *Hum Mutat*. 2003 Jun;21(6):622-9.
22. Smith CJ, Crock PA, King BR, Meldrum CJ, Scott RJ. Phenotype-genotype correlations in a series of wolfram syndrome families. *Diabetes Care*. 2004 Aug;27(8):2003-9.
23. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061-73.
24. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*. 2010 Jan;42(1):30-5.
25. Rendtorff ND, Lodahl M, Boulahbel H, et al. Identification of p.A684V missense mutation in the WFS1 gene as a frequent cause of autosomal dominant optic atrophy and hearing impairment. *Am J Med Genet A*. 2011 Jun;155A(6):1298-313.
26. Eiberg H, Hansen L, Kjer B, et al. Autosomal dominant optic atrophy associated with hearing impairment and impaired glucose regulation caused by a missense mutation in the WFS1 gene. *J Med Genet*. 2006 May;43(5):435-40.
27. Borgna-Pignatti C, Marradi P, Pinelli L, Monetti N, Patrini C. Thiamine-responsive anemia in DIDMOAD syndrome. *J Pediatr*. 1989 Mar;114(3):405-10.
28. Lagarde WH, Underwood LE, Moats-Staats BM, Calikoglu AS. Novel mutation in the SLC19A2 gene in an African-American female with thiamine-responsive megaloblastic anemia syndrome. *Am J Med Genet A*. 2004 Mar 15;125A(3):299-305.
29. Rindi G, Casirola D, Poggi V, De Vizia B, Patrini C, Laforenza U. Thiamine transport by erythrocytes and ghosts in thiamine-responsive megaloblastic anaemia. *J Inher Metab Dis*. 1992;15(2):231-42.

CHAPTER IV:
**Targeted exome sequencing reveals HSD17B4-deficiency
in a male with cerebellar ataxia, peripheral
neuropathy, azoospermia, and hearing loss**

Contributors: Daniel S. Lieber*, Steven G. Hershman*, Nancy G. Slate, Sarah E. Calvo,
Katherine B. Sims, Jeremy D. Schmahmann, Vamsi K. Mootha

** These authors contributed equally*

Abstract

Objective: To describe a patient with a previously undiagnosed neurological disorder whose diagnosis was informed through targeted exome sequencing with copy number variant analysis.

Design: Case report.

Setting: Tertiary referral center.

Patient: 35-year-old man with cerebellar ataxia, peripheral neuropathy, hearing loss, cognitive impairment, and azoospermia.

Main Outcome Measures: Clinical features, neuroimaging and metabolic findings, muscle biopsy with biochemical and histochemical analysis, and genetic studies.

Results: Commercially available genetic tests for inherited ataxias and mitochondrial disorders were negative. Serum, urine, and muscle biopsy findings pointed to a mitochondrial disorder while elevated ratios of pristanic:phytanic acid and arachidonic:docosahexaenoic acid were consistent with a peroxisomal disorder. Targeted exome sequencing and a computational algorithm to infer copy number variants revealed compound heterozygous mutations in *HSD17B4*, which encodes the peroxisomal D-bifunctional protein, including a 12kb in-frame deletion of exons 10-13 and a missense variant (p.A196V) within an NAD⁺ binding domain at a residue conserved to bacteria. Recessive mutations in *HSD17B4* are a known cause of peroxisomal D-bifunctional protein deficiency (OMIM #261515) and have been reported in sisters with ataxia, hearing loss, and ovarian dysgenesis (Perrault syndrome; OMIM #233400). Our patient's phenotypic overlap with previous cases, severity of mutations in *HSD17B4*, and metabolic evidence of peroxisomal dysfunction, led to a diagnosis of Perrault syndrome due to HSD17B4 deficiency.

Conclusions: This is the first reported case of a male with ataxia and infertility due to recessive HSD17B4 mutations, expanding the phenotype of Perrault syndrome. Our study highlights the importance of copy number variants in the diagnosis of neurological disorders, and points to potential crosstalk between mitochondria and peroxisomes in the pathogenesis of HSD17B4 deficiency and Perrault syndrome.

Introduction

Patients with mitochondrial disorders frequently present with cerebellar ataxia, often together with other neurological and non-neurological symptoms.¹ Such disorders can be caused by mutations in the mitochondrial DNA or by recessive, dominant, or X-linked mutations in nuclear-encoded genes essential to mitochondrial respiratory chain function.² Due to genetic and phenotypic heterogeneity, single gene testing is often ineffective and targeted sequencing approaches have been utilized to identify the genetic basis of disease in infantile- and adult-onset cases of suspected mitochondrial disease.³ Here, we performed targeted exome analysis in an adult with cerebellar ataxia and other neurological and non-neurological features.

Patient

A 35-year-old man presented for evaluation of a gait disorder progressing since childhood, cognitive impairment, and sensorineural hearing loss. He had mildly delayed developmental milestones, including sitting at 8 months, walking at 14 months, talking at 18 months, and difficulty reading at age 6. Running was impaired at age 7, and sports impossible by age 9. Progressive ataxia necessitated a cane by age 18, wheelchair-dependence by 29. He needed assistance academically through school, and graduated from college in his 30's. Sensorineural hearing loss was noted at age 34. Review of medical charts revealed documented azoospermia. He was the product of neurologically healthy, non-consanguineous parents, with one healthy, fertile sister. The patient's father died of cancer in his early 60's.

Neurological examination revealed high arched feet, hammer toes, and normal secondary sexual characteristics. Oculomotor examination showed square wave jerks at rest, gaze-evoked nystagmus, saccadic intrusions into pursuit, hypermetric saccades, and failure to suppress the vestibular ocular reflex. There was mild dysarthria, moderate upper and lower extremity dysmetria,

and moderate gait ataxia (Brief Ataxia Rating Scale: 13/30). Deep tendon reflexes were normal in the arms, exaggerated in the legs. Plantar responses were flexor. He had impaired pin sense in the feet. Hearing was intact on office testing. He had an average IQ, bland affect, reduced visual motor processing speed, and was impaired on a card-sorting test.

Brain MRI from age 14 through 35 showed progressive cerebellar volume loss (Figure 4.1). Nerve conduction studies revealed absent long latency responses in the legs, symmetrically reduced motor conduction velocities in arms and legs with normal amplitudes; needle examination was normal. Waking electroencephalogram was normal. Audiology testing revealed mild decrease in hearing for higher frequencies bilaterally, decreased speech intelligibility, and word recognition 86% on the right, and 94% on the left.

Pertinent normal laboratory studies included thyroid functions, vitamins E and B12, copper, ceruloplasmin, serum protein electrophoresis, serum tissue transglutaminase and gliadin antibodies, RPR, ANA, and lactate. Normal genetic tests included SCA1, 2, 3, 5, 6, 7, 8, 10, 14, 17, DRPLA, FRDA, AOA1, 2, POLG, MELAS, MERRF and NARP. Endocrine tests with normal results were LH 11 mIU/mL (2-12), somatomedin 112 ng/mL (114-492), and prolactin 11 ng/mL (2-19).

Abnormal laboratory results included low testosterone at 164 ng/dL (270-1100), increased FSH at 15 mIU/mL (1-8), and elevated pyruvate 0.18 (0.08-0.16). A number of serum amino acids were elevated, including alanine 638 (146-494), asparagine 121 (26-92), glutamic acid 75 (6-62), isoleucine 103 (39-90), lysine 260 (119-243), methionine 38 (13-37), proline 310 (97 – 297), and valine 365 (172 – 335). Elevated urine organic acids were present including small amounts of lactic, succinic, and 2-oxoglutaric acids. Tests of long- and very long-chain fatty acids revealed elevated total plasma ω -9 fatty acids. The percentage of C18:2 ω 6 (linoleic) was low and the ratio of pristanic acid / phytanic acid, percentage of arachidonic acid, and arachidonic/ docosahexenoic acid ratio were elevated.

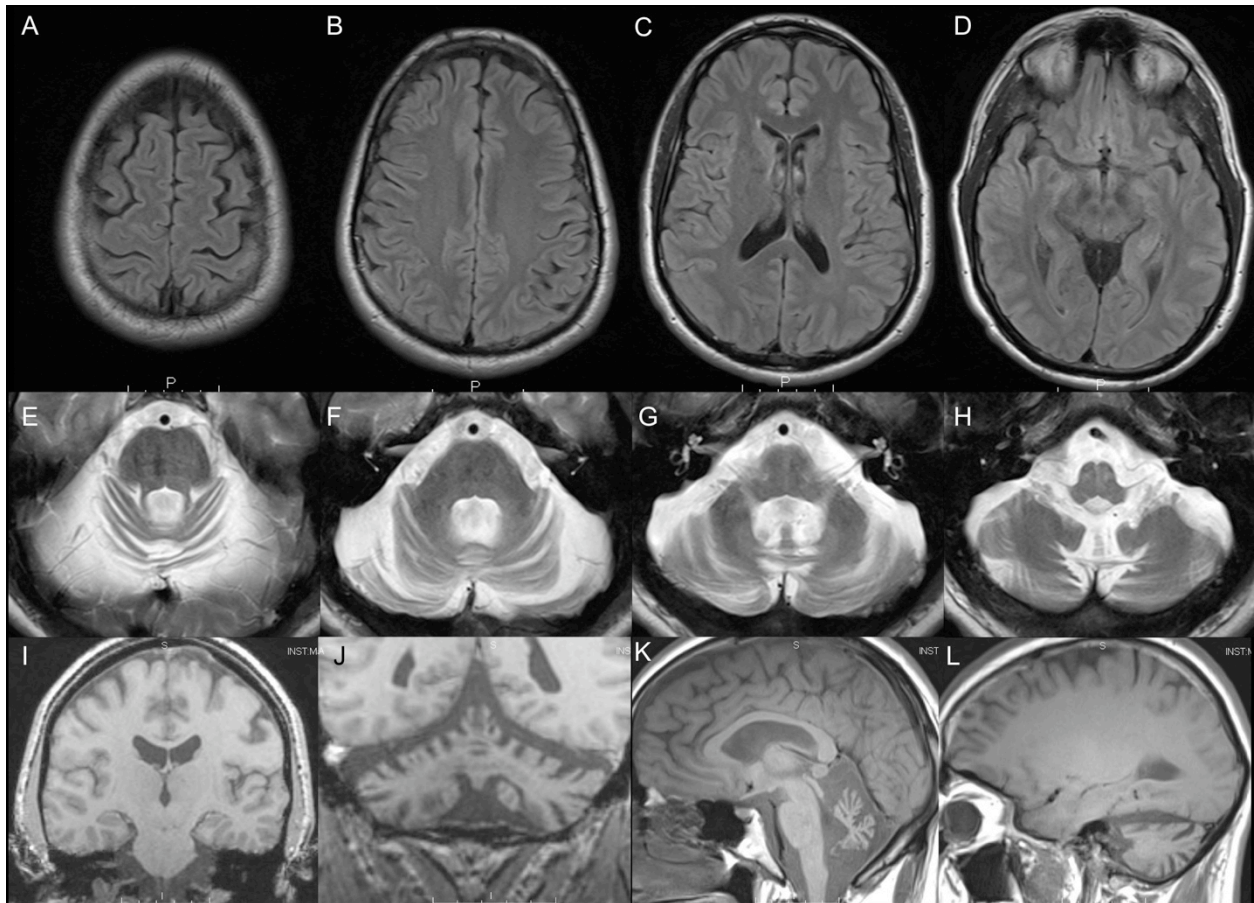


Figure 4.1. MRI at age 35 demonstrating cerebellar volume loss with preservation of cerebral hemispheres. A-D) FLAIR axial; E-H) T2 axial; I-J) Coronal T1; K-L) Sagittal T1

Muscle biopsy was normal on light and electron microscopy. Muscle coenzyme Q10, free and total carnitine, and acylcarnitine were normal, but electron transport chain testing revealed mild Complex I deficiency (22% Complex I+III after normalization to citrate synthase).

Given a classification of “probable” mitochondrial disease based on clinical, metabolic, and biochemical features,⁴ informed consent was obtained from the patient and his family to undergo targeted DNA sequencing for suspected mitochondrial disease.

Methods

Genotyping

PCR of deletion (hg19 chr5:118825016-118837334del_insA; NM_000414:c.715-1207del) was performed using AccuPrime kit (Life Technologies) with extension time 80s and annealing temperature of 55 degrees Celsius. The following primers were used, as depicted in Figure 4.2E: GGGAAATGAATGGTACCCAGAT (red), GCTTTCATTTTGAACATGGTTG (blue), TTTTACCATCCACAGGCTCAC (black).

AccuPrime PCR of SNV (NM_000414:c.587C>T) was performed using an extension time of 1 minute and an annealing temperature of 54 degrees Celsius. The following primers were used: AGTTGCTTTTGGATAGGTGCAG (forward) & CCCCATTTGTGTA AAACAAAAA (reverse).

Multiple sequence alignment

Multiple alignment of HSD17B4 protein in Figure 4.2A was created by aligning the following Uniprot IDs with ClustalW⁵: P51659, P51660, Q98TA2, Q9NKW1, Q9VXJ0, Q02207, Q9I4V1.

Variant annotation

Genetic variants were annotated with predictions from PolyPhen2⁶, and variant frequencies from Exome Variant Server⁷ and 1000 Genomes⁸.

Results

Targeted capture of mitochondrial DNA and the exons of 1598 nuclear genes implicated in mitochondrial biology, mitochondrial disease, or neurological disorders with phenotypic overlap was performed as previously described.³ Variant calling through the Genome Analysis Toolkit identified 1,569 single nucleotide variants (SNVs) and small insertions/deletions (indels) in the patient's sample with 205x mean coverage and 95.1% of targeted bases covered at >10x. We applied filters to identify potentially pathogenic variants including mtDNA variants previously known to be pathogenic as well as rare, protein-modifying variants that were hemizygous X-linked, autosomal recessive, or present in dominant-acting disease genes, as previously described.³ No variants passed our filters and a genetic basis of disease remained unclear.

As our previous analyses focused only on SNVs and indels, we subsequently applied CONIFER to identify large copy number variants (CNVs) in our targeted exome data.⁹ CNV analysis revealed a potential 13kb heterozygous in-frame deletion of exons 10-13 of *HSD17B4* (c.715-1207del; p.239_403del) (Figure 4.2). Review of patient exome data revealed that the patient also harbored a rare, heterozygous missense variant in *HSD17B4* (c.587C>T; p.A196V). This variant was not present in EVS nor 1000 Genomes, altered a residue conserved to bacteria within a predicted NAD-binding domain, and is predicted to be “probably damaging” by PolyPhen2. Familial genotyping confirmed the presence of an inherited heterozygous deletion and revealed that the variants were compound heterozygous in the proband: the deletion, which was heterozygous in the proband's healthy sister, was likely inherited from the father, while the missense variant was inherited from the mother.

Recessive mutations in *HSD17B4* cause D-bifunctional protein deficiency (OMIM #261515), a severe disorder of peroxisomal fatty acid beta-oxidation that is generally fatal within the first 2 years of life,¹⁰ and have recently been identified in two sisters diagnosed with Perrault

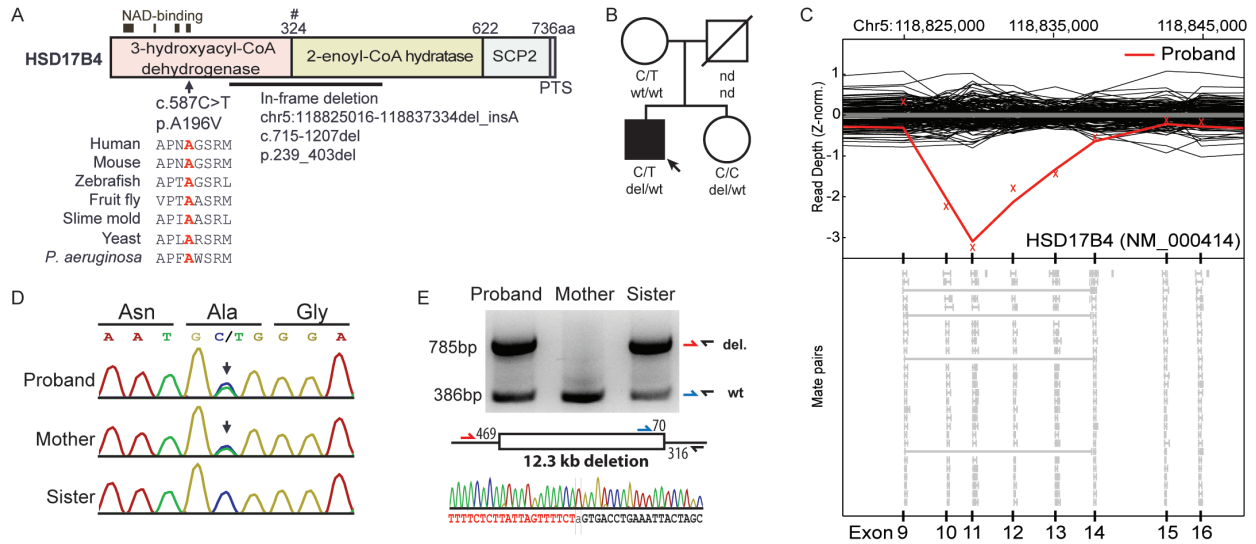


Figure 4.2. Compound heterozygous variants in *HSD17B4*.

A) Schematic of *HSD17B4* (NP_000405) with multiple sequence alignment. SCP2, sterol carrier protein 2; #, cleavage site. B) Pedigree denoting genotypes. C) Normalized depth of coverage at each exon (“x”); lines represent smoothed data from proband (red) and controls (black). Below, reads showing deletion-spanning mate pairs. D) Sanger trace of c.587C>T. E) DNA gel indicating genotypes at deletion site, schematic of primers, and Sanger trace across deletion in proband.

Abbreviations: wt, wildtype; nd, no data; del, deletion, norm, normalized.

syndrome (OMIM #233400), who exhibited severe sensorineural deafness, ovarian dysgenesis, peripheral neuropathy and ataxia.¹¹ Due to the phenotypic overlap with previous cases, the predicted severity of the *HSD17B4* mutations, and evidence of altered peroxisomal very-long-chain fatty acid metabolism, we reached a diagnosis of Perrault syndrome in our patient.

Discussion

HSD17B4, also known as D-bifunctional protein (DBP), is a peroxisomal enzyme that catalyzes multiple steps of beta-oxidation of very long chain fatty acids (VLCFA). Although named for its homology to a steroid-converting enzyme, the protein's role in steroid hormone metabolism is unclear. The protein consists of three domains: an N-terminal dehydrogenase domain, a hydratase domain, and a sterol carrier protein (SCP) domain. The three amino acids at the C-terminus ("AKL") constitute the peroxisomal targeting signal (PTS). After peroxisomal import, the 79-kDa full-length protein is proteolytically cleaved to yield a 35-kDa dehydrogenase subunit and a 45-kDa hydratase subunit containing the hydratase and SCP domains.

Recessive mutations in HSD17B4 are known to cause DBP-deficiency, an early-onset neurological disorder characterized by neonatal hypotonia, seizures, visual impairment, and psychomotor retardation.¹⁰ Patients with the disorder generally die in infancy, although a small number survive into their teens. Like our patient, individuals with DBP deficiency often show an elevated ratio of pristanic to phytanic acid in plasma, reflective of dysfunctional beta-oxidation in the peroxisome.¹²

DBP-deficiency has been classified into subtypes based on the affected enzymatic activities of HSD17B4. Type I patients have deficiencies of both the hydratase and dehydrogenase subunits whereas types II and III have isolated hydratase or dehydrogenase deficiency, respectively. Recently, a Type IV category was proposed to describe mildly affected patients with compound heterozygous

mutations affecting both subunits.¹³ Our case may be classified as type IV due to the mutations affecting both subunits and the phenotypic similarity between our patient and two teenage brothers categorized as type IV, including the presence of cerebellar ataxia, polyneuropathy, pes cavus, and hearing loss.¹³

In addition to causing DBP-deficiency, recessive mutations in *HSD17B4* have been identified in two sisters diagnosed with Perrault syndrome (OMIM #233400). Perrault syndrome was first defined in 1951 as ovarian dysgenesis and sensorineural hearing loss in females.¹⁴ Brothers of Perrault syndrome females have been reported with hearing loss but typically are fertile.¹⁵ While autosomal recessive inheritance has been observed, the syndrome is genetically heterogeneous.¹⁶ Recessive mutations in three additional genes have been implicated in Perrault syndrome: *HARS2*, *LARS2*, and *CLPP*, that respectively encode the mitochondrial histidyl-tRNA synthetase, leucyl-tRNA synthetase, and a mitochondrial ATP-dependent protease.¹⁷⁻¹⁹

The phenotype and genotype of our patient resemble those of the two Perrault syndrome sisters.¹¹ Both families presented with infertility, cerebellar ataxia, demyelinating polyneuropathy, pes cavus, and hearing loss. Both had compound heterozygous *HSD17B4* variants consisting of a missense mutation predicted to affect NAD-binding in the dehydrogenase domain and a predicted loss-of-function mutation in the hydratase subunit. Our patient had azoospermia rather than complete gonadal dysgenesis, and it is noteworthy that *HSD17B4* knockout mice, which accumulate VLCFA but show no neurological abnormalities, exhibit male-specific sterility with testicular lipid accumulation.²⁰

While *HSD17B4* is listed in the MitoCarta inventory due to its strong co-expression and homology with *bona fide* mitochondrial proteins and its proteomic identification in mitochondrial isolates,²¹ prior experimental studies have physically localized the protein to the peroxisome.²² It is known that abnormal peroxisomal fatty acid catabolism can lead to secondary mitochondrial

dysfunction.²³ This case, together with prior examples of disease caused by dual-localized mitochondrial/peroxisomal proteins (e.g., AMACR), raises the hypothesis that defects within the peroxisome can lead to secondary defects within mitochondria.

The discovery of a multi-exon deletion compounded with a rare missense mutation suggests that CNV detection may be a crucial addition to standard exome analysis. Our patient represents the first male patient with infertility and ataxia due to recessive *HSD17B4* mutations, and the second unrelated case with hearing loss and infertility due to recessive *HSD17B4* mutations.^{11, 16} Identification of additional patients with Perrault syndrome due to *HSD17B4* deficiency should further expand the phenotypic spectrum of the disorder.

Acknowledgements

We thank Scott Vafai, Declan McGuone, and Matthew Frosch for discussion. This work was supported by the NSF Graduate Research Fellowship (DSL), DOD National Defense Science and Engineering Graduate Fellowship (SGH), grants RC2HG005556 and R01GM97136 from the NIGMS/NIH (VKM), and support from the Birmingham and MINDlink Foundations (JDS).

References

1. Zeviani M, Simonati A, Bindoff LA. Ataxia in mitochondrial disorders. Handbook of clinical neurology / edited by PJ Vinken and GW Bruyn. 2012;103:359-72.
2. Chinnery PF. Mitochondrial Disorders Overview. In: Pagon RA, Bird TD, Dolan CR, Stephens K, Adam MP, editors. GeneReviews. Seattle (WA)1993.
3. Lieber DS, Calvo SE, Shanahan K, et al. Targeted Exome Sequencing of Suspected Mitochondrial Disorders. Neurology. 2013;In press.
4. Bernier FP, Boneh A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. Neurology. 2002 Nov 12;59(9):1406-11.
5. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]. 2002 Aug;Chapter 2:Unit 2.3.
6. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010 Apr;7(4):248-9.

7. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>) [January 2012].
8. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. Oct 28;467(7319):1061-73.
9. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome research*. 2012 Aug;22(8):1525-32.
10. Ferdinandusse S, Denis S, Mooyer PA, et al. Clinical and biochemical spectrum of D-bifunctional protein deficiency. *Annals of neurology*. 2006 Jan;59(1):92-104.
11. Pierce SB, Walsh T, Chisholm KM, et al. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *American journal of human genetics*. 2010 Aug 13;87(2):282-8.
12. Paton BC, Sharp PC, Crane DI, Poulos A. Oxidation of pristanic acid in fibroblasts and its application to the diagnosis of peroxisomal beta-oxidation defects. *The Journal of clinical investigation*. 1996 Feb 1;97(3):681-8.
13. McMillan HJ, Worthylake T, Schwartzentruber J, et al. Specific combination of compound heterozygous mutations in 17beta-hydroxysteroid dehydrogenase type 4 (HSD17B4) defines a new subtype of D-bifunctional protein deficiency. *Orphanet journal of rare diseases*. 2012 Nov 22;7(1):90.
14. Perrault M, Klotz B, Housset E. [Two cases of Turner syndrome with deaf-mutism in two sisters]. *Bulletins et memoires de la Societe medicale des hopitaux de Paris*. 1951 Jan 26-Feb 2;67(3-4):79-84.
15. Christakos AC, Simpson JL, Younger JB, Christian CD. Gonadal dysgenesis as an autosomal recessive condition. *American journal of obstetrics and gynecology*. 1969 Aug 1;104(7):1027-30.
16. Jenkinson EM, Clayton-Smith J, Mehta S, et al. Perrault syndrome: further evidence for genetic heterogeneity. *Journal of neurology*. 2012 May;259(5):974-6.
17. Pierce SB, Chisholm KM, Lynch ED, et al. Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 2011 Apr 19;108(16):6543-8.
18. Pierce SB, Gersak K, Michaelson-Cohen R, et al. Mutations in LARS2, Encoding Mitochondrial Leucyl-tRNA Synthetase, Lead to Premature Ovarian Failure and Hearing Loss in Perrault Syndrome. *American journal of human genetics*. 2013 Mar 26.
19. Jenkinson EM, Rehman AU, Walsh T, et al. Perrault Syndrome Is Caused by Recessive Mutations in CLPP, Encoding a Mitochondrial ATP-Dependent Chambered Protease. *American journal of human genetics*. 2013 Mar 26.
20. Huyghe S, Schmalbruch H, De Gendt K, et al. Peroxisomal multifunctional protein 2 is essential for lipid homeostasis in Sertoli cells and male fertility in mice. *Endocrinology*. 2006 May;147(5):2228-36.
21. Pagliarini DJ, Calvo SE, Chang B, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell*. 2008 Jul 11;134(1):112-23.
22. Adamski J, Normand T, Leenders F, et al. Molecular cloning of a novel widely expressed human 80 kDa 17 beta-hydroxysteroid dehydrogenase IV. *The Biochemical journal*. 1995 Oct 15;311 (Pt 2):437-43.
23. Busanello EN, Amaral AU, Tonin AM, et al. Experimental evidence that pristanic acid disrupts mitochondrial homeostasis in brain of young rats. *Journal of neuroscience research*. 2012 Mar;90(3):597-605.

CHAPTER V:

Operon analysis implicates C6orf57 in Complex II biology

Contributors: Daniel S. Lieber, Vishal M. Gohil, Nicole P. Morales, Joshua M. Baughman, Roland Nilsson, Sarah E. Calvo, Patricia L. Dahia, Vamsi K. Mootha

Abstract

Complex II, also known as succinate dehydrogenase (SDH), is central to mitochondrial energy metabolism, with roles in both the mitochondrial respiratory chain and the Krebs cycle. In humans, mutations in Complex II subunits and assembly factors can cause hereditary tumor syndromes, such as paraganglioma and pheochromocytoma, as well as disorders of energy metabolism, such as Leigh's disease and infantile leukoencephalopathy. We have developed a computational approach utilizing the information inherent in bacterial operons to predict novel components of mitochondrial pathways and protein complexes. This method has implicated the uncharacterized mitochondrial protein C6orf57 and its yeast homologue, FMP21, in a role in Complex II activity. Studies of isolated biochemical activity in *FMP21* deletion strains reveal compromised Complex II and Complex II+III activity. Sequencing of germline DNA from 72 patients with pheochromocytoma or paraganglioma of unknown genetic etiology revealed no rare coding variants in C6orf57. Future work hopes to elucidate the specific role of the protein in Complex II biology and to confirm the finding in mammalian systems. This study suggests that C6orf57 may be a novel assembly factor for Complex II and is a candidate gene for future studies of paraganglioma and pheochromocytoma.

Introduction

Succinate dehydrogenase (SDH), also known as Complex II of the mitochondrial respiratory chain, is the only enzyme complex involved in both oxidative phosphorylation and the Krebs cycle, making it a central component of mitochondrial energy metabolism. SDH catalyzes the oxidation of succinate to fumarate and couples the reaction to the reduction of ubiquinone. There are four core subunits of Complex II (SDHA, SDHB, SDHC, SDHD), which are encoded in the nuclear genome in mammals and in the yeast *Saccharomyces cerevisiae*. SDHA, a flavoprotein, and SDHB, and iron-sulfur protein, form the catalytic subunits while SDHC and SDHD represent membrane anchor subunits. In yeast, the homologues of SDHA-D are SDH1-4, respectively.

In mammals, two complex II assembly factors have been identified: SDHAF1, also known as SDH6/YDR379C-A in yeast, which has been implicated in infantile leukoencephalopathy ¹, and SDHAF2, also known as SDH5/EMI5 in yeast, ² that has been linked to paraganglioma. SDH5 is required for flavination of SDHA, ³ while SDH6 is thought to affect the insertion or retention of iron-sulfur clusters within Complex II. ¹ The yeast protein TCM62, which has homology to the mitochondrial chaperonin HSP60 in humans, has also been identified as a complex II assembly factor ⁴, although it is unclear to what extent the protein's role is specific to Complex II assembly.

Succinate dehydrogenase deficiency can result in a number of human diseases. Mutations in *SDHA* can cause Leigh syndrome ⁵ and mutations in *SDHAF1* can cause infantile leukoencephalopathy ¹. In addition, germline mutations in *SDHB*, *SDHC*, *SDHD*, and *SDHAF2* can lead to rare neuroendocrine tumor conditions, accounting for 20-40% of patients with paraganglioma or pheochromocytoma. ⁶ While our understanding of the genetic basis for these tumors has improved dramatically over the last decade, no likely deleterious germline mutations in known susceptibility genes are identified in 30-60% of cases of paraganglioma/pheochromocytoma, suggesting that causal genetic loci remain to be discovered. ⁶

The evolutionary relationship between mitochondria and bacteria enables computational inference of mitochondrial gene function through analysis of bacterial gene neighborhoods. It is believed that mitochondria are derived from a bacterial ancestor whose close symbiosis with its eukaryotic host eventually led to the complete integration of the two organisms⁷. Over the course of evolution, many of the genes encoded by the mitochondrial genome were transferred to the eukaryotic host genome such that today only 13 of the 1,500 estimated human mitochondrial proteins are encoded by the mitochondrial DNA. The remainder of the mitochondrial proteome is encoded in the nuclear genome and 60% of mammalian mitochondrial proteins have close bacterial homologues, enabling researchers to study protein function using strategies developed for bacteria⁸.

Bacterial genomes are organized into operons, which are functional units of genomic DNA containing a cluster of genes under the control of a single regulatory signal or promoter. Often, operons contain multiple genes involved in a particular metabolic pathway or protein complex. For example, all thirteen subunits of Complex I, also known as NADH dehydrogenase or NADH:ubiquinone oxidoreductase, are encoded by the *nuo* operon in *E. coli*.⁹ The phenomenon of spatial clustering of functionally related genes has enabled the inference of bacterial gene function, an approach referred to as operon analysis or inference by genomic context.^{10,11} Currently, there are over 700 sequenced bacterial genomes freely available through NCBI for which the genes have been assigned to operons¹². The wealth of operon data provides a resourceful way to discover relationships among bacterial proteins, and by extension, amongst eukaryotic proteins with close bacterial homologues.

To identify novel components of Complex II, and therefore potential paraganglioma candidate genes, we have developed a computational approach utilizing the extensive homology between mitochondrial and bacterial proteomes as well as the information inherent in bacterial operons.

Methods

Operon analysis

Sequences for 675 prokaryotic species (Bacteria and Archaea) were downloaded from KEGG in May 2009. Protein sequences for 1098 mouse mitochondrial proteins were obtained from <http://www.broadinstitute.org/pubs/MitoCarta/>.⁸ BlastP with 1e-3 e-value cutoff was used to find homology between MitoCarta and prokaryotic proteomes.¹³ Prokaryotic operon predictions were downloaded from DOOR¹². For each pair of MitoCarta proteins, we counted the number of genera (based on species name) that (a) contained homologues of both genes predicted to be in the same operon or (b) contained homologues of both genes in the genome. Each pair of genes received an Op/Cop score defined as (a)/(b). For each pair of genes, we then calculated the Pearson correlation over the Op/Cop scores across all genes. This matrix was then hierarchically clustered with Matlab (clusterdata) using default parameters.

Yeast Co-expression

FMP21 was selected as input to SPELL version 2.03 (<http://spell.princeton.edu/>), which was used with default settings and a maximum of 100 results.¹⁴

Phylogenetic profiling

121 eukaryotic proteomes, 68 archaeal, and 942 bacterial proteomes were downloaded from NCBI. PFAM domains were predicted by PfamScan¹⁵ using default parameters. For each species, the number of proteins with a domain was counted. The Pearson correlation of these scores across all bacterial species was determined between each PFAM protein domain and DUF1674 (PF07896).

Respiratory phenotype in yeast

Data for the respiratory phenotype of the yeast deletion collection was downloaded from (http://www-deletion.stanford.edu/YDPM/Download_Data/Yeast_Deletion_Project/Regression_Tc2_hom.txt).¹⁶

For each deletion strain, the Pearson correlation across all growth conditions was calculated relative to the *fmp21* (YBR269C) deletion.

Biochemical assays in yeast

Isolated Complex II (DCPIP reduction) and II+III (cytochrome c reduction) assays were performed as previously described.¹⁷ Citrate synthase was measured using the citrate synthase assay kit (Sigma CS0720). Protein quantitation was performed with the BCA assay kit (Thermo Scientific Pierce 23225).

Strain creation and media conditions

Wildtype strains and haploid yeast knockout strains of *fmp21* (YBR269C), *sdh2*(YLL041C) and *emi5/sdh5* (YOL071W) were obtained from yeast knockout collection (courtesy of C. Kiely and F. Winston). Wildtype background strain is BY4741 (*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*). Gene deletion was confirmed through colony PCR using primers “TTTCAAAGATATATTTTCAGCGAGG” and “AAAAGCACTAGAAAGACTGTCTCCA”.¹⁸ Strains were grown in YPD media unless otherwise specified. Mitochondrial isolation was performed as previously described.¹⁹

FMP21 was amplified by PCR from WT yeast using primers “GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCAAAAATGTTGTGCGCCATCAAAAAGC” and “GGGGACCACTTTGTACAAGAAAGCTGGGTGCGAAATCCGTTACTCTTCCGTTAAAC” and cloned into pYES-DEST52 plasmid (Life Technologies) under a GAL1 promoter. Plasmid containing FMP21 or empty plasmid was transformed using one-step transformation of yeast.²⁰ Yeast colonies transformed with the plasmid were selected and tested in SC-URA media containing 20% galactose.

Germline sequencing of tumor syndrome patients

Germline DNA from 72 patients with pheochromocytoma or paraganglioma and clinical or transcriptional features consistent with pseudohypoxic-type tumors was tested for mutations in all three exons of *C6orf57* including 3' and 5' UTR. Informed consent was obtained according to a University of Texas Health Science Center at San Antonio (UTHSCSA) Institutional Review Board (IRB)-approved protocol. All patients had previously tested negative for mutations in *SDHB*, *SDHC*, *SDHD*, *VHL*, *RET*, *NF1* and *TMEM127*. Most cases were also negative for mutations in *SDHAF2/SDH5*, but they have not been checked for *SDHA* mutations. Sequencing was performed at Beckman Genomics and analyzed using the Mutation Surveyor Software (Softgenetics).

The following PCR primers were used to amplify *C6orf57* exons and UTRs: 1F, CCAGGGATGCCTCTAACTCC; 1R, GGGAGGCGCTTCTCTCTG; 2F, TTTCACCTCTGCGTCCTTTTG; 2R, TCTTTCTTTCACTCATTCTCTCA; 3F, CTGGGCAACAGAGGGAGA; 3R, GCAGCATATTTCTTCCAAAGC; 3UTR-F, TTCTTTCTTTATATCCTTTATGTCGTG; 3UTR-R, CCTCCTGGGTTTAAGCCATT.

Results

Gene neighborhoods for bacterial homologues of C6orf57

We used protein homology along with predictions of bacterial operons across 675 prokaryotic genomes to determine the pairs of mitochondrial proteins whose bacterial homologues consistently share gene neighborhoods across multiple unrelated species. Clustering this data revealed modules of mitochondrial protein homologues that tended to co-locate within bacterial operons. Manual inspection of these clusters revealed coherent clusters of proteins involved in diverse complexes and pathways including respiratory chain Complexes I, II, III/IV, and V, mitochondrial ribosome, Fe-S cluster biogenesis, fatty acid metabolism, ABC transporters, Krebs cycle, and glycine cleavage.

We manually inspected each cluster in order to determine whether there were any novel

candidate genes not known to associate with other members of the cluster. Only the Complex II cluster contained a gene of unknown function with operon evidence linking it to the majority of genes within the cluster. In addition to the known SDH subunits with strong bacterial homology (SDHA, SDHB, and SDHC), this cluster contained C6orf57, a completely uncharacterized protein.

A closer look at the bacterial homologues of C6orf57 revealed that the protein is found in a conserved operon in a number of bacterial species along with all four subunits of Complex II (Figure 5.1). Among all bacterial proteins, the closest bacterial homologue of human C6orf57 is xccb100_2045 of *Xanthomonas campestris*, a gammaproteobacteria. This gene lies in an operon conserved across at least twelve species of gammaproteobacteria and one species of betaproteobacteria. Interestingly, this operon and nearly all of the homologous operons also contain the recently identified Complex II assembly factor SDH5 (also known as SDHAF2 in humans, and SDHE in bacteria), which has been shown to play a role in the flavination of SDHA.^{2,3}

While the closest bacterial homologue to C6orf57 lies in an operon with Complex II subunits and an assembly factor, there are other bacterial homologues whose gene neighborhoods may shed light into the role and function of C6orf57. In multiple species of alphaproteobacteria including *Rickettsia prowazekii*, the species considered to be the closest relative of the mitochondrial ancestor, a DUF1674 domain protein (RP167 in *R. prowazekii*) lies adjacent to a protein containing a polyketide cyclase domain (PF03364), which is involved in polyketide synthesis or the binding & transport of lipids. Interestingly, these polyketide cyclase domain-containing proteins are homologous to human proteins COQ10A and COQ10B and to yeast COQ10. In yeast, COQ10 is a ubiquinone-binding protein thought to function in the delivery of Q₆ to its proper location in the mitochondrial inner membrane.^{9 21} The observation that C6orf57 homologues lie in operons with homologues of Complex II subunits and ubiquinone transport factors lie suggests that perhaps

Bacterial genomes	<i>C6orf57</i>	<i>SDHC</i>	<i>SDHD</i>	<i>SDHA</i>	<i>SDHB</i>	<i>SDHAF2</i>
<i>Xanthomonas campestris</i> pv. <i>campestris</i> B100	xccb100_2045	xccb100_2046	xccb100_2047	xccb100_2048	xccb100_2050	xccb100_2051
<i>Xanthomonas campestris</i> pv. <i>campestris</i> 8004	XC_1979	XC_1981	XC_1982	XC_1983	XC_1985	XC_1986
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913	XCC2134	XCC2132	XCC2131	XCC2130	XCC2129	XCC2128
<i>Pseudoxanthomonas suwonensis</i>	Psesu_1295	Psesu_1296	Psesu_1297	Psesu_1298	Psesu_1299	-
<i>Xanthomonas axonopodis</i> pv. <i>citrumelo</i> F1	XACM_2180	XACM_2182	XACM_2183	XACM_2184	XACM_2186	XACM_2187
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A	PXO_00605	PXO_00604	PXO_00603	PXO_00602	PXO_00600	PXO_00599
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> 306	XAC2073	XAC2075	XAC2076	XAC2077	XAC2078	XAC2079
<i>Methylomonas methanica</i>	Metme_2138	Metme_2139	Metme_2140	Metme_2141	Metme_2142	Metme_2143
<i>Xanthomonas albilineans</i>	XALc_1827	XALc_1826	XALc_1825	XALc_1824	XALc_1823	XALc_1822
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i>	XCV2240	XCV2241	XCV2242	XCV2243	XCV2245	XCV2246
<i>Pseudoxanthomonas spadix</i>	DSC_09595	DSC_09590	DSC_09585	DSC_09580	DSC_09575	DSC_09570
<i>Alkalimnicola ehrlichei</i>	Mlg_1331	Mlg_1332	Mlg_1333	Mlg_1334	Mlg_1335	Mlg_1336
<i>Methylomicrobium alcaliphilum</i>	MEALZ_2683	MEALZ_2682	MEALZ_2681	MEALZ_2680	MEALZ_2679	MEALZ_2678
<i>Cycloclasticus</i> sp. P1	Q91_1061	Q91_1062	Q91_1063	Q91_1064	Q91_1065	Q91_1066
<i>Xylella fastidiosa</i> subsp. <i>fastidiosa</i> GB514	XFLM_07120	XFLM_07125	XFLM_07130	XFLM_07135	XFLM_07140	XFLM_07145
<i>Xylella fastidiosa</i> M23	XfasM23_0341	XfasM23_0342	XfasM23_0343	XfasM23_0344	XfasM23_0345	XfasM23_0346
<i>Xylella fastidiosa</i> Temecula1	PD0349	PD0350	PD0351	PD0352	PD0353	PD0354
<i>Stenotrophomonas maltophilia</i> R551-3	Smal_1534	Smal_1535	Smal_1536	Smal_1537	Smal_1538	Smal_1539
<i>Xylella fastidiosa</i> M12	Xfasm12_0377	Xfasm12_0378	Xfasm12_0379	Xfasm12_0380	Xfasm12_0381	Xfasm12_0382
<i>Xylella fastidiosa</i> 9a5c	XF1069	XF1070	XF1071	XF1072	XF1073	XF1074
<i>Pusillimonas</i> sp. T7-7	PT7_2480	PT7_2481	PT7_2482	PT7_2483	PT7_2484	-

Figure 5.1. Conserved operon encoding homologues of *C6orf57*, *SDHA-D*, and *SDHAF2*.

Bacterial gene identifiers are listed below their homologues in human. Arrows represent transcriptional directionality of bacterial operons. Most of these operons also encode a homologue of the recently identified Complex II assembly factor *SDHAF2* (also known as *SDH5*). All bacterial species represent gamma-proteobacteria except for *Pusillimonas* sp. T7-7, which is beta-proteobacterial. Species are sorted in order of decreasing homology between human *C6orf57* and the bacterial DUF1674 homologue. Dashes represent the absence of a homologue in the operon.

DUF1674 proteins are involved in the interplay between Complex II and ubiquinone, which is needed to transfer electrons from Complex II to Complex III.

Localization of C6orf57

C6orf57 is amongst the strongest predictions of being mitochondrial localized according to MitoCarta⁸. The protein is predicted by TargetP and MitoPred to have a mitochondrial localization signal, has a mitochondrial homologue in yeast (*FMP21*), has a homologue in *Rickettsia prowazakei* (RP167), is strongly co-expressed with other mitochondrial proteins (11/50 nearest neighbors in co-expression space are known mitochondrial proteins), and is found in mitochondrial fractions by MS/MS in three mouse tissues: adipose, small intestine, & testis. The Human Protein Atlas (<http://www.proteinatlas.org/>) has experimentally shown mitochondrial localization of the protein across multiple cell types.¹⁷ Altogether, there is strong computational and experimental evidence that the protein is localized to the mitochondrion.

Structure and conservation of C6orf57 protein

In humans, the C6orf57 protein is 108 amino acids and consists of an N-terminal mitochondrial targeting sequence (amino acids 1-22) and a C-terminal DUF1674 domain at aa 40-108 (Figure 5.2A). DUF1674 (PF07896) is a domain of unknown function approximately 60 amino acids in length that is found in many eukaryotic and bacterial organisms. The domain is found at the C-terminus across nearly all species with the domain. There are a number of conserved residues near the C-terminus (Figure 5.2B), including a stretch of conserved residues forming the motif “GPEPTRY” and the terminal two amino acids, “DF”. A range of 3D structures was determined experimentally for a bacterial homologue of C6orf57, Rhr2 in the alphaproteobacteria *Rhodobacter sphaeroides* (PDB 2K5K). While some models show the protein in a toric conformation, a number of

other conformations suggest a high degree of structural flexibility (Figure 5.2C). While the global conformation varies widely between the different 3D models of the protein, the local conformations consistently suggest the presence of a kinked or helical C-terminal tail.

Presence of DUF1674 in Evolution

Evolutionarily, the DUF1674 (PF07896) domain is present in a number of alpha-, beta-, and gamma-proteobacteria, as well as in some protists and most eukaryotes. The domain is almost always C-terminal (426/427 species in PFAM database) and the proteins are typically small (typically <100 amino acids in bacteria, <200 in eukaryotes). The domain is present in >87% of 121 eukaryotic species sequenced to date, with notable exceptions including some microsporidian and diatom species. The protein is present in at least two species of protists (*Phytophthora ramorum*, *Phytophthora sojae*), 88 alpha-proteobacterial genomes (75% of 117), 1 beta-proteobacterial species (*Pusillimonas sp. T7-7*), and 12 gamma-proteobacterial genomes (5% of 238).

Phylogenetic profiling across bacterial genomes reveals that the presence of DUF1674 in a genome is correlated with the presence, in that genome, of a number of other PFAM domains, including HIG_1_N (PF04588), DUF1982 (PF09326), NDUFA12 (PF05071), ATP12 (PF07542), ETC_C1_NDUFA4 (PF04800), NADH5_C (PF06455), and CybS (PF05328). Many of these represent core structural domains of known respiratory complex or super-complex assembly factors, including HIG2A and ATPAF2.^{18,22} Interestingly, CybS, which is the primary domain in human SDHD, is not found in most bacterial species, which instead have an SDHD subunit containing the Sdh_cyt domain (PF01127). Given the other evidence of association between DUF1674 and Complex II subunits, this observation raises the hypothesis that DUF1674 may be functionally related to CybS-containing forms of SDHD.

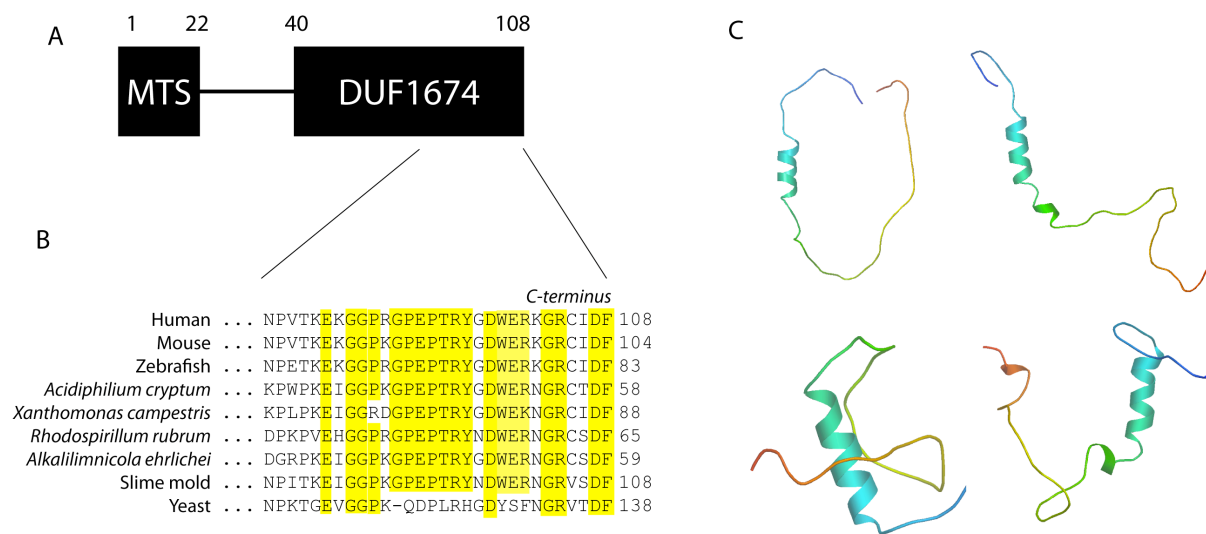


Figure 5.2. Structure of DUF1674-containing proteins.

A) Schematic of human C6orf57 protein showing N-terminal mitochondrial targeting sequence (MTS) and C-terminal DUF1674 domain. B) Multiple sequence alignment of C6orf57 homologues in select species, highlighting residues of high homology. Uniprot IDs: Q5VUM1, Q8BTE0, Q6ZM06, A5G2G6, Q8P8V1, Q2RNC3, Q0A907, Q54Y25, P38345. C) Four models of 3D structure of Rhr2 in the alphaproteobacteria *Rhodobacter sphaeroides* (PDB 2K5K). N-terminus to C-terminus is colored blue to red, with DUF1674 at the C-terminus.

FMP21: respiratory phenotype and co-expression in yeast

Many species of fungi possess homologues of C6orf57, including the budding yeast *S. cerevisiae* in which the homologue is FMP21 (YBR269C). This protein has been localized to the mitochondrion through proteomic analyses²⁰ and has 41% identity to the human protein across the DUF1674 domain. Other than being identified in a screen as one of 84 genes required for protection from hyperoxia toxicity,¹² the function and role of FMP21 is unknown.

Multiple large-scale studies have shown that deletion of *FMP21* results in a respiratory phenotype.^{16,23} Interestingly, a reanalysis of the Steinmetz et al. data shows that the respiratory phenotype of this gene is remarkably similar to that of other Complex II subunits, in that these deletion strains showed a severe growth deficiency when grown in glycerol or lactate as the carbon source media but a milder phenotype when grown in ethanol. Analysis of the respiratory phenotype across nine media conditions reveals that the fifteen genes whose knockout respiratory profile is most correlated with *FMP21*'s include all five Complex II-related proteins for which there is data (*SDH5*, *SDH4*, *TCM62*, *SDH1* and *SDH2*; no data for *SDH3* and *YDR379C-A*). While this list of genes also includes a number of other proteins involved in the Krebs cycle (*FUM1*, *CIT1*), there is a strong enrichment for Complex II-related genes ($p < 1e-15$).

There is also co-expression evidence linking FMP21 to components and assembly factors of Complex II. According to SPELL (Hibbs et al., 2007), five out of the seven known Complex II subunits/assembly factors are found within the top 150 genes co-expressed with *FMP21*: *SDH4* (#7), *SDH2* (#34), *EMI5/SDH5* (#41), *SDH3* (#99), and *SDH1* (#146). Co-expression of *FMP21* (YBR269C) and *SDH4* (YDR178W) has also been observed previously through automated clustering algorithms.²⁴

Complex II activity of FMP21 deletion

The operon, co-expression, and respiratory phenotype analyses support the hypothesis that *C6orf57* and its yeast homologue, *FMP21*, play a role in Complex II biology. To test this hypothesis experimentally, we obtained *FMP21* knockouts and measured Complex II activity, comparing the activity to that of wildtype strains or deletion strains for Complex II subunits and assembly factors. Deletion of *FMP21* in a haploid strain resulted in approximately two-fold reduction in Complex II and Complex II+III activity (Figure 5.3A). While the two-fold reduction is statistically significant, knockouts of *sdh2* and *emi5* showed a ten-fold decrease in residual activity in both Complex II+III and Complex II activity compared to wildtype. We expressed a plasmid with *FMP21* under a *GAL1* promoter in the *fmp21* deletion strain and rescued Complex II+III activity to 90% of wildtype (Figure 5.3B), demonstrating that the reduction in Complex II+III activity is due to the absence of *FMP21*. Complex II activity has not yet been tested in this rescue system. Our results demonstrate that deletion of *FMP21* significantly impairs Complex II activity, but to a lesser extent than deletion of Complex II subunits or assembly factors.

Sequencing patients with paraganglioma and pheochromocytoma

Given the computational and experimental evidence of *C6orf57*'s involvement in Complex II activity, and the role of Complex II subunits and assembly factors in rare neuroendocrine tumors, we sequenced the exons of *C6orf57* in germline DNA from 72 patients with paraganglioma or pheochromocytoma of unknown etiology. While several known polymorphisms were detected in *C6orf57*, including rs1048886 and rs13654, no rare coding mutations were detected in the patients.

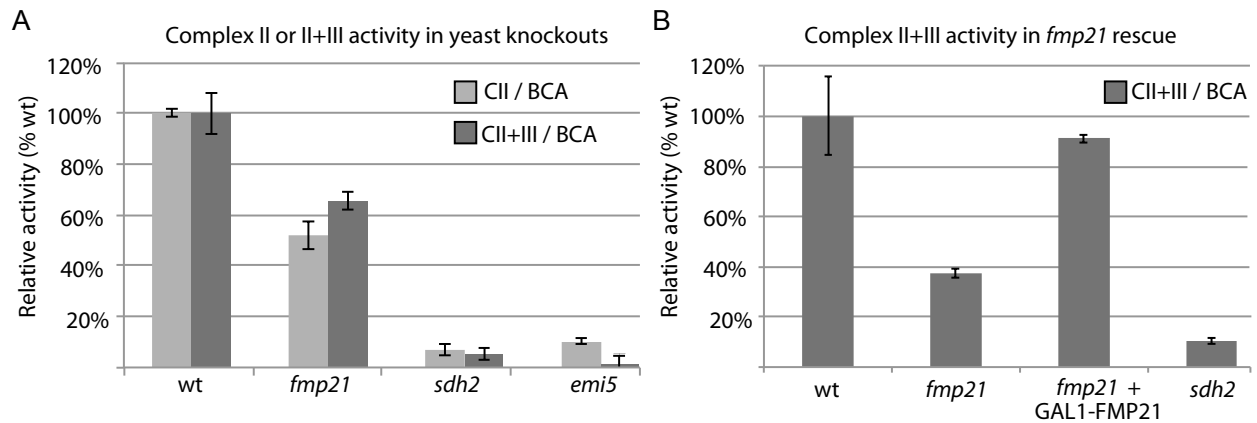


Figure 5.3. Complex II activity in yeast knockouts.

Complex II or Complex II+III activity was measured in whole cell lysate and normalized to protein content (BCA assay). This was plotted as percent of wildtype (wt) activity. Error bars represent standard error (two technical replicates). A) Complex II or II+III activity assay in yeast knockout strains B) Complex II+III assay in yeast knockouts or knockout + rescue grown in SC-URA media +10% galactose.

Discussion

Through a computational screen utilizing bacterial operons, we identified a potential functional relationship between the uncharacterized protein C6orf57 and Complex II of the mitochondrial respiratory chain, also known as succinate dehydrogenase. The operon of the closest bacterial homologue of C6orf57 contains all four subunits of Complex II and the homologue of SDHAF2, a recently identified Complex II assembly factor. This operon is conserved across 13 proteobacterial species.

Experimental and computational evidence involving *FMP21*, the *C6orf57* homologue in yeast, support the hypothesis that C6orf57 is involved in Complex II function. *FMP21* is highly co-expressed with four out of five SDH subunits or assembly factors and the deletion strain shows a two-fold reduction in Complex II+III activity, which can be rescued by expression of the gene from a plasmid, and a two-fold reduction in Complex II activity. The experimental evidence suggests that *FMP21* is required for maximal Complex II and II+III activity but does not result in as severe a Complex II phenotype as would *SDH2* or *SDH5* deletion strains, since such strains reduced Complex II activity to 10% residual activity. Such a result is reminiscent of Chen et al., in which Rcf1 is described as a protein not essential for basal mitochondrial function but essential for optimally efficient ETC function and respiration.²² In fact, phylogenetic profiling shows that the HIG1 domain of Rcf1 and the DUF1674 domain of FMP21 are highly correlated across bacteria ($r=0.81$).

Whereas SDHAF2 was found to catalyze the flavination of SDHA, there is evidence suggesting that C6orf57's role may be related to the interface between Complex II and ubiquinone. First, *FMP21* is closely co-expressed with *SDH4*, both relative to all proteins as well as to among Complex II subunits and assembly factors. Second, DUF1674 has a correlated phylogenetic profile across bacteria with the CybS domain of SDHD. Third, the DUF1674 protein of *Rickettsia prowazekii*

lies adjacent to a homologue of COQ10, which is involved in ubiquinone binding and transport. These observations raise the hypothesis that *C6orf57* impacts the mitochondrial inner membrane, possibly by affecting the interaction between *SDHD* and ubiquinone.

Germline mutations in *SDHB*, *SDHC*, *SDHD*, and *SDHAF2* can cause increased risk of paraganglioma and pheochromocytoma (PGL/PCC) and we therefore hypothesized that *C6orf57* may represent a novel candidate. However, sequencing the exons of *C6orf57* in 72 PGL samples did not uncover any rare, non-synonymous mutations. Our result suggests that germline *C6orf57* mutations are not a common cause of PGL/PCC. It should be noted, however, that only a single pathogenic variant in *SDHAF2* has been identified to date, and this variant has only been reported in two PGL families, despite sequencing of over 300 individuals.^{2,25} This implies that mutations in PGL/PCC genes can be quite rare, requiring germline DNA from hundreds of patients before uncovering mutations in a particular susceptibility gene. In addition to sequencing germline DNA from more individuals, somatic DNA should also be analyzed for mutations in *C6orf57*.

If *C6orf57* does not play a role in tumorigenesis, it is possible that it plays a role in other disease contexts. Recessive mutations in *SDHA* and *SDHAF1* can cause mitochondrial respiratory chain disorders including Leigh syndrome and infantile leukodystrophy, and *C6orf57* is similarly a candidate for such severe disorders. In mice, the homologue of *C6orf57* (*1110058L19Rik*) has been implicated in innate immunity, as RNAi knockdown of the gene blocks the production of IL-6 in J774A.1 macrophages.²⁶ In humans, the SNP rs1048886 (*C6orf57*: p.Q46R) was found to be associated with risk of type 2 diabetes in South Asians at genome-wide significance.^{27 28} However, this result has not been replicated and furthermore, rs1048886 is in linkage disequilibrium with variants in neighboring gene *FAM135A*, making it unclear which of the two genes is implicated in the association.

In conclusion, we have identified a putative role for C6orf57 in Complex II of the mitochondrial respiratory chain. In yeast, deletion of its homologue, *FMP21*, results in a two-fold decrease in Complex II activity. However, it is possible that FMP21 and C6orf57 may affect the function of multiple respiratory chain complexes and this should be investigated further. Future work should seek to further characterize the *FMP21* deletion strain and to determine whether the protein can bind lipids *in vitro*. In addition, the role of C6orf57 should be studied in mammalian systems using RNAi knockdown and knockout mice. Finally, germline and somatic DNA from PGL/PCC patients should be sequenced at this locus to determine whether mutations in *C6orf57* may play a role in tumorigenesis.

Acknowledgements

This work was supported by a National Science Foundation Graduate Research Fellowship (DSL).

Author contributions

Conceived and designed the study: DSL, VKM. Performed experiments: DSL, VMG, JMB. Performed the computational analyses: DSL, RN, SEC. Performed patient sequencing: NPM, PLD. Wrote the manuscript: DSL, VKM.

References

1. Ghezzi D, Goffrini P, Uziel G, et al. SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy. *Nat Genet.* 2009 Jun;41(6):654-6.
2. Hao HX, Khalimonchuk O, Schraders M, et al. SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science.* 2009 Aug 28;325(5944):1139-42.
3. McNeil MB, Clulow JS, Wilf NM, Salmond GP, Fineran PC. SdhE is a conserved protein required for flavinylation of succinate dehydrogenase in bacteria. *J Biol Chem.* 2012 May 25;287(22):18418-28.
4. Dibrov E, Fu S, Lemire BD. The *Saccharomyces cerevisiae* TCM62 gene encodes a chaperone necessary for the assembly of the mitochondrial succinate dehydrogenase (complex II). *J Biol Chem.* 1998 Nov 27;273(48):32042-8.
5. Horvath R, Abicht A, Holinski-Feder E, et al. Leigh syndrome caused by mutations in the flavoprotein (Fp) subunit of succinate dehydrogenase (SDHA). *Journal of neurology, neurosurgery, and psychiatry.* 2006 Jan;77(1):74-6.
6. Cascon A, Inglada-Perez L, Comino-Mendez I, et al. Genetics of pheochromocytoma and paraganglioma in Spanish pediatric patients. *Endocrine-related cancer.* 2013 Feb 12.
7. Margulis L. Origin of eukaryotic cells. New Haven,: Yale University Press; 1970.
8. Pagliarini DJ, Calvo SE, Chang B, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell.* 2008 Jul 11;134(1):112-23.
9. Barros MH, Johnson A, Gin P, Marbois BN, Clarke CF, Tzagoloff A. The *Saccharomyces cerevisiae* COQ10 gene encodes a START domain protein required for function of coenzyme Q in respiration. *J Biol Chem.* 2005 Dec 30;280(52):42627-35.
10. Strong M, Mallick P, Pellegrini M, Thompson MJ, Eisenberg D. Inference of protein function and protein linkages in *Mycobacterium tuberculosis* based on prokaryotic genome organization: a combined computational approach. *Genome Biol.* 2003;4(9):R59.
11. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 2000 Sep 15;28(18):3442-4.
12. Outten CE, Falk RL, Culotta VC. Cellular factors required for protection from hyperoxia toxicity in *Saccharomyces cerevisiae*. *The Biochemical journal.* 2005 May 15;388(Pt 1):93-101.
13. Dennis DA, Komistek RD, Mahfouz MR, Outten JT, Sharma A. Mobile-bearing total knee arthroplasty: do the polyethylene bearings rotate? *Clinical orthopaedics and related research.* 2005 Nov;440:88-95.
14. Hibbs MA, Hess DC, Myers CL, Huttenhower C, Li K, Troyanskaya OG. Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics.* 2007 Oct 15;23(20):2692-9.
15. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2010 Jan;38(Database issue):D211-22.
16. Steinmetz LM, Scharfe C, Deutschbauer AM, et al. Systematic screen for human disease genes in yeast. *Nat Genet.* 2002 Aug;31(4):400-4.
17. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol.* 2010 Dec;28(12):1248-50.
18. De Meirleir L, Seneca S, Lissens W, et al. Respiratory chain complex V deficiency due to a mutation in the assembly gene ATP12. *Journal of medical genetics.* 2004 Feb;41(2):120-4.
19. Daum G, Bohni PC, Schatz G. Import of proteins into mitochondria. Cytochrome b2 and cytochrome c peroxidase are located in the intermembrane space of yeast mitochondria. *The Journal of biological chemistry.* 1982 Nov 10;257(21):13028-33.
20. Reinders J, Zahedi RP, Pfanner N, Meisinger C, Sickmann A. Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *Journal of proteome research.* 2006 Jul;5(7):1543-54.

21. Busso C, Tahara EB, Ogusucu R, et al. *Saccharomyces cerevisiae* coq10 null mutants are responsive to antimycin A. *The FEBS journal*. 2010 Nov;277(21):4530-8.
22. Chen YC, Taylor EB, Dephoure N, et al. Identification of a protein mediating respiratory supercomplex stability. *Cell metabolism*. 2012 Mar 7;15(3):348-60.
23. Hibbs MA, Myers CL, Huttenhower C, et al. Directing experimental biology: a case study in mitochondrial biogenesis. *PLoS computational biology*. 2009 Mar;5(3):e1000322.
24. Han F, Rao N, editors. Mining Co-regulated Genes Using Association Rules Combined with Hash-tree and Genetic Algorithms. *Communications, Circuits and Systems, 2007 ICCAS 2007 International Conference on; 2007: IEEE*.
25. Bayley JP, Kunst HP, Cascon A, et al. SDHAF2 mutations in familial and sporadic paraganglioma and pheochromocytoma. *The lancet oncology*. 2010 Apr;11(4):366-72.
26. Yang IV, Wade CM, Kang HM, et al. Identification of novel genes that mediate innate immunity using inbred mice. *Genetics*. 2009 Dec;183(4):1535-44.
27. Sim X, Ong RT, Suo C, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS genetics*. 2011 Apr;7(4):e1001363.
28. Wang X, Chua HX, Chen P, et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2013 Feb 27.

CHAPTER VI:

Implications and Future Directions

Through the development of MitoExome sequencing and its application to multiple sets of patients with suspected or confirmed mitochondrial disease, we have gained insight into the genetic architecture of the disease and into the benefits and limitations of next-generation sequencing for the molecular diagnosis of mitochondrial disorders.

The efficacy of next-generation sequencing approaches in mitochondrial disease

Most rare disease sequencing studies report only the successes – i.e., individuals or patient cohorts in which sequencing successfully informed the genetic basis of disease. As a result, the efficacy of next-generation sequencing-based approaches at identifying the genetic basis of disease remains an open question.

The diagnostic efficacy of next-generation sequencing will likely differ among diseases and our laboratory has focused on patients with suspected mitochondrial disorders. We have analyzed two distinct sets of patients in the course of our research. Prior to the work described in this dissertation, our laboratory sequenced and analyzed DNA from a set of 42 severe infantile cases with strong biochemical evidence of mitochondrial disease.¹ As described in Chapter II-IV, we also

sequenced a diverse set of 102 patients referred to Massachusetts General Hospital (MGH) with suspicion of a mitochondrial disorder based on clinical and biochemical criteria. For the set of 42 severe infantile patients, we found that MitoExome sequencing enabled molecular diagnoses in 26% of cases lacking a prior molecular diagnosis, with an additional 29% harboring strong recessive candidate genes, including three genes that have been independently validated or confirmed as disease genes since our publication (*EARS2*, *NDUFB3*, *AGK*).

The diagnostic efficacy for the set of diverse MGH patients was much lower than in the infantile set, however, with only seven new molecular diagnoses amongst 84 patients without prior molecular diagnoses (8%). Although 26 additional patients harbored variants of unknown significance, including 10 patients with recessive mutations in mitochondrial localized proteins or in genes associated with monogenic disorders, it is unclear whether any of these mutations are causal. Our study demonstrates that clinical sequencing currently yields new molecular diagnoses that may not have been achieved through traditional genetic testing methods, however the diagnostic efficacy is far from perfect, currently around 10-30% for mitochondrial disorders. Our results demonstrate that the diagnostic efficacy of sequencing is highly dependent on the set of patients tested, i.e. the denominator in the fraction of molecular diagnoses achieved. Not surprisingly, sequencing-based diagnostic approaches are more likely to be effective in patients with early-onset disease and strong biochemical evidence of an enzymatic or metabolic deficiency. Our results inform the expectations surrounding exome sequencing and its ability to reveal new molecular diagnoses for patients with suspected genetic disorders.

An important question to consider is why there is a difference in the diagnostic efficacy between the severe infantile and diverse MGH patient sets. While there is no clear answer, there are a number of possible explanations. The first is that the diverse MGH cases likely contain a larger fraction of non-mitochondrial diseases, due to the selection criteria of strong biochemical evidence

of mitochondrial dysfunction in the infantile cohort. While we sequenced 200 genes encoding non-mitochondrial proteins known to be associated with monogenic neurological or metabolic disorders -- and in fact found several causal mutations in these genes -- there are hundreds of other Mendelian disease genes that were not sequenced. Furthermore, it could be that a larger number of MGH patients harbored causal mutations in non-mitochondrial proteins not yet known to cause disease, although such mutations will be admittedly hard to identify as pathogenic given the large number of non-mitochondrial proteins relative to our sample size. If feasible, whole-exome sequencing should be performed on the patients who currently lack a molecular diagnosis in order to determine whether such an approach would increase the diagnostic yield. Another possibility is that while severe infantile cases are caused by rare, protein-modifying, highly conserved variants, less severe cases may be caused by mutations that are less rare, less conserved, or may have subtler effects on gene function (e.g., regulatory variants). Such variants may not be fully penetrant. Finding such causal variants will be more challenging and will require many more patients than analyzed at present, due to the higher background frequency of such variants in healthy individuals.

Non-mitochondrial proteins implicated in mitochondrial disease

One of the most striking aspects of our work was the repeated discovery of causal genetic defects that would not classically be considered causes of mitochondrial disorders. Of the seven genes implicated in new molecular diagnoses (*ATP5A1*, *POLG2*, *NDUFV1*, *DPYD*, *KARS*, *WFS1*, and *HSD17B4*), only three encode proteins that are uniquely localized to the mitochondria (*ATP5A1*, *POLG2*, and *NDUFV1*). The others are either dual-localized (*KARS*), or are currently believed to localize outside the mitochondria (*DPYD*, *WFS1*, *HSD17B4*). *KARS* is dual-localized to the cytoplasm and mitochondria, *DPYD* localizes to the cytoplasm, *WFS1* localizes to the endoplasmic reticulum (ER), and *HSD17B4* localizes to the peroxisome. Previous studies have also

identified mutations in non-mitochondrial proteins associated with mitochondrial dysfunction. Two siblings with refractory epilepsy and mitochondrial dysfunction were recognized through exome sequencing to have homozygous nonsense mutations in GM3 synthase (*ST3GAL5*), a lysosomal protein.² Multiple patients with MEGDEL syndrome (3-methylglutaconic aciduria with sensorineural deafness, encephalopathy, and Leigh-like syndrome) were found to carry recessive loss-of-function mutations in *SERAC1*, which localizes to the mitochondrial-ER interface.³ Our study and others demonstrate that secondary mitochondrial dysfunction may be more common than previously thought.

Our study suggests that patients with suspected mitochondrial disorders may frequently harbor causal genetic defects in non-mitochondrial proteins. Given current technological capabilities and their marginal costs, it is likely worthwhile to sequence all previously established Mendelian disease genes in these patients, rather than just those genes canonically associated with mitochondrial disorders. In order to keep up with the continually expanding list of previously established disease genes, it may even be worthwhile for clinical testing laboratories to sequence the patient's whole-exome but to limit the analysis to known disease genes. It is important for both physicians and patients to understand that a patient with clinical, biochemical, or metabolic features suggestive of mitochondrial disease may turn out not to have a primary genetic lesion in a mitochondrial protein. The percentage of suspected mitochondrial disease caused by primary versus secondary mitochondrial dysfunction is currently unknown. Knowing this information could have broad implications for the diagnosis and treatment of suspected mitochondrial disorders.

Future directions

The insights gained through our work have raised as many questions as they have answered. Further study is needed to pursue many of these findings.

The contribution of non-mitochondrial genes to mitochondrial dysfunction

A promising area of mitochondrial disease research is to determine the ways by which organellar crosstalk may lead to mitochondrial dysfunction. As whole-exome sequencing studies are performed on patients with suspected mitochondrial disorders, we expect an increase in the number of non-mitochondrial proteins implicated in these disorders. For the non-mitochondrial proteins implicated in this study (*DPYD*, *HSD17B4*, and *WFS1*), clinical and experimental studies are needed to uncover whether and how their disruption leads to mitochondrial dysfunction. A systematic study of the biochemical, metabolic, mtDNA, and histologic measurement of mitochondrial function in Wolfram syndrome, DPD deficiency, and HSD17B4-deficiency would be informative. For example, it would be worthwhile to study the effect of HSD17B4 knockdown on mitochondrial function and determine whether alterations in fatty acid supplementation may alter mitochondrial function or cell viability.

Genetic and phenotypic data aggregation

Both for mitochondrial disease patients and for patients suspected of other genetic disorders, a key next step is to build platforms enabling shared discoveries among physicians and scientists across the world. Dozens of patients whose genomes have been analyzed by our laboratory have harbored potentially deleterious genetic variants that, due to their rarity, have never before been seen before by our laboratory. However, it is likely that another patient with a similar phenotype and genotype will be found somewhere in the world at some point in the future. There is a need to build a platform that is able to connect the researchers or clinicians observing these patients, and even to connect the patients themselves. A simple solution is the creation of a platform, perhaps one similar to the site www.craigslist.org, whereby clinicians and researchers can post a short summary of the clinical presentation and the genetic findings, and then could be

contacted by others who are interested in the genetic or phenotypic findings. Such an approach would safeguard the anonymity of the information and could greatly increase the flow of information across the globe. An alternative to this approach, which would be more involved but also more powerful, would be to build a shared repository of datasets that can be queried by the research community. Such a platform would greatly accelerate the pace of genetic research and could have a large impact on our understanding of human disease.

Automated phenotype-genotype correlation

Another computational task would be to create an algorithm that could, in an automated fashion, interpret the patient's genetic information in the context of their clinical information. The algorithm would predict the likelihood of a variant or variant set of contributing to disease given a particular clinical phenotype. Much of my PhD was spent manually poring over genetic and phenotypic data of patients and determining whether there was sufficient genotype-phenotype correlation to merit a molecular diagnosis. For example, two of our patients had recessive mutations in *WFS1*, however only one had symptoms consistent with Wolfram syndrome. Another example would be our patient who harbored recessive *HSD17B4* mutations and, though the patient was male, shared many phenotypic traits with a pair of sisters previously diagnosed with Perrault syndrome due to *HSD17B4*-deficiency.

Experimental systems for variant interpretation

Although a number of computational algorithms exist to predict whether a variant may deleterious to protein function, there has been relatively little work done in setting up experimental systems to quantify the impact of a variant. The Duke Center for Human Disease Modeling has shown a number of successful applications of using a zebrafish model to test variants in a wide

variety of genes affecting development,^{4,5} yet this approach is limited to mostly developmental phenotypes and can be quite laborious. One idea is to use cellular models such as the yeast *S. cerevisiae*, the amoeba *Dictyostelium discoideum*, or even mammalian tissue culture models to model cellular phenotypes at a larger scale. For example, given a strong, quantifiable respiratory phenotype upon deletion of the yeast homologue of *POLG*, one could imagine rescuing the deletion strain with every *POLG* allele observed in humans in order to assess variant impact. One could even test potentially dominant-acting variants by expressing *POLG* variants in a wildtype strain, rather than in the deletion, and assessing variant impact. In yeast, there are over 600 mitochondrial proteins with homology to human, of which 200 are known to be associated with a respiratory phenotype. This system could thus be generalized to much of the mitochondrial proteome.

Conclusion

While our understanding of genetics and its relationship to human disease is still limited, it is my hope that the experimental and computational approaches described in this dissertation provide a foundation for future methodologies and discoveries. As the price of sequencing continues to drop and as computational and experimental tools are developed to pinpoint the variants most likely to cause disease, the pace of research is likely to accelerate greatly. Within the next few years, we are likely to uncover the vast majority of genes underlying monogenic diseases and may start to uncover more complex patterns of genetic inheritance. There has never been a more exciting time to participate in biomedical research and I look forward to the discoveries of the coming years.

References

1. Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med*. 2012 Jan 25;4(118):118ra10.
2. Fragaki K, Ait-El-Mkadem S, Chausseot A, et al. Refractory epilepsy and mitochondrial dysfunction due to GM3 synthase deficiency. *European journal of human genetics : EJHG*. 2012 Sep 19.
3. Wortmann SB, Vaz FM, Gardeitchik T, et al. Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. *Nature genetics*. 2012 Jul;44(7):797-802.
4. Davis EE, Zhang Q, Liu Q, et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nature genetics*. 2011 Mar;43(3):189-96.
5. Zaghoul NA, Katsanis N. Zebrafish assays of ciliopathies. *Methods in cell biology*. 2011;105:257-72.

APPENDIX A

Supplementary Materials for Chapter II: Targeted exome sequencing of suspected mitochondrial disorders

Appendix A-1: Supplementary Methods

MitoExome sequencing

Targeted gene transcripts from the RefSeq¹ and UCSC known gene² collections were downloaded from the UCSC genome browser³ assembly hg19 (February 2009) corresponding to all genes listed in Table A-1. Single 120bp baits were synthesized on a chip (Agilent) for each target region, or tiled across targets that exceeded 120bp. Two Agilent bait sets were synthesized: design1 contained 42,119 baits targeted to mtDNA and 1,560 nuclear genes (coding regions and 5' UTRs, single bait per target), and design2 contained 25,963 baits targeted to mtDNA and a superset of 1,598 nuclear genes (coding regions only, two baits per nuclear target and single bait per mtDNA target). Genomic DNA was sheared, ligated to Illumina sequencing adapters, and selected for lengths between 200-350bp. This “pond” of DNA was hybridized with an excess of Agilent baits in solution. The “catch” was pulled down by magnetic beads coated with streptavidin, then eluted.^{4,5} Index tags were added to 60 samples to facilitate “barcoded” multiplex sequencing. All samples were sequenced at the MGH Next-Generation Sequencing Core (<http://nextgen.mgh.harvard.edu>). 42 samples were sequenced in singleplex (one sample per lane, Agilent bait design1) and 60 barcoded samples were multiplexed (4-12 samples per lane, Agilent bait design2).

Variant detection

Sequence alignment, variant detection, and variant annotation for the nuclear and mitochondrial genomes were performed as described previously,⁶ with modifications described here. Illumina reads were aligned to the GRCh37 reference human genome assembly using BWA⁷ and variants were detected and annotated using the Genome Analysis Toolkit (GATK) software package version 1.0.5973⁸ and custom scripts. The ReadBackedPhasing algorithm was used to determine the phase of heterozygous variants. For the analysis of mtDNA, all reads were separately aligned to the mtDNA revised Cambridge Reference Sequence (NC_012920) and heteroplasmy level for SNVs was estimated based on the GATK pooled variant caller (Unified Genotyper v1.0.2986) assuming a mixture of 500 mtDNA chromosomes (-poolSize 500). mtDNA indels were detected using the standard Unified Genotyper and heteroplasmy was estimated as the percentage of aligned reads containing the variant. mtDNA variants supported by $\geq 1\%$ of aligned reads were detected. Variants present at $< 20\%$ heteroplasmy were deemed tentative until confirmed at higher heteroplasmy in an independent DNA sample or a maternal relative. We developed an algorithm to exclude mtDNA variants derived from sequence artifacts or nuclear sequences of mitochondrial origin (NuMTs)⁹, based on filtering out low-heteroplasmy variants ($< 10\%$) that met the following criteria: (i) were present in at least 10% of samples, (ii) showed apparent paternal transmission based on sequencing of five trios, or (iii) exhibited correlation with the nuclear:mitochondrial DNA content across the 102 patients ($p < 0.01$ Bonferroni-corrected, t-test of Spearman correlation). The nuclear:mitochondrial DNA ratio was estimated by the ratio of reads aligning to chromosome 1 to reads aligning to mtDNA, respectively. Large mtDNA deletions or rearrangements were detected

through *de novo* mtDNA assembly using a modified ALLPATHS assembly approach, as previously described.⁶

Confirmatory genotyping and phasing

All variants underlying molecular diagnoses were independently validated by Sanger sequencing, and compound heterozygous variants were phased through read-backed phasing (GATK version 1.0.5973), sequencing familial DNA, or haplotype inference using genotype data from 1000 Genomes. For haplotype inference, two rare variants in the same gene observed together in multiple individuals were considered unlikely to be compound heterozygous and thus excluded, whereas two rare variants in the same gene observed to be mutually exclusive across multiple individuals were considered to be likely compound heterozygous.

Heteroplasmy levels for three known pathogenic mtDNA mutations (MELAS, MERRF, NARP) were quantified at in DNA derived from whole blood and/or patient lymphoblastoid cell lines (LBLs) using ARMS quantitative PCR (Medical Genetics Laboratories, Baylor College of Medicine, Test ID 3006).¹⁰ Dihydropyrimidine dehydrogenase (DPD) deficiency was confirmed by a purine and pyrimidine panel (Mayo Clinic, Test ID 81420).

Modeling the ATP5A1 p.Y321C mutation in yeast

An *atp1* null mutant was created in the W303-1A *S. cerevisiae* yeast strain (*a*, *ade2-1*, *his3-1*, *15*, *leu2-3*, *112*, *trp1-1*, *ura3-1*) by homologous recombination of PCR products using the KanMX resistance marker, as previously described.¹¹ pRS305 plasmids expressing the wildtype or p.Y315C allele of *ATP1* were created via Quikchange (Agilent). Residue numbering is based on the precursor peptide. Multiple sequence alignment was performed using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) with the following UniProt IDs: P25705, Q03265, P35381, P07251, and P0ABB0.

dbGaP accessions

Patient identifiers from this study are listed in dbGaP phs000339 with the prefix “MGO”.

Appendix A-2: Supplementary Notes

MitoExome sensitivity and specificity

For each individual, we detected SNVs and indels compared to the reference human genome. On average, individuals had 24 homoplasmic mtDNA variants, 1470 nuclear SNVs, and 56 nuclear indels. Over 95% of nuclear SNVs in each individual represented common polymorphisms in existing databases.

To assess sensitivity and specificity, we performed MitoExome sequencing on two HapMap individuals (NA12878, NA12891) with publicly available genotype data (HapMap Phase 3 release 2) and deep-coverage whole-genome data (1000 Genomes Project pilot 1¹²). Nuclear SNVs were estimated to be 97.9% sensitive and 99.9% specific, based on 797 and 1655 sites genotyped using independent technology. We detected 78/82 (95%) of indels ranging between 1 and 13bp in size in our HapMap samples, based on comparisons using the same GATK indel detection algorithm on deep-coverage whole genome sequence data, although we note that this algorithm may have a significant false negative rate. Reproducibility was assessed by independent hybrid selection and

sequencing of HapMap sample NA12878, and showed 98% genotype concordance at heterozygous sites, consistent with our reported sensitivity.

The particularly deep coverage of the mtDNA enabled us to detect mtDNA variants with both high and extremely low ($\geq 1\%$) levels of heteroplasmy. mtDNA SNV detection was estimated at 98% sensitivity and 99.5% specificity, based on reports for 29 individuals with independent whole-mtDNA sequence data (Medical Genetics Laboratories, Baylor College of Medicine, Test ID 3055). Additionally, we detected 11/12 sites independently assessed by quantitative PCR (qPCR) from the same tissue sample, including 5/6 variants present at $<10\%$ heteroplasmy (Table A-4). At one site (m.3243A>G) assessed across seven individuals with MELAS, we observed strong correlation ($R^2=0.98$) between the levels of heteroplasmy in LBLs as estimated by NGS read count versus qPCR (Figure A-6 panel A). However, the estimates of heteroplasmy consistently differed by two-fold between these two methods. We note that qPCR validation was performed at only three mtDNA sites and that other study designs, such as sequencing pre-defined mixtures of mtDNA species,¹³ are required to reliably assess the accuracy of NGS-based quantitation of heteroplasmy.

Heteroplasmy in LBL versus whole blood

Previous studies have shown that the levels of heteroplasmy depend on the source of the DNA.¹⁴ Comparison of m.3243A>G heteroplasmy levels by qPCR in LBL versus whole blood revealed consistently higher heteroplasmy levels in whole blood (Figure A-6 panel B). This emphasizes the importance of tissue selection in future NGS studies of mitochondrial disease.

Loss-of-function variants in LBLs from MELAS patients

We observed that lymphoblastoid cell lines derived from MELAS patients carrying the m.3243A>G mutation harbored a large number of low-heteroplasmy loss-of-function (LOF) variants (i.e. frameshift indels and nonsense mutations). Specifically, we observed a tenfold enrichment of LOF mutations: 5 such mutations in samples from 7 MELAS patients (Table A-4) versus 8 in all 102 patients ($p < 1e-3$, binomial exact test). No enrichment was observed for missense or synonymous variants. The five LOF mutations in MELAS patients included four in complex I subunit and two that were previously reported as somatic mutations (m.5260G>A and m.4788G>A).^{15, 16} Interestingly, somatic LOF mutations in Complex I subunits, including the m.5260G>A mutation, have been previously associated with oncocytic thyroid tumors.¹⁷ Further studies are required to evaluate whether the m.3243A>G MELAS mutation causes a preferential accumulation of somatic loss-of-function mtDNA mutations either *in vivo* or in cell culture.

Prioritized variants of unknown significance (pVUS) in mtDNA

Six patients harbored pVUS in mtDNA that did not meet the requirements for molecular diagnosis, as described here:

- Patients 1008 and 1038 harbored pVUS in *RNR1* (m.1555A>G, m.1494C>T) that were not consistent with the previously reported phenotype of aminoglycoside-induced hearing loss.^{18, 19}
- Patient 1011 harbored two heteroplasmic pVUS (*COX1*:p.G140X 42% heteroplasmy, *COX2*:p.W106X, 16% heteroplasmy) that were Sanger-confirmed in DNA from LBLs but were not detected in an independent blood sample from the patient or from an affected daughter, likely representing cell-line mutations.

- Patients 1051, 1085, and 1014 harbored low-heteroplasmy variants that have not yet been confirmed in an independent DNA sample (Table A-5). These include two mutations previously reported in patients with MELAS^{20,21} and a *COX2* nonsense mutation also detected in Patient 1011.

Prioritized variants of unknown significance (pVUS) in nuclear disease genes

11 patients harbored pVUS in established nuclear disease loci that did not meet the requirements for molecular diagnosis, as summarized here:

- Eight patients harbored pVUS that were not consistent with previously reported phenotypes of the relevant genes. Five of these patients harbored rare recessive variants in established autosomal recessive disease genes (*DNAI1*, *DPYD*, *ETFB*, *SECISBP2*, *WFS1*), and four patients harbored rare heterozygous variants in dominant-acting disease genes (*POLG*, *POLG2*).
- Three patients harbored X-linked pVUS that did not have enough support to warrant diagnosis:
 - A hemizygous missense *AIFM1*:p.G53A variant was detected in male patient 1104. The variant was not present in the Exome Variant Server (EVS) nor 1000 Genomes. Previously, a hemizygous frameshift mutation in *AIFM1* was identified in a male with severe encephalopathy.²² More recently, a hemizygous p.G308E missense mutation was identified in a patient with early prenatal ventriculomegaly and postnatal developmental delay.²³ The patient phenotype was not similar enough to either previous report to warrant molecular diagnosis.
 - A hemizygous *HCCS*:p.A174V variant located near a splice site was detected in male patient 1002. The variant was present in 0/2443 male and 1/4060 female X chromosomes in EVS. Previous reports indicate that deletions, nonsense mutations, and *de novo* missense mutations in *HCCS* can cause male-lethality and X-linked dominant microphthalmia in females (OMIM #309801).^{24, 25, 26} The patient phenotype was not similar enough to previous reports to warrant molecular diagnosis
 - A heterozygous *NDUFA1*:p.G32R variant was detected in female patient 1100 with complex I deficiency (22% normalized complex I+III activity). The variant was found to be hemizygous in an affected brother and DNA from a second affected brother was unavailable. This mutation has previously been reported in the heterozygous state in a single female with complex I deficiency and in the hemizygous state in multiple males with complex I deficiency.^{27,28} However this allele is present at 1% allele frequency (AF) in European Americans in the Exome Variant Server and therefore may represent a polymorphism.

Recessive KARS mutations

Patient 1098 harbored compound heterozygous missense mutations in *KARS* (Figure A-2), which were phased through parental genotyping. The p.T587M variant was absent from dbSNP, 1000 Genomes, and Exome Variant Server (EVS)²⁹ and altered a residue conserved to *E. coli*. The p.P228L variant was positioned within the anti-codon binding domain and altered a residue conserved to *C. elegans*, similar to the p.L133H mutation previously shown to impair aminoacylation

in vitro.³⁰ The p.P228L variant was present in 2/12994 chromosomes in EVS. Both patient mutations were predicted to be “probably damaging” by PolyPhen2.³¹

Recessive *KARS* mutations have been associated with peripheral neuropathy (OMIM #613641) based on a single patient with peripheral neuropathy, developmental delay, dysmorphic features, and vestibular Schwannoma.³⁰ Patient 1098, who died by age 4, had no reports of peripheral neuropathy and no EMG or nerve conduction studies were performed to our knowledge. In addition to the single published case, compound heterozygous *KARS* mutations were recently identified in two sibs who presented in infancy with microcephaly, cortical blindness, developmental delay, and seizures (Abstract #701F, 62nd Annual Meeting of the American Society of Human Genetics).³² We hypothesize that peripheral neuropathy could be a later-onset feature of recessive *KARS* mutations, just as recessive mutations in *DARS2* have been associated with childhood- or adolescent-onset peripheral neuropathy in addition to early-onset neurological symptoms such as tremors and ataxia.^{33, 34} Further investigation of patients with recessive *KARS* mutations will likely expand the phenotypic spectrum associated with this disorder.

Hypothesized mechanism of ATP5A1 mutation causing combined respiratory chain deficiency

How might a mutation in the complex V subunit *ATP5A1* lead to combined respiratory chain deficiency and mtDNA depletion? Yeast experiments have shown that nearby mutations in this protein lead to loss of membrane potential and to mtDNA instability.¹¹ We hypothesize that human *ATP5A1* mutations may act in a similar manner. The mutation may cause uncoupling of complex V, which could lead to decreased membrane potential and subsequent loss of mtDNA by an undefined mechanism. The mtDNA loss would lead to decreased synthesis of mtDNA-encoded respiratory chain subunits and thus deficiencies of multiple respiratory chain complexes, as observed in the patient and her affected sister. Future experiments in mammalian cells are needed to explore this hypothesis.

Table A-1. Nuclear gene products targeted in this study

AADAT	ALDH5A1	C5orf33	COX6A2	EIF2B5	GPD2	ITPR1	MRPL21	NDUFA12	PKD2	PSTK	SLC7A7	THEM5
ABCA9	ALDH6A1	C6orf57	COX6B2	EIF3F	GPI	IVD	MRPL22	NDUFA13	PTCD1	PTCD1	SLC7A4	THEM1L
AARS	ALDH7A1	C6orf72	COX6B2	EIF3K	GPI	IYD	MRPL23	NDUFAB1	PKD4	PTCD2	SLC9A6	TIMM8A
AARS2	ALDH9A1	C6orf81	COX6C	ELAC2	GPR56	KARS	MRPL24	NDUFAF1	PDLM2	PTCD3	SLC12A7	TIMM8B
AARS5D1	ALDH18A1	C6orf125	COX7A1	ELN	GPRCS5C	KCNKA1	MRPL27	NDUFAF2	PDP1	PTEN	SLC13A2	TIMM9
AASS	ALDOA	C6orf136	COX7A2	ENDOG	GPT	KCNK3	MRPL28	NDUFAF3	PDP2	PTGES2	SLC16A1	TIMM10
ABAT	ALDOB	C6orf203	COX7A2L	ENO3	GP2	KCNJ10	MRPL30	NDUFAF4	PDPR	PTGR2	SLC17A5	TIMM13
ABCA1	ALKBH7	C7orf10	COX7B	EPRS	GPX1	KIAK0141	MRPL32	NDUFB1	PEX3	PTPMT1	SLC19A2	TIMM17A
ABCD1	ALPL	C7orf55	COX7C	ESR1	GPX4	KIAA0174	MRPL33	NDUFB2	PEX3	PTRH1	SLC22A4	TIMM17B
ABCB1	AMACR	C8orf38	COX8A	ESR2	GR-IPR	KIAA0564	MRPL34	NDUFB3	PDK1	PTRH2	SLC22A5	TIMM22
ABCB6	AMN	C8orf55	COX8C	ETFA	GRN	KIAA0664	MRPL35	NDUFB4	PEBP1	PTS	SLC25A1	TIMM23
ABCB7	AMT	C8orf82	COX10	ETFB	GRPEL1	KIF1B	MRPL36	NDUFB5	PECK1	PUS1	SLC25A2	TIMM44
ABCB8	ANGEL2	C9orf46	COX11	ETFDH	GRPPEL2	KMO	MRPL37	NDUFB6	PECR	PXMP2	SLC25A3	TIMM50
ABCB9	ANKRD13C	C9orf89	COX15	ETHE1	GRSF1	KRT5	MRPL38	NDUFB7	PEPD	PXMP3	SLC25A4	TK2
ABCB10	ANKRD16	C9orf98	COX16	EXO3	GSR	KRT8	MRPL39	NDUFB8	PET112L	PYCR1	SLC25A5	THEM11
ABCC12	ANKRD26	C9orf123	COX17	FABP1	SSS	KRT18	MRPL40	NDUFB9	PEX1	PYGM	SLC25A6	THEM14C
ABCD1	APPB1	C10orf10	COX18	FABP3	GSTA3	KYNU	MRPL41	NDUFB10	PEX3	QDPR	SLC25A10	THEM14E
ABCD2	APBB1	C10orf10	COX19	FADH1	GSTK1	L2HGDDH	MRPL42	NDUFB11	PEX5	QRSL1	SLC25A11	THEM22
ABCD3	APEX2	C10orf58	CPDX	FAMD2A	GSTO1	LACE1	MRPL43	NDUFC1	PEX6	RAB1A	SLC25A12	THEM65
ABCD4	APOA1BP	C11orf48	CP51	FAHD2B	GSTZ1	LACTB	MRPL44	NDUFC2	PEX7	RAB1B	SLC25A13	THEM66
ABCE1	APOL	C12orf10	CPT1A	FAM36A	GFTH25	LACTB2	MRPL45	NDUFS1	PEX10	RAB2B	SLC25A14	THEM70
ABCF2	APTX	C12orf62	CPT1B	FAM38A	GTPBP3	LAMP2	MRPL46	NDUFS2	PEX11B	RAB32	SLC25A15	THEM85
ABHD5	ARAF	C12orf65	CPT1C	FAM65C	GTPBP5	LAP3	MRPL47	NDUFS3	PEX12	RAB3C	SLC25A16	THEM126A
ABHD10	ARG2	C14orf22	CPT2	FAM81A	GTBPB8	LARS2	MRPL48	NDUFS4	PEX13	RAB3D	SLC25A17	THEM143
ABHD11	ARHGEF	C14orf69	CRAT	FAM82A1	GTPBP10	LETM1	MRPL49	NDUFS5	PEX14	RAB4B	SLC25A18	THEM147
ABO	ARHGAP17	C14orf153	CRLS1	FAM82A2	GUCY2D	LNC2	MRPL50	NDUFS6	PEX16	RAB8B	SLC25A19	THEM160
ACAA1	ARMC1	C14orf156	CROT	FAM82B	GUF1	LCT	MRPL51	NDUFS7	PEX19	RAB11A	SLC25A20	THEM186
ACAA2	ARMC4	C14orf159	CRY1	FAM134B	GUSB	LDHA	MRPL52	NDUFS8	PEX26	RAB11B	SLC25A21	THEM205
ACACA	ARMC10	C15orf48	CS	FAM136A	GSY1	LDHALB6B	MRPL53	NDUFS9	PFAS	RAB11FIP5	SLC25A22	THEM223
ACACB	ARMS2	C15orf61	CSF2RB	FAM162A	GSY2	LDHB	MRPL54	NDUFV2	PKKM	RAB18	SLC25A23	TMLHE
ACAD8	ARSA	C17orf42	CTH	FAM167B	HPD	LDHC	MRPL55	NDUFV3	PGAM2	RAB22A	SLC25A24	TMC1C
ACAD9	ARSB	C17orf81	CTNS	FAM195A	HADH	LDHD	MPS2	NEDD8	PGM1	RAB24	SLC25A25	TOM1L2
ACAD10	ARSG	C17orf82	CYP22	FARS2	HADA	LETM1	MRP5	NELL1	PDS1	RAC31	SLC25A26	TOM5
ACAD11	AS3MT	C17orf90	CSTA	FASN	HADHB	LETMD1	MRP56	NEU1	PHB	RAB32	SLC25A27	TOMM7
ACADL	ASAH1	C18orf19	CTSB	FASTK	HAGH	LGSN	MRP57	NEU4	PHB2	RAB35	SLC25A28	TOMM20
ACADM	ASAH2	C18orf55	CTU1	FASTKD1	HAL	LIAS	MRP59	NFS1	PHGDH	RAB39B	SLC25A29	TOMM22
ACADS	ASL	C19orf47	CUBN	FASTKD2	HAMP	LIFR	MRP510	NFU1	PHKA1	RABL3	SLC25A30	TOMM34
ACADSB	ASMTL	C19orf70	CXorf23	HAO1	LIP1T		MRP511	NFXL1	PHKA2	RABL4	SLC25A31	TOMM40
ACADVL	ASD1	C20orf7	CYB5B	FBXO34	HAO2	LIP2T	MRP512	NIF3L1	PHKB	RAF1	SLC25A32	TOMM40A
ACAT1	ATTB1	C20orf24	CYB5R1	FDP5	HARS	LMBRD1	MRP514	NIPSNAP1	PHKG2	RARS	SLC25A33	TOMM47L
ACAT2	ATAD3A	C21orf72	FDX1	FMR2	HARS2	LETM1C33328	MRP515	NIPSNAP3A	PHKH2	RARS2	SLC25A34	TOM1MT
ACBD3	ATAD3B	C21orf53	CYB5R3	FDX1L	HAX1	LONP1	MRP516	NIPSNAP3B	PHT1	RBM16	SLC25A35	TP53
ACLY	ATCAY	C21orf57	CYC1	FDXR	HBXIP	LONP2	MRP517	NT1	PHYH	RDH13	SLC25A36	TP11
ACN9	ATG5	C22orf32	CYCS	FECH	HCCS	LCPAT1	MRP518A	NKIRAS1	PHYHPL	RDH14	SLC25A37	TPMT
ACO1	ATIC	CASA	CYP7B1	FGF14	HCLS1	LRPPRC	MRP518B	NLN	PIGY	REEP1	SLC25A38	TPO
ACO2	ATM	CAB5B	CYP11A1	FGF20	HDDC2	LSM7	MRP518C	NLRX1	PINK1	REXO2	SLC25A39	TPP1
ACOT1	ATN1	CABC1	CYP11B1	FH	HDDH3	LYPLA1	MRP521	NME1	PIPOX	RFK	SLC25A40	TRAP1
ACOT2	ATP1A1	CACNA1A	CYP11B2	FIBP	HEBP1	LYRM1	MRP522	NME1-NME2	PISO	RGM7D1	SLC25A41	TRAPP2L
ACOT4	ATP1A2	CACNA1B	CYP17A1	FIBS1	HEMK1	LYRM2	MRP523	NME2	PITRM1	RHBD1D1	SLC25A42	TRAPP4C
ACOT7	ATP5B	CANX	CYP24A1	FITM2	HEXA	LYRM4	MRP524	NME3	PLAG21B	RHOA	SLC25A43	TRIAP1
ACOT8	ATP5C1	CARKD	CYP27A1	FKBP8	HEXB	LYRM5	MRP525	NME4	PLAG2GA	RHOB	SLC25A44	TRIM37
ACOT9	ATP5D	CARS2	CYP27B1	FLJ35220	HFE2	LYRM7	MRP526	NME6	PLAG2G15	RHOC	SLC25A45	TRIT1
ACOT11	ATP5E	CASP8	D2HGDDH	FLJ40434	HGD		MRP527	NMT1	PLEKHG4	RHOT1	SLC25A46	TRMT2B
ACOT12	ATP5F1	CASQ1	DACT2	FMO3	HGSNAT	MAN2B1	MRP528	NNT	PLIN5	RHOT2	SLC27A3	TRMU
ACOT13	ATP5G1	CBARA1	DAP3	FMR1	HIBADH	MANBA	MRP530	NOP10	PM20D1	RILP	SLC30A6	TRNT1
ACOX1	ATP5G2	CBRA	DARS	FOXRED1	HIBCH	MANBA	MRP531	NPC1	PM20D1	RMN2D1	SLC35C1	TRP6
ACOX3	ATP5G3	CBR3	DAR2	DAR2	HIBCH	MANBA	MRP532	NPC2	PM20D1	RMN2E1	SLC35C2	TRP6
ACP6	ATP5H	CCBL1	DBH	FRMD3	HIGD2A	MAP1D	MRP534	NPEPPS	PMPCA	RNASEH1	SLC37A4	TSGA10
ACSF2	ATP5I	CCBL2	DBI	FTCD	HINT1	MAP3K5	MRP535	NPHS2	PMPCB	RNF130	SLC46A1	TSH23
ACSF3	ATP5J	CCDC51	DBT	FTH1	HINT2	MAPK9	MRP536	NPIPL3	PNKD	RNMTL1	SLMO1	TSPO
ACSL1	ATP5J2	CCDC56	DCAF5	FTMT	HINT3	MAPT	MRF	NPL	PNP	ROBLD3	SLMO2	TST
ACSL3	ATP5L	CCDC58	DCAKD	FTSJ2	HK1	MARCH5	MRS2	NR0B1	PNPLA8	ROMO1	SMCP	TTBC
ACSL4	ATP5O	CCDC90A	DCI	FUCA1	HK2	MARS	MSRA	NR4A2	PNPT1	RPIA	SMCR7L	TTC19
ACSL5	ATP5S	CCDC90B	DCXR	FUNDQ2	HK3	MARS2	MSRB2	NRD1	POLDIP2	RPL10A	SMIP1	TTCS2
ACSL16	ATP5L	CCDC109A	DCXB	DJAUQ	HADG1	MAT2A	MRP533	NPC1	RPLC1A	RPL10B	SLC18A1	TTCS3
ACSM1	ATP7B	CCDC109B	DDC	FXN	HMGCL	MAT2B	MTCH1	NSDL	POLG2	RPL28	SNCAIP	TTPA
ACSM2A	ATP10D	CCDC136	DDT	FXD2	HMGCLL1	MAVS	MTCH2	NT5C	POLM	RPL34	SND1	TTYH1
ACSM3	ATPAF1	CCS	DDX28	G6PC	HMGCS2	MCACT1	MTCP1	NT5C3	POLQ	RPL35A	SNPH	TUFM
ACSM4	ATPAF2	CCCT2	DEC1	GAA	HP	MCAT	MTERF	NT5DC2	POLR2F	RPP14	SNRPN	TXN
ACSM5	ATPIF1	CCCT5	DEC2	GAD1	HPD	MCCC1	MTERFD1	NT5DC3	POLR2G	RPS6KA3	SOD1	TXN2
ACSS1	ATXN1	CCCT7	DELXQ28	GADD45GIP1	HPDL	MCCC2	MTERFD3	NT5M	POLRMT	RPS14	SOD2	TXNDC6
ACSS3	ATXN2	CCCL46	GALC	GALCS	HPC2	MCCP2	MTFM1	NTL1	POR1	RPS15A	SPTA7	TXNRD1
ACX3	ATXN3	CDKL5	DHCR24	GALNS	HPRT1	MCL1	NHPR1	NULP1	PRR1	RPR8	SPATA19	TXNRD2
ACY2P	ATXN7	CDKN2A	DHDP5L	GAMT	HP51	MCOLN1	MTG1	NUD1T	PPA2	RPU5D4	SPEF2	TYMP
ADAMTSL2	ATXN10	CDS2	DHDDH	GARS	HP53	MDH1	MTFHD1	NUD2	PPARGC1A	RPM2B	SPG7	UBR5
ADCK1	AUH	CECR5	DHRS1	GART	HP54	MDH2	MTFHD1L	NUD8	PPDF	RSAD1	SPINK5	UCHL1
ADCK2	AURKAIP1	CEND1	DHRS4	GATC	HP55	ME1	MTFHD2	NUD13	PPIF	RTN4P1	SPR	UCP1
ADCK4	B4GALT5	CERK	DHRS11	GAMT	HP56	ME2	MTFHD2L	NUD19	PPM1B	SACS	SPRYD4	UCP2
ADCK5	BAD	CHAT	DHTKD1	GBA	HRSR12	ME3	MTFR	OAS1	PPM1J	SAMM50	SPTBN2	UCP3
ADHC	BALM	CHAD1	DIABLO	GBE1	HDC2	MCP2	MTFRS	OSR2	PAF1K	SARDH	SFC2	UCRC
ADH5	BAX	CHCHD2	DIABLO	GBE1	HDC2	MCP2	MTFRS	OAT	PPDX	SARS	SCORL	UFC1
ADHFE1	BBOX1	CHCHD3	DIS3	GCAT	HSD3B2	METT11D1	MTIF3	OC1AD1	PPP1CC	SARS2	SSBP1	UGT1A6
ADO	BCAT1	CHCHD4	DLAT	GDCH	HSD3B7	METT15	MTO1	OCRL	PPP2R2B	SCCPDH	STAR	UNG
ADSL	BCAT2	CHCHD6	DLB	GCK	HSD11B1	MFN1	MTP18	OGDH	PPT1	SCN5A	STAR3D3	UPB1
AFG3L1	BCKDHA	CHCHD7	DLST	GCSH	HSD17B3	MFN2	MTPAP	OGDHL	PPTC7	SCO1	STAR4D	UQC
AFG3L2	BCKDHB	CHCHD10	DMDGH	GFER	HSD17B4	MFN5B	MTRF1	OGFDL2	PRDX1	SCO2	STAR5D	UQCR
AGA	BCKDK	CHDH	DNAI1	GFMI	HSD17B8	MGST1	MTRF1L	OGG1	PRDX2	SCP2	STAT5B	UQCRCB
AGK	BCL2	CHDEA	DNAA3	GFME2	HSD17B10	MGST3	MTRR	OLA1	PRDX3	SDHA	STAT3	UQCRC1
AGL	BCL2L1	CISD1	DNAAJ4	GPS1	HSD17B14	MIFER	MTRR	OMA1	PRDX5	SDHAF1	STOM1	UQCRC2
AGMAT	BCL2L2	CKMT1A	DNAAJ11	GGT1	HSDL1	MLYCD	MTX2	OMP	PRDX5	SDHAF2	STOML2	UQCRCF1
AGPAT5	BCL2L13	CKMT1B	DNAAJ15	GHITM	HSDL2	MMAA	MTX3	OPA1	PRDX6	SDHB	STS	UQCRCF
AGPS	BCS1L	CKMT2	DNAAJ19	GHR	HSPA9	MMAB	MUC5B	OPA3	PRELID1	SDHC	SUCLA2	UQCRCQ
AGR2	BHD1	CLIC4	DNAAJ27	GIMAP5	HSPA13	MMAHC	MUL1	OSBPL1A	PRELID2	SDHD	SUCLG1	URO5
AGXT	BHD2	CLN3	DNAAJ30	GK	HSPB7	MMAHDC	MUT	OSGEPL1	PREPL	SDSL	SUCLG2	USMG5
AGXT2	BEAN	CLPB	DNL2	GK2	HSPD1	MOBP	MUTYH	OTC	PRICKLE4	SECISBP2	SUMF1	USP24
AGXT2L2	BLP	DNM1L	GLA	HSP1	HSP1	MOC51	MVK	OXAL1	PRKAG2	SEPT4	SUXO	USP30
AHSA1	BLQC1S1	CLPX	DPY30	GLB1	HTATIP2	MOC52	MYCBP	OXCT1	PKAR2B	SERH2L	SUPV3L1	UXS1
AIFM1	BLCC1S3	CLYBL	DPYD	GLDC	HTRA2	MOSC1	MYL10	OXCT2	PKCA	SERPINA6	SURF1	VAMP1
AIFM2	BNIP3	CM1A	DPY5	GLOD4	HYAL1	MOSC2	MYO5A	OXNAD1	PRKCG	SETX	SYNE1	VAMP8
AIFM3	BNIP3L	CMC1	DRD4	GLRX	IARS	MPG	NADK	OXR1	PRNP	SET3B	SYNJ2BP	VARS2
AK1	BOLA1	CMKP1	DRG2	GLRX2	IARS2	MPST	NAGA	OXSM	PRODHD	SFTPB	TACO1	VAT1
AK2	BPFL	COL2A1	DTNBP1	GLRX3	ICT1	MPV17	NAGLU	PACRG	PRODHD	SFTPC	TAP1	VDAC1
AK3	BRCA1	COMT	DUS2L	GLRX5	ID2	MPV17L	NAGS	PAH	PROSC	SFXN1	TAP1T	VDAC2
AKL1	BRP44L	COPB1	DUS2L	GLS1	ID3	MPEP	NAF1	PAB1	PRSC1	SFXN2	TAP2	VDAC3
AKP1	BRP44L	COP3	DUSP28	GLS2	IDH1	MRE11A	NAPEPLD	PAM16	PRRSL	SFXN3	TARS2	VDR
AKAP10	BSG	COP5A	DUT	GLUD1	IDH2	MRM1	NARS	PANK2	PRSS35	SFXN4	TARS2L	VPS25
AKR1B10	BZRAP1	COP5S	DYM	GLUD2	IDH3A	MRP63	NARS2	PARK7	PSAP	SFXN5	TAT	VARS2
AKR1B15	C1orf31	COP5P	EAR52	GLYAT	IDH3B	MRPL1	NBN	PARL	PSMA1	SGSH	TADN3	WASF1
AKR1C1	C1orf89	COO2	EBP	GLYATL1	IDH3G	MRPL2	NCOA4	PARS2	PSMA2	SH3BP5	TAZ	WBSR16

Table A-2. Variant details for molecular diagnoses and pVUS

This table includes all prioritized variants in the 102 sequenced patients.

Column descriptions for table below	
gene symbol	Official gene symbol from NCBI
chrpos	Chromosome in GRCh37 assembly and position in chromosome. GrCh37 is similar to hg19 except that mitochondrial DNA is revised Cambridge Reference Sequence (CRS; Genbank NC_012920) rather than the chromosome displayed in the UCSC browser.
REF	Reference allele on + strand of GRCh37 reference genome. Note indels are reported using the VCF format conventions.
ALT	Alternate allele on + strand, observed in patient. Note indels are reported using the VCF format conventions.
variant type	nSNP indicates nuclear single-nucleotide variant, mtSNP indicates mtDNA single-nucleotide variant, indel indicates insertion/deletion
position type	CDS indicates coding sequence, utr3 indicates 3' UTR, utr5 indicates 5' UTR
coding coordinate	Coordinate of variant within coding region of transcript
protein coordinate	Coordinate of variant within protein region of transcript (# amino acids)
transcript id	Accession number of RefSeq or UCSC transcript
strand	Strand of transcript on reference assembly
AF (1KG)	Minor allele frequency of the variant, based on 1000 Genomes low-coverage whole-genome data (http://www.1000genomes.org/); missing value represents variant not present in 1KG
AF (EVS_EA)	Minor allele frequency of the variant in European Americans, based on the Exome Variant Server (January 2012) (http://evs.gs.washington.edu/EVS/); missing value represents variant not present in EVS
AF (EVS_AA)	Minor allele frequency of the variant in African Americans, based on the Exome Variant Server (January 2012) (http://evs.gs.washington.edu/EVS/); missing value represents variant not present in EVS
AF (mtDB)	Minor allele frequency of the mtDNA variant, based on mtDB (http://www.mtDB.igp.uu.se/); missing value represents variant not present in mtDB
functional class	Indicates nonsense, missense, synonymous, etc. Includes evolutionary conservation of sites for missense variants (used for the "protein-modifying" filter)
pathogenicity	Predicted/known pathogenicity of variant
PolyPhen2	Predicted consequence of nuclear variants using PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/bgi.shtml) with assembly=hg19, model=HumDiv, transcripts=Canonical, annotations=All

Table A-2 (continued)

gene symbol	chrpos	REF	ALT	variant type	position type	coding coordinate	protein coordinate	transcript id	strand	AF (HG)	AF (EVS EA)	AF (EVS AA)	AF (mtDB)	functional class	pathogenicity	PolyPhen2
DPYD	1:3770547	G	T	rSNP	CDS	c.2303C>A	p.768K	NM_000110	-	0.54%	0.56%	0.11%	missense, highly conserved	likely_deleterious, rare	possibly damaging	
DPYD	1:97915614	C	T	rSNP	intron	c.1905+1		NM_000110	-				splice-donor_1	known_deleterious, rare		
HAO2	1:119934738	T	A	rSNP	CDS	c.777T>A	p.D259E	NM_001005783	+	0.00%	0.00%	0.00%	missense, highly conserved	likely_deleterious, high	benign	
ANGEL2	1:213185682	G	A	rSNP	CDS	c.238C>T	p.P26S	NM_144567	-	0.00%	0.00%	0.00%	missense, highly conserved	likely_deleterious, high	possibly damaging	
ANGEL2	1:213186613	G	A	rSNP	CDS	c.197_206del	p.I66Xfs	NM_144567	-				frameshift, indel	likely_deleterious, high	possibly damaging	
RNA5EH1	2:3598030	A	G	rSNP	CDS	c.442T>C	p.C149R	NM_002936	-				missense, highly conserved	likely_deleterious, high	probably damaging	
RNA5EH1	2:3598048	C	T	rSNP	CDS	c.426G>A	p.V142I	NM_002936	-				missense, highly conserved	likely_deleterious, high	probably damaging	
HADHA	2:26461799	T	C	rSNP	intron	c.180+3		NM_000182	-				splice-donor_3	likely_deleterious, moderate	probably damaging	
WFS1	4:6302689	T	G	rSNP	CDS	c.1167T>G	p.D389E	NM_001145853	+	0.16%	0.05%		missense, highly conserved	likely_deleterious, high	benign	
WFS1	4:6302799	G	A	rSNP	CDS	c.1277G>A	p.C226Y	NM_001145853	+	0.04%	0.00%		missense, moderately conserved	likely_deleterious, moderate	benign	
THG1L	5:157161603	G	A	rSNP	CDS	c.1672C>T	p.R559C	NM_001145853	+	0.14%	0.00%		missense, highly conserved	likely_deleterious, high	probably damaging	
ABCG1	7:87175195	T	A	rSNP	CDS	c.386G>A	p.A130T	NM_017872	-	0.14%	0.07%	0.03%	missense, highly conserved	likely_deleterious, high	benign	
ABCG1	7:87196427	T	C	rSNP	CDS	c.1871A>G	p.K624R	NM_000927	-		0.01%	0.00%	missense, moderately conserved	likely_deleterious, high	benign	
DNAH1	9:3448201	C	T	rSNP	CDS	c.143C>T	p.T48I	NM_000927	-		0.00%		missense, moderately conserved	likely_deleterious, moderate	benign	
DNAH1	9:3452699	T	C	rSNP	CDS	c.245T>C	p.682T	NM_012144	+	0.15%	0.00%	0.03%	missense, highly conserved	likely_deleterious, high	possibly damaging	
SECSBP2	9:1940462	G	GA	indel	CDS	c.304delA	p.N102MfsX23	NM_024077	+				frameshift, indel	likely_deleterious, high	benign	
SECSBP2	9:1943699	G	A	rSNP	CDS	c.699C>A	p.W231	NM_024077	+	0.14%	0.01%	0.21%	missense, moderately conserved	likely_deleterious, moderate	benign	
Ax8	9:135601154	T	A	rSNP	CDS	c.161A>G	p.N454S	NM_152572	-	0.03%	0.00%		missense, moderately conserved	likely_deleterious, moderate	possibly damaging	
Ax8	9:13568127	G	A	rSNP	CDS	c.1015C>T	p.R339C	NM_152572	-	0.42%	0.52%	0.24%	missense, highly conserved	likely_deleterious, high	probably damaging	
NDUFV1	11:67376961	C	T	rSNP	CDS	c.338C>T	p.P113L	NM_001166102	+				missense, moderately conserved	likely_deleterious, moderate	probably damaging	
NDUFV1	11:67377094	CT		indel	CDS	c.472delI	p.S158PfsX15	NM_001166102	+				frameshift, indel	likely_deleterious, high	probably damaging	
ACACB	12:109634819	C	T	rSNP	CDS	c.488C>T	p.R330W	NM_001093	+	0.18%	1.28%	0.27%	missense, highly conserved	likely_deleterious, high	probably damaging	
ACACB	12:109673448	A	T	rSNP	CDS	c.444A>T	p.D1481V	NM_001093	+		0.01%	0.00%	missense, highly conserved	likely_deleterious, high	probably damaging	
POLG	15:8965011	G	A	rSNP	CDS	c.255A>T	p.R852C	NM_001126131	-		0.00%		missense, moderately conserved	known_deleterious, rare	probably damaging	
POLG	15:89665023	C	T	rSNP	CDS	c.2542G>A	p.G848S	NM_001126131	-		0.01%	0.05%	missense, moderately conserved	known_deleterious, rare	probably damaging	
POLG	15:89668750	C	T	rSNP	CDS	c.1860G>A	p.R627Q	NM_001126131	-		0.01%	0.00%	missense, moderately conserved	known_deleterious, rare	probably damaging	
POLG	15:89668809	C	T	rSNP	CDS	c.1825G>A	p.G609S	NM_001126131	-		0.16%	0.08%	missense, moderately conserved	likely_deleterious, moderate	probably damaging	
POLG	15:89670432	C	T	rSNP	CDS	c.1399G>A	p.A67T	NM_001126131	-	0.10%	0.19%	0.03%	missense, highly conserved	known_deleterious, rare	probably damaging	
POLG	15:89871763	G	C	rSNP	CDS	c.1174C>G	p.L392V	NM_001126131	-				missense, highly conserved	known_deleterious, rare	probably damaging	
POLG	15:89876498	G	A	rSNP	CDS	c.488C>T	p.P163I	NM_001126131	-				missense, highly conserved	likely_deleterious, high	probably damaging	
POLG	15:89876558	G	A	rSNP	CDS	c.428C>T	p.A143V	NM_001126131	-				missense, highly conserved	known_deleterious, rare	probably damaging	
POLG	15:89876994	C	T	rSNP	CDS	c.332G>A	p.G11D	NM_001126131	-				missense, moderately conserved	known_deleterious, rare	benign	
NDUFAB1	16:23593662	G	A	rSNP	CDS	c.394G>T	p.D132Y	NM_005003	-				missense, highly conserved	likely_deleterious, high	probably damaging	
KARS	16:75662486	G	C	rSNP	CDS	c.1760C>T	p.L587M	NM_001130089	-				missense, highly conserved	likely_deleterious, high	probably damaging	
KARS	16:7566980	G	A	rSNP	CDS	c.683C>T	p.P228L	NM_001130089	-	0.03%	0.00%		missense, highly conserved	likely_deleterious, high	possibly damaging	
ACACA	17:35444292	C	T	rSNP	CDS	c.7111G>A	p.V237I	NM_198834	-				missense, highly conserved	likely_deleterious, high	benign	
ACACA	17:35627755	T	G	rSNP	intron	c.1009-3		NM_198834	-	0.19%	0.37%	0.08%	splice-acceptor_3	likely_deleterious, moderate	benign	
RSAD1	17:48558270	C	T	rSNP	CDS	c.25C>T	p.R9C	NM_018346	+				missense, moderately conserved	likely_deleterious, moderate	benign	
POLG2	17:62476433	T	G	rSNP	CDS	c.1265A>C	p.Q422P	NM_007215	-				missense, moderately conserved	likely_deleterious, moderate	benign	
POLG2	17:62481850	T	C	rSNP	CDS	c.1056A>G	p.R369G	NM_007215	-				missense, highly conserved	likely_deleterious, high	probably damaging	
ATP5A1	18:43661788	T	C	rSNP	CDS	c.962A>G	p.Y321C	NM_001001937	-				missense, highly conserved	likely_deleterious, high	probably damaging	
MAN2B1	19:12763287	C	T	rSNP	CDS	c.1886G>A	p.R613Q	NM_000628	-				missense, highly conserved	likely_deleterious, high	possibly damaging	
MAN2B1	19:12776557	G	T	rSNP	CDS	c.222C>A	p.D74E	NM_000528	-				missense, highly conserved	likely_deleterious, high	probably damaging	
ECHT1	19:39306573	G	A	rSNP	CDS	c.806C>T	p.P298L	NM_001398	-	0.00%	0.03%		missense, highly conserved	likely_deleterious, high	probably damaging	
ECHT1	19:39322045	T	C	rSNP	CDS	c.164A>G	p.Y55C	NM_001398	-	0.00%	0.64%		missense, highly conserved	likely_deleterious, high	probably damaging	
ETFB	19:5180170	G	A	rSNP	CDS	c.854C>T	p.L285M	NM_00104763	-	0.01%	0.00%		missense, highly conserved	likely_deleterious, high	probably damaging	
ETFB	19:51857874	C	A	rSNP	CDS	c.19G>T	p.V7F	NM_00104763	-	0.29%	0.08%		missense, moderately conserved	likely_deleterious, moderate	benign	
CERK	22:47091143	A	C	rSNP	CDS	c.1013T>G	p.V386G	NM_022766	-		0.02%	0.00%	missense, moderately conserved	likely_deleterious, moderate	possibly damaging	
ASMTL	X:1558040	G	C	rSNP	intron	c.52-3		NM_001173473	-				splice-acceptor_3	likely_deleterious, moderate	possibly damaging	
HCCS	X:1138740	C	T	rSNP	CDS	c.52T>C	p.A174V	NM_001122608	+	0.00%	0.00%	0.03%	splice-donor_1	likely_deleterious, moderate	benign	
MAOA	X:43603735	A	G	rSNP	CDS	c.1559A>G	p.K520R	NM_000240	+	0.07%	0.00%		missense, highly conserved	likely_deleterious, high	benign	
APEX2	X:58028663	C	T	rSNP	CDS	c.421C>T	p.R141C	NM_014481	+	0.60%	0.06%		missense, unknown conservation	likely_deleterious, moderate	benign	
ACSL4	X:108921620	G	A	rSNP	CDS	c.680C>T	p.T227M	NM_004458	-	0.23%	0.06%		missense, unknown conservation	likely_deleterious, moderate	benign	
NDUFA1	X:119005988	C	C	rSNP	CDS	c.94G>C	p.G52R	NM_004541	+	1.05%	0.06%		missense, highly conserved	known_deleterious, rare	possibly damaging	
AIFM1	X:129289210	G	rSNP	CDS	non_coding, exon	c.158G>C	p.G53A	NM_145812	-				missense, moderately conserved	likely_deleterious, moderate	possibly damaging	
RNR1	CRS:1494	C	T	mtSNP	non_coding, exon			RNR1	+	0.04%			non-coding	known_deleterious, rare	benign	
RNR1	CRS:1555	A	G	mtSNP	non_coding, exon			RNR1	+	0.44%			non-coding	known_deleterious, rare	benign	
RNR1	CRS:3243	A	G	mtSNP	non_coding, exon			RNR1	+				non-coding	known_deleterious, rare	benign	
RNR1	CRS:3271	T	C	mtSNP	non_coding, exon			RNR1	+				non-coding	known_deleterious, rare	benign	
ND1	CRS:3949	T	C	mtSNP	CDS	c.648T>C	p.Y216H	ND1	+				missense, highly conserved	likely_deleterious, high	probably damaging	
ND1	CRS:3659	G	A	mtSNP	CDS	c.353G>A	p.W118*	ND1	+				nonsense	likely_deleterious, high	benign	
ND2	CRS:4788	G	A	mtSNP	CDS	c.319G>A	p.G107*	ND2	+				nonsense	likely_deleterious, high	benign	
ND2	CRS:5260	G	A	mtSNP	CDS	c.791G>A	p.W264*	ND2	+				nonsense	likely_deleterious, high	benign	
COX1	CRS:6321	G	A	mtSNP	CDS	c.418G>A	p.G140*	COX1	+				nonsense	likely_deleterious, high	benign	
TRNS1	CRS:7472	A	T	mtSNP	non_coding, exon			TRNS1	-	0.04%			non-coding	known_deleterious, rare, site	likely_deleterious, high	
COX2	CRS:7902	G	A	mtSNP	CDS	c.317G>A	p.W106*	COX2	+				nonsense	likely_deleterious, high	benign	
TRNK	CRS:8344	A	G	mtSNP	non_coding, exon			TRNK	+	0.04%			non-coding	known_deleterious, rare	benign	
ATP6	CRS:8993	T	C	mtSNP	CDS	c.467T>C	p.L156P	ATP6	+				missense, highly conserved	known_deleterious, rare	benign	
ATP6	CRS:8993	T	G	mtSNP	CDS	c.467T>G	p.L156R	ATP6	+				missense, highly conserved	known_deleterious, rare	benign	
ND4	CRS:11866	A	AC	mtindel	CDS	c.1108delC	p.P370PfsX12	ND4	+	0.04%			frameshift, indel	likely_deleterious, high	benign	
ND6	CRS:1484	A	C	mtSNP	CDS	c.190A>G	p.M64V	ND6	+	0.26%			missense, poorly conserved	known_deleterious, rare	benign	
CYTB	CRS:15683	C	T	mtSNP	CDS	c.937C>T	p.Q313*	CYTB	+				nonsense	likely_deleterious, high	benign	

Table A-3: MitoExome sequencing statistics

Median values per patient are shown with range across patients indicated in brackets (N=102).

Targeted DNA	Nuclear DNA	mtDNA
# gene loci	1589	37
target size (bp)	2,297,893	16,569
coverage		
mean coverage	226 [105-401]	12,680 [3,805-19,253]
% targeted bp covered $\geq 1X$	97 [95-98]	100 [100]
% targeted bp covered $\geq 10X$	95 [90-97]	100 [100]
% targeted bp covered $\geq 20X$	93 [88-96]	100 [100]
# variants compared to reference DNA*	1521 [1427-1822]	55 [21-291]
# rare variants	40 [23-94]	9 [0-51]
# rare, likely deleterious variants	21 [11-40]	2 [0-10]
# genes with 2 rare, likely deleterious variants	0 [0-3]	NA

* mtDNA variants: $\geq 1\%$ heteroplasmy

Table A-4. Quantification of mtDNA heteroplasmy in patients with prior mtDNA diagnoses

Percentages represent heteroplasmy levels derived from qPCR or from next-generation sequencing (NGS) read counts. Previously reported somatic mutations are indicated in bold (MITOMAP³⁵). Abbreviations: LBL, lymphoblastoid cell line; MELAS, mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes; MERRF, myoclonic epilepsy with ragged red fibers; hom., homoplasmic; het., heteroplasmic; nd, no data; KSS, Kearns-Sayre Syndrome; NARP, Neuropathy, ataxia, and retinitis pigmentosa;

Patient ID	Patient Dx	OMIM #	mtDNA mutation (Prior Dx)	Observed heteroplasmy			Other prioritized variants
				qPCR (Blood)	qPCR (LBL)	NGS (LBL)	
1107	MELAS	540000	m.3243A>G	37%	7%	2%	-
1110	MELAS	540000	m.3243A>G	47%	9%	4%	m.3659G>A 2% (ND1;p.W118X); m.5260G>A 1% (ND2;p.W264X)
1111	MELAS	540000	m.3243A>G	72%	52%	26%	-
1119*	MELAS	540000	m.3243A>G	27%	3%	0.7%	m.11866A>AC 16% (ND4;p.P370PfsX12)
1120	MELAS	540000	m.3243A>G	37%	35%	14%	m.4788G>A 2% (ND2;p.G107X)
1122	MELAS	540000	m.3243A>G	39%	9%	3%	m.15683C>T 1% (CYTB;p.Q313X)
1123	MELAS	540000	m.3243A>G	14%	4%	2%	-
1108	MERRF	545000	m.8344A>G	14%	0%	1%	-
1116	MERRF	545000	m.8344A>G	53%	67%	91%	-
1121	MERRF	545000	m.8344A>G	Hom	nd	71%	-
1112	NARP	551500	m.8993T>G	Hom	nd	84%	-
1124	NARP	551500	m.8993T>C	97%	nd	99%	-
1012	KSS	530000	m.3643_15569del	Het	nd	18%	-

* Variant not called by NGS pipeline due to heteroplasmy <1%

Table A-5. Prioritized variants of unknown significance (pVUS) in 84 patients without prior molecular diagnoses

Prioritized variants (defined in Methods) are listed using Human Genome Variation Society (HGVS) nomenclature along with the patient genotype (hom., homozygous; het., heterozygous; hem., hemizygous; hp, heteroplasmy) and brief clinical description. Age represents age at diagnosis. Underline denotes established disease gene as in Figure 2.2. Variants separated by “+” are compound heterozygous. Variants separated by “,” are potential compound heterozygous (unphased). Genomic coordinates are in Table A-2.

Abbreviations: DD, developmental delay; DM, diabetes mellitus; F, female; FTT, failure to thrive; GI, gastrointestinal; HL, hearing loss; HTN, hypertension; IUGR, intrauterine growth restriction; LD, Leigh's disease; M, male; mo, month-old; PDD, pervasive developmental delay; PEO, progressive external ophthalmoplegia; SNHL, sensorineural hearing loss; SS, short stature.

^a Variant present at 28% in maternal DNA

^b Not detected by Sanger sequencing in whole blood of patient or affected daughter

^c Patient harbored variant in *POLG2* underlying diagnosis of PEO, see Table 2.1

^d Variant found to be hemizygous in an affected brother; other brother untested

Patient ID	Clinical description	Prioritized variants of unknown significance (pVUS)
1002	57 yo M w/ PEO, ptosis, dysphagia, muscle weakness, exercise intolerance, mild HL	<u>HCCS</u> :p.A174V (hem.)
1003	17 yo F w/ LD, myopathy, migraine, exercise intolerance, SS, GI dysmotility	<i>ACACB</i> :p.R830W+p.D1481V
1007	58 yo M w/ myopathy, exercise intolerance, sleep disorder, GI dysmotility	<u>DNAI1</u> :p.T481+p.I682T
1008	4 yo M w/ GI dysmotility, dyspnea, exercise intolerance, dysphagia, SS	<u>RNR1</u> m.1555A>G (3% hp) ^a
1011	61 yo F w/ leukodystrophy, progressive dementia, gait difficulties, incontinent bladder and bowel; deceased; family history (sister and two daughters with leukodystrophy)	<u>COX1</u> :p.G140X (42% hp) ^b <u>COX2</u> :p.W106X (16% hp) ^b
1014	6 yo M w/ loss of skills, seizures, GI dysmotility, autonomic dysfunction	<u>COX2</u> :p.W106X (1% hp)
1015	5 yo F w/ PDD/Autism, GI dysmotility	<u>SECISBP2</u> : p.M233I,p.N102MfsX23
1019 ^c	53 yo F w/ bilateral ptosis, PEO, ataxia, fatigue, swallowing dysfunction;family history (3 siblings with PEO)	<i>RNASEH1</i> :p.V142I+p.C148R
1020	3 mo F w/ FTT, microcephaly, encephalopathy, IUGR, hypotonia, pulmonary hypertension, heart failure; d. 3 mo; consanguinity (parents first cousins); family history (affected sister)	<i>ATP5A1</i> :p.Y321C (hom.) <i>HAO2</i> :p.D259E (hom.)
1025	17 yo F w/ myoclonus, myopathy, migraines, exercise intolerance, GI dysmotility	<u>POLG2</u> :p.Q422P (het.)
1026	33 yo F w/ myoclonic epilepsy, migraines, exercise intolerance, encephalopathy	<i>ANGEL2</i> :p.P80S+p.Y68Xfs; <i>MAN2B1</i> :p.R613Q,p.D74E; <u>POLG</u> :p.G609S (het.)
1032	23 yo M w/ brainstem stroke, lactic acidemia, SNHL, optic neuropathy, exercise intolerance	<i>NDUFAB1</i> :p.D132Y (hom.) <u>DPYD</u> :p.T768K (hom.)

Table A-5 (continued)

1038	56 yo M w/ spastic paraparesis, optic atrophy, HL, bladder/bowel dysfunction, dysphagia	<i>RNR1</i> m.1494C>T (98% hp)
1040	31 yo F w/ visual loss, myopathy, GI dysmotility, tachycardia, exercise intolerance	<i>AK8</i> :p.N454S+p.R339C
1044	51 yo F w/ myopathy, SNHL, sleep disorder, GI dysmotility/reflux, DM, autonomic dysfunction	<i>ACACA</i> :p.E227K+c.1009-3T>G
1046	37 yo F w/ retinopathy, SNHL, exercise intolerance, GI dysmotility, headache	<i>THG1L</i> :p.A130T (hom.)
1051	2 yo F w/ GI dysmotility, fatigue, dystonia, DD, hypotonia, stroke-like events, migraines	<i>ND1</i> m.3949T>C (3% hp)
1067	30 yo F w/ ataxia, hypogonadotropic hypogonadism, hyperhidrosis, cognitive decline, migraine	<i>ETFB</i> :p.T285M+p.V7F <i>ABCB1</i> :p.K624R+p.I221V <i>POLG</i> :p.L392V (het.)
1072	20 yo F w/ migraines, exercise intolerance, GI dysmotility, autonomic dysfunction, myoglobinuria	<i>CERK</i> :p.V338G (hom.) <i>WFS1</i> :p.D389E+p.C426Y
1083	3 yo M w/ mild language delay, acute hypoxic encephalopathy	<i>APEX2</i> :p.R141C (hem.)
1085	53 yo F w/ DM, muscle weakness, exercise intolerance	<i>TRNL1</i> m.3271T>C (2% hp)
1088	6 yo M w/ pseudo obstruction, glucose intolerance, B cell dysfunction, DD, SS	<i>MAOA</i> :p.K520R (hem.)
1091	19 yo F w/ GI dysmotility, migraine, hypoplasia cerebellum, vision loss, LD, seizures, balance problems, autonomic dysfunction, hypotonia, immune deficiencies	<i>POLG</i> :p.P163L (het.)
1092	1 yo F w/ encephalopathy, vision loss, HL, myoclonus, dystonia, seizures, GI dysmotility; d. 1 yo	<i>ECH1</i> :p.P269L,p.Y55C
1100	15 yo F w/ GI dysmotility, cyclical vomiting, fatigue, exercise intolerance, headaches; family history (two affected brothers)	<i>NDUFA1</i> :p.G32R (het.) ^d
1104	1 yo M w/ global DD, dystonia, GI dysmotility, migraines, seizures	<i>ACSLA</i> :p.T227M (hem.); <i>AIFM1</i> :p.G53A (hem.); <i>ASMTL</i> :c.52-3G>C (het.)

Table A-6. Prior molecular diagnoses in nuclear DNA recovered by targeted exome sequencing

Abbreviations: SANDO, sensory ataxia neuropathy dysarthria and ophthalmoplegia; MIRAS, mitochondrial recessive ataxia syndrome; PDHAD, Pyruvate dehydrogenase E1 alpha deficiency; het., heterozygous; hom., homozygous; cmpd het., compound heterozygous; Mol. Dx, molecular diagnosis

Patient ID	Previous diagnosis	OMIM	Gene	Type	Recovered mutations underlying Mol. Dx	Other prioritized variants
1117	Mitochondrial Trifunctional Protein Deficiency	609015	<i>HADHA</i>	hom.	c.180+3A>G	-
1118	SANDO	607459	<i>POLG</i>	cmpd het.	p.R852C + [p.R627Q; p.G11D]	<i>RSAD1</i> :p.R9C (hom.)
1113	SANDO	607459	<i>POLG</i>	cmpd het.	p.G848S + p.A143V	-
1115	MIRAS	607459	<i>POLG</i>	hom.	p.A467T	-
1114	PDHAD	312170	<i>PDHA1</i>	hemizygous	c.963_977dup21	-

Table A-7. Diagnostic yield in patient subsets

Rate of molecular diagnosis (Mol. Dx) within five subsets of patients: (i) ETC activity $\leq 30\%$ in any complex, normalized to citrate synthase, (ii) “Definite” Morava score, (iii) Infantile age of diagnosis ≤ 2 years, (iv) Family history, based on first-degree relative with same phenotype, and (v) Deceased status. Patients with molecular diagnoses included all those listed in Table 2.1, Table A-4, and Table A-6. P-value calculated with one-tailed binomial exact test relative to category “All” (first-row). Abbreviations: ns, not significant at $p < 0.01$

Category	Patients without prior Mol. Dx (N=84)			All patients (N=102)		
	# Mol. Dx/ # in category	% Mol. Dx	P-value	# Mol. Dx/ # in category	% Mol. Dx	P-value
All	6/84	7%		24/102	24%	
ETC $\leq 30\%$	3/52	6%	ns	0/0	-	-
Definite	4/24	17%	ns	8/28	29%	ns
Infantile	2/11	22%	ns	0/0	-	-
Family history	3/6	50%	<0.01	8/11	73%	<0.01
Deceased	2/5	40%	ns	3/7	43%	ns

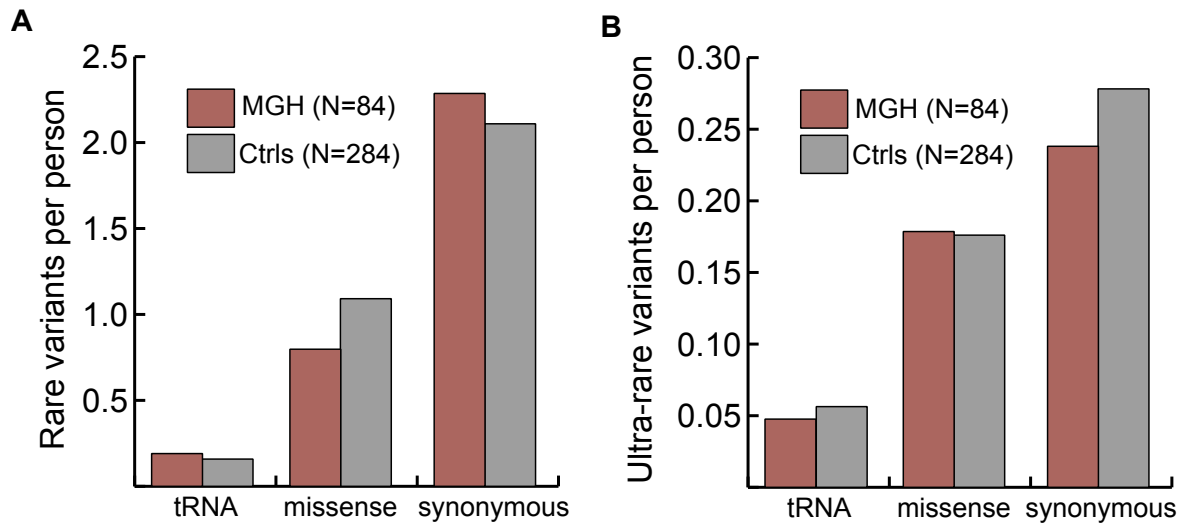


Figure A-1. Number of rare mtDNA variants in cases and controls

(A) Mean number of rare mtDNA variants, defined as allele frequency (AF) <0.3% in mtDB and AF<10% in sequenced cases.

(B) Mean number of ultra-rare mtDNA variants, defined as not present in mtDB and AF<2% in sequenced cases.

MGH refers to 84 cases lacking previous molecular diagnoses. Ctrls refers to 284 European controls from 1000 Genomes Project. Analysis limited to high heteroplasmy variants (at least 50% heteroplasmy based on read count) that are not known to cause disease (MITOMAP).

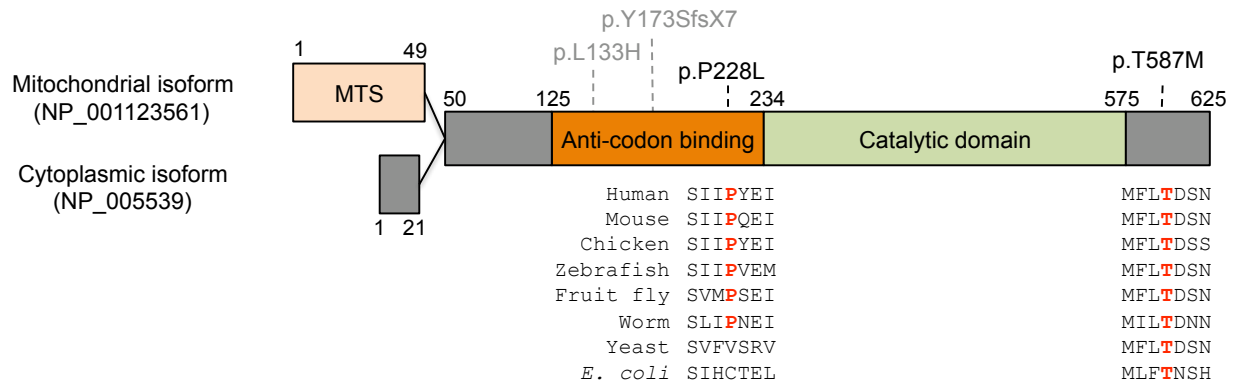


Figure A-2. Mutations in the KARS protein

Schematic diagram depicts two splice isoforms of KARS protein that differ at the N-terminus, with coordinates above listed relative to mitochondrial isoform. Black text indicates mutations detected in patient 1098, gray text indicates previously reported mutations.³⁰ ClustalW2 multiple sequence alignment shown for each patient mutation; UniProt identifiers: Q15046, Q99MN1, Q5ZKP8, F1QSF6, Q8SXM8, Q22099, P15180, P0A8N5. Abbreviations: MTS, mitochondrial targeting sequence

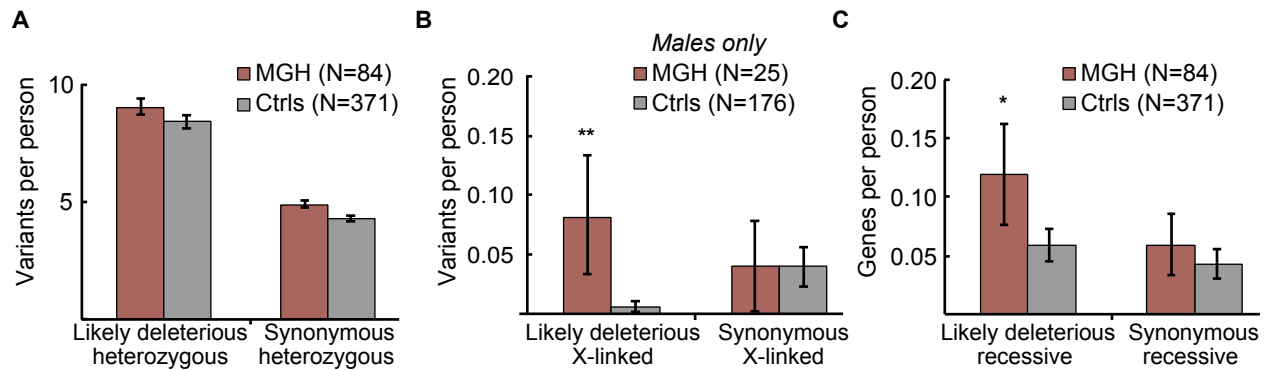


Figure A-3. Number of variants in genes not previously linked to mitochondrial disease

(A) Mean number of rare, likely deleterious, heterozygous variants.

(B) Mean number of rare, likely deleterious, hemizygous, X-linked variants in males.

(C) Mean number of autosomal genes containing rare, likely deleterious, recessive (i.e. homozygous or two heterozygous) variants.

MGH refers to 84 cases lacking previous molecular diagnoses. Ctrl refers to 371 healthy individuals from the NIMH Control repository. Rare and likely deleterious are as defined in Methods. The relevant number of rare, synonymous variants is shown in each panel as a control. Analysis limited to MitoCarta genes not previously linked to mitochondrial disease. Asterisks indicate significance based on binomial exact test (* $p < 0.05$, ** $p < 0.01$). Error bars represent standard error of the mean.

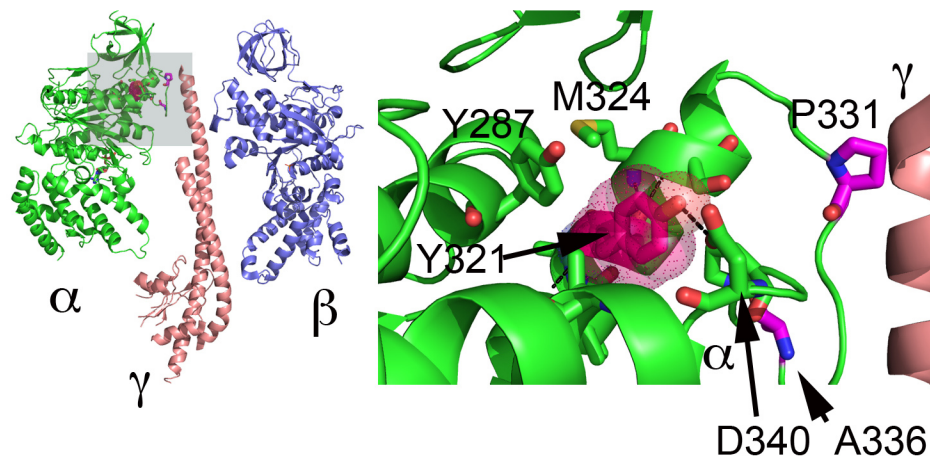


Figure A-4. Position of ATP5A1 patient mutation relative to yeast *mg1* mutations

Schematic of F₁ ATPase 3D structure based on x-ray crystal structures of the bovine³⁶ and yeast³⁷ complexes. Red indicates position of the ATP5A1:p.Y321 residue in the alpha subunit (coordinates based on human precursor peptide). Zoomed inset shows that patient mutation (α Y321) forms a H-bond with α D340 and is within 4Å of α Y287, α Q317- α S325, α R330, α P338, α D340. Note α P331 and α A336 are known *mg1* mutation sites.¹¹ These residues lie near the a-g interface and are thought to be important for coupling of the F₁ and F_o sub-complexes.³⁸ Visualization was performed using Pymol³⁹ with coordinates of the bovine and yeast F₁ ATPase (pdb, 1W0J and 2HLD).

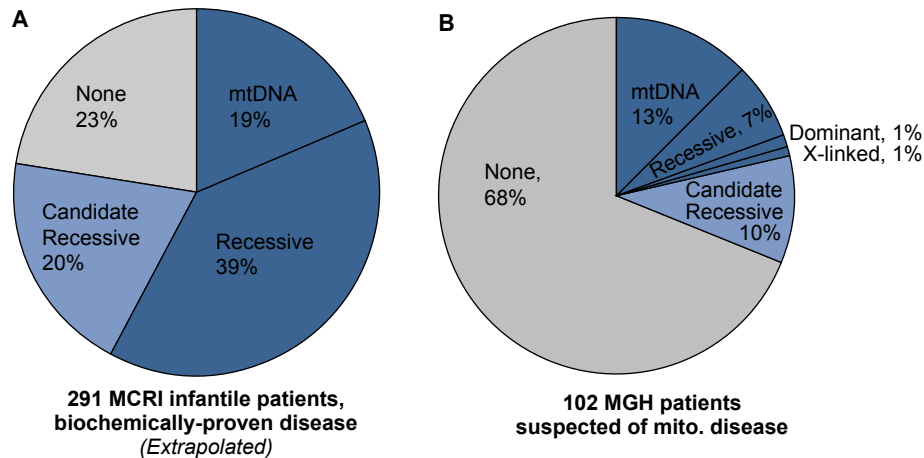


Figure A-5. Molecular diagnoses and recessive candidates in two patient groups

Comparison of MitoExome sequence results from our previous study⁶ and current study shows the fraction of patients with firm molecular diagnoses (dark blue) and autosomal recessive candidate genes (light blue). Comparison was limited to the 1,034 genes sequenced in both studies (thus excludes 212 DDx genes). Unlike Figure 2.4, this comparison shows all cases, including those with prior molecular diagnoses from traditional genetic testing.

(A) Murdoch Childrens Research Institute (MCRI) cohort of 291 infantile patients with severe ETC defects⁶: 124 with molecular diagnosis based on traditional sequencing, 38 with MitoExome sequencing, and 129 without MitoExome sequencing. MitoExome sequence results were extrapolated from the 38 patients with MitoExome sequencing to the 129 patients without MitoExome sequencing. New diagnoses from the extrapolation were then added to 124 prior diagnoses.

(B) 102 MGH patients: 18 with molecular diagnosis based on traditional sequencing (11 of which had diagnoses prior to referral to MGH), and 84 patients without prior molecular diagnosis (selected from a larger set of 159 patients as described in Methods).

As a caveat, we note that different methods of ascertainment were used for each group. Future prospective studies will be required to explore differences in the genetic architecture between these two groups.

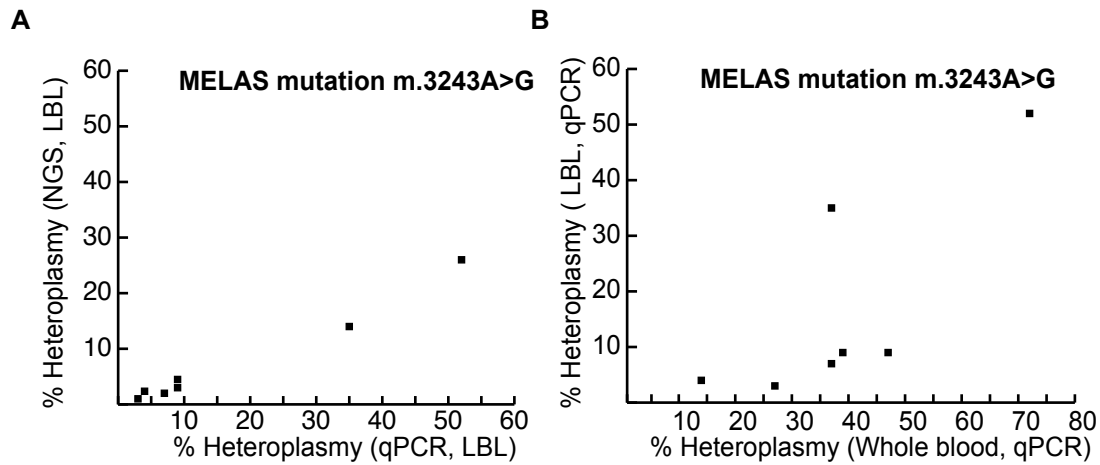


Figure A-6. Heteroplasmy of m.3243A>G across 7 patients with MELAS

- (A) Quantification of heteroplasmy based on qPCR versus next-generation sequencing (NGS) read-counts in the same LBL-derived DNA sample.
- (B) Quantification of heteroplasmy based on qPCR of DNA from whole blood versus lymphoblastoid cell line (LBL).

References

1. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D61-5.
2. Fujita PA, Rhead B, Zweig AS, et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D876-82.
3. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res.* 2002 Jun;12(6):996-1006.
4. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* 2009 Feb;27(2):182-9.
5. Fisher S, Barry A, Abreu J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 2011 Jan 4;12(1):R1.
6. Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science translational medicine.* 2012 Jan 25;4(118):118ra10.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009 Jul 15;25(14):1754-60.
8. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010 Sep;20(9):1297-303.
9. Tourmen Y, Baris O, Dessen P, Jacques C, Malthiery Y, Reynier P. Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics.* 2002 Jul;80(1):71-7.
10. Wang J, Venegas V, Li F, Wong LJ. Analysis of mitochondrial DNA point mutation heteroplasmy by ARMS quantitative PCR. *Curr Protoc Hum Genet.* 2011 Jan;Chapter 19:Unit 19 6.
11. Wang Y, Singh U, Mueller DM. Mitochondrial genome integrity mutations uncouple the yeast *Saccharomyces cerevisiae* ATP synthase. *The Journal of biological chemistry.* 2007 Mar 16;282(11):8228-36.
12. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* Oct 28;467(7319):1061-73.
13. Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *American journal of human genetics.* 2010 Aug 13;87(2):237-49.
14. Shanske S, Pancrudo J, Kaufmann P, et al. Varying loads of the mitochondrial DNA A3243G mutation in different tissues: implications for diagnosis. *American journal of medical genetics Part A.* 2004 Oct 1;130A(2):134-7.
15. Porcelli AM, Ghelli A, Ceccarelli C, et al. The genetic and metabolic signature of oncocyctic transformation implicates HIF1alpha destabilization. *Human molecular genetics.* 2010 Mar 15;19(6):1019-32.
16. Taylor RW, Barron MJ, Borthwick GM, et al. Mitochondrial DNA mutations in human colonic crypt stem cells. *J Clin Invest.* 2003 Nov;112(9):1351-60.
17. Gasparre G, Porcelli AM, Bonora E, et al. Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocyctic phenotype in thyroid tumors. *Proceedings of the National Academy of Sciences of the United States of America.* 2007 May 22;104(21):9001-6.
18. Estivill X, Govea N, Barcelo E, et al. Familial progressive sensorineural deafness is mainly due to the mtDNA A1555G mutation and is enhanced by treatment of aminoglycosides. *American journal of human genetics.* 1998 Jan;62(1):27-35.
19. Zhao H, Li R, Wang Q, et al. Maternally inherited aminoglycoside-induced and nonsyndromic deafness is associated with the novel C1494T mutation in the mitochondrial 12S rRNA gene in a large Chinese family. *American journal of human genetics.* 2004 Jan;74(1):139-52.

20. Goto Y, Nonaka I, Horai S. A new mtDNA mutation associated with mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS). *Biochimica et biophysica acta*. 1991 Oct 21;1097(3):238-40.
21. Kirby DM, McFarland R, Ohtake A, et al. Mutations of the mitochondrial ND1 gene as a cause of MELAS. *Journal of medical genetics*. 2004 Oct;41(10):784-9.
22. Ghezzi D, Sevrioukova I, Invernizzi F, et al. Severe X-linked mitochondrial encephalomyopathy associated with a mutation in apoptosis-inducing factor. *American journal of human genetics*. 2010 Apr 9;86(4):639-49.
23. Berger I, Ben-Neriah Z, Dor-Wolman T, et al. Early prenatal ventriculomegaly due to an AIFM1 mutation identified by linkage analysis and whole exome sequencing. *Molecular genetics and metabolism*. 2011 Dec;104(4):517-20.
24. Prakash SK, Cormier TA, McCall AE, et al. Loss of holocytochrome c-type synthetase causes the male lethality of X-linked dominant microphthalmia with linear skin defects (MLS) syndrome. *Human molecular genetics*. 2002 Dec 1;11(25):3237-48.
25. Wimplinger I, Morleo M, Rosenberger G, et al. Mutations of the mitochondrial holocytochrome c-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. *American journal of human genetics*. 2006 Nov;79(5):878-89.
26. Wimplinger I, Shaw GM, Kutsche K. HCCS loss-of-function missense mutation in a female with bilateral microphthalmia and sclerocornea: a novel gene for severe ocular malformations? *Molecular vision*. 2007;13:1475-82.
27. Potluri P, Davila A, Ruiz-Pesini E, et al. A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. *Molecular genetics and metabolism*. 2009 Apr;96(4):189-95.
28. Mayr JA, Bodamer O, Haack TB, et al. Heterozygous mutation in the X chromosomal NDUFA1 gene in a girl with complex I deficiency. *Molecular genetics and metabolism*. 2011 Aug;103(4):358-61.
29. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>) [January 2012].
30. McLaughlin HM, Sakaguchi R, Liu C, et al. Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. *American journal of human genetics*. 2010 Oct 8;87(4):560-6.
31. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010 Apr;7(4):248-9.
32. Geraghty M, Nowaczyk M, Humphreys P, et al. Exome sequencing reveals that mutations in the genes encoding aminoacyl tRNA synthetases (ARS) cause a variety of clinical syndromes. 62nd Annual Meeting of the American Society of Human Genetics (Abstract #701F); November 7, 2012; San Francisco, CA.
33. Yamashita S, Miyake N, Matsumoto N, et al. Neuropathology of leukoencephalopathy with brainstem and spinal cord involvement and high lactate caused by a homozygous mutation of DARS2. *Brain & development*. 2012 Jun 5.
34. Isohanni P, Linnankivi T, Buzkova J, et al. DARS2 mutations in mitochondrial leukoencephalopathy and multiple sclerosis. *Journal of medical genetics*. 2010 Jan;47(1):66-70.
35. Ruiz-Pesini E, Lott MT, Procaccio V, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res*. 2007 Jan;35(Database issue):D823-8.
36. Kagawa R, Montgomery MG, Braig K, Leslie AG, Walker JE. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *EMBO J*. 2004 Jul 21;23(14):2734-44.
37. Kabaleeswaran V, Puri N, Walker JE, Leslie AG, Mueller DM. Novel features of the rotary catalytic mechanism revealed in the structure of yeast F1 ATPase. *EMBO J*. 2006 Nov 15;25(22):5433-42.
38. Arsenieva D, Symersky J, Wang Y, Pagadala V, Mueller DM. Crystal structures of mutant forms of the yeast F1 ATPase reveal two modes of uncoupling. *The Journal of biological chemistry*. 2010 Nov 19;285(47):36561-9.
39. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.