UNIVERSITY OF CALIFORNIA,
IRVINE

Assessing the impact of potential alternative splicing on phenotypic differences among patients with mitochondrial complex I deficiency

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Genetic Counseling

by

Bethany Larson Berg

Thesis Committee:
Professor Virginia Kimonis, MD, MRCP, Chair
Professor Moyra Smith, MD, PhD
Professor Klemens Hertel, PhD

2017

DEDICATION

To the families of patients with rare diseases who commit their lives to pursuing answers and providing hope to patients and medical providers alike – to you I am forever thankful and inspired

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There have been several individuals throughout the last two years that have made my graduate experience truly special and worthwhile. The first person I'd like to thank is Pam Flodman. Before I was accepted into this program, Pam kindly offered to meet with me in person and offered me a glimpse into what it would be like to train as a genetic counselor at UC Irvine. From that moment, I grew that much more motivated to apply to this program, largely because I knew I'd have the opportunity to be mentored by Pam who exemplified the intelligent, encouraging, and compassionate genetic counselor that I hoped to become. I'll always be thankful for her willingness to set aside everything for her students and for being such a foundation for this program. I'd also like to thank Katie Singh for pushing us to work hard in our studies and to go the extra mile for our patients. Her knowledge of genetic diseases is incredible (!) but even more amazing is her obvious compassion for her patients – this specific passion is something I will always aim to model as I begin my career outside of UC Irvine.

I will always be thankful to have been introduced to a project that, though small, had a connection to my previous research studies at UC San Diego, research that I didn't think would ever have direct clinical application. My committee members have been very supportive and have certainly helped me learn an immense amount throughout the course of my project. Dr. Kimonis has not only shown me the power of collaboration in furthering research but she also demonstrates a gentle and caring demeanor with her patients that I will always appreciate. Dr. Smith's passion for research in genetics is inspiring, to say the least. I'll always be thankful that she was there to share in my excitement of results and to patiently show me how to think about bioinformatics data from different perspectives. Dr. Hertel has been kind, insightful, and supportive when it came to interpreting my results or re-thinking experimental design.
I also want to especially thank Dr. Lan Weiss who was a tremendous support for this project, helping culture patient fibroblasts and extract RNA while also helping me troubleshoot all of the many PCR and gel electrophoresis experiments. She also has a very kind and friendly spirit that makes me thankful for all of the time that I had to spend with her in the lab. In addition, I want to say a special thanks to Kathy Hall for introducing me to this project and for her assistance in establishing my thesis hypothesis. She has been an integral part to my training not only by helping me with my thesis but also by training me as an intern. I learned an incredible amount from her about the details that go into running a genetics clinic, something I know I will take with me in my career wherever I go.

The families whom raise money for our *NUBPL* research project will always be an inspiration to me. How they dedicate their lives to providing day-to-day care for their children and pursuing research in new discoveries and treatments is something that will always motivate me as a healthcare provider.

To all of my supervisors at UC Irvine and Kaiser Permanente, thank you for providing me with positive support while also challenging me and showing me how to improve upon my weaknesses. Your guidance and encouragement to become the best genetic counselor I can be continually reminds me how thankful I am to be in such a field and to have graduated from such a highly respected program. Additionally, I'd like to thank all of the office staff at the Tower and

at each of the UCI and Long Beach clinics for patiently working with us students in our training!

A special thank you goes out to my classmates. I knew that going into such a small field meant that I would be able to become close with a small group of classmates. However, I never imagined that my classmates would mean so much to me by accepting all of my flaws, laughing with me, and supporting me when I needed it most. I was pretty lucky to have been accepted into a class of such incredible women and I can't wait to keep in touch with all of you as we dive into our careers. You all are incredibly strong, smart, and compassionate and are truly going to be some of the best genetic counselors I know!

I lastly want to thank my family and friends. My parents have been my biggest support throughout my life, exemplifying how to always strive for your best while making sure to leave time to help others along the way. And to my husband, Chris, none of this would have happened had you not crushed the doubt I had in myself and pushed me to pursue this career. I couldn't be thankful enough to have you as my best friend.

ABSTRACT OF THE THESIS

Assessing the impact of potential alternative splicing on phenotypic differences among patients with mitochondrial complex I deficiency

By

Bethany Larson Berg

Master of Science in Genetic Counseling

University of California, Irvine, 2017

Professor Virginia Kimonis, MD, MRCP, Chair

With the onset of clinical whole exome sequencing, novel gene mutations have been identified in numerous patients with previously undiagnosed diseases, including those with mitochondrial disease of unknown etiology. Complex I deficiency is one of the most common mitochondrial disorders and is both clinically and genetically heterogeneous. Many mutations within the nuclear and mitochondrial genomes have been thoroughly studied for their impact on complex I function based on deleterious alterations in protein function and prevalence among affected individuals. However, many mutations, including splicing mutations, require further analyses of mRNA transcripts in order to understand the impact on disease pathology and potential phenotypic differences.

This study was designed to assess the potential affect of alternative splicing among three patients (two siblings and one unrelated patient) with complex I deficiency and compound heterozygous mutations in *NUBPL*. Each of them share the same branch-site mutation in *NUBPL*, c.815-27T>C, but have differences in disease severity. Using primers designed to separately encompass this mutation as well as a different splice mutation in the unrelated patient (c.693+1 G>A), reverse transcription polymerase chain reaction (RT-PCR) was performed on

RNA from patient fibroblasts. Gel electrophoresis demonstrated alternative splicing occurring in all three patients following RT-PCR with the branch site mutation primers. In addition, whole exome data from the two siblings was re-analyzed and potential modifier genes affecting *NUBPL* were studied. Our results support the importance of investigating potentially pathogenic mutations using RNA studies given that alternative splicing can significantly impact phenotypic expression across different tissues and different individuals.

1 Introduction

1.1 Rare diseases

Across the globe, over 7,000 rare diseases have been reported, collectively affecting 30 million people in the United States alone (NIH, U.S. National Library of Medicine). Despite the fact that an individual rare disease is defined as affecting 200,000 people or less globally, the number of individuals diagnosed with rare diseases as a whole is increasing every year. Many of these rare diseases are caused by genetic alterations, and with the advancement of genetic sequencing technologies such as whole exome sequencing, families are finding the diagnostic odyssey to be much shorter than in previous decades. The onset of clinical whole exome sequencing has given clinicians the power to use gene sequencing of an individual's exome to identify genetic causes for previously undiagnosed patients. For many patients, this has meant having a direct answer as to the genetic cause of their disease, educating the patient and families with reasoning behind their clinical signs and symptoms and providing them with more access to appropriate medical care. In addition, families potentially have more access to clinical trials for drug therapies as well as support groups that represent individuals with any rare disease, such as Global Genes, The Rare Disease Clinical Research Network, and the National Organization for Rare Disorders (NORD). For many families, there is a sense of empowerment to connect with academic researchers to increase the repertoire of information surrounding their genetic alteration and the biological impact of the associated gene(s). Given that rare diseases are affecting millions of people world-wide and the identification of the genes involved in a majority of them is rapidly increasing, there is an increased need for not only the physiological role of individual genes, but also the role of modifier genes and alterations at the level of RNA and proteins. By contributing to this research in rare diseases, there is not only increased knowledge

behind one particular gene and one particular disease, but there is an increase in the comprehensive understanding of how our genes play a role in human development and disease pathology as a whole.

1.2 Previous work on a rare disease in two siblings

In previous whole exome sequencing studies by Hall, et al. (2014), two siblings were found to have a rare cause of complex I deficiency due to compound heterozygous mutations in the gene *NUBPL* (Hall, et al. 2014). Complex I deficiency, which is the most frequent mitochondrial disorder presenting in childhood, is both clinically and genetically heterogeneous with symptoms ranging from fatal neonatal disease to adult-onset progressive neurodegenerative disorders. In these studies, two female siblings were found to have cerebellar atrophy, ataxia, developmental delay, cerebellar hypoplasia, and pontine hypoplasia, where one sibling was found to have a more severe phenotype than the other. Their 12-year journey for a diagnosis involved numerous laboratory analyses, imaging studies, as well as mitochondrial and nuclear DNA sequencing panels for known ataxia conditions. Ultimately, with the use of clinical whole exome sequencing, each sibling was found to have one allele with a pathogenic branch-site mutation (c.815-27T>C) in the *NUBPL* gene and one allele with a pathogenic missense mutation (c.311T>C) *in trans*, giving them a molecular diagnosis of complex I deficiency. A brief review of this investigation, including clinical findings behind identifying these two variants are summarized below.

1.2.1 Summary of clinical cases 1 and 2 (Siblings A and B)

Per summary of case reports by Hall et al (2014), Sibling A (Patient 1) failed to reach developmental milestones starting at the age of 3 months. At 3 months of age, she started having

tremulousness and increased rigidity. At 9 months of age, she was unable to sit with support, had poor head control, was unable to crawl, and could not pull to stand. She stood independently at 30 months and walked at 4 years. Her first words were spoken at 30 months and, at the age of 15 years, she was making five-word sentences. At this time, evaluations included analysis of serum amino acids, urine organic acids, and a lysosomal storage screen, all of which were normal. She also had a magnetic resonance imaging (MRI), where findings showed diffuse abnormality of cerebellar cortices bilaterally with normal white matter and normal supratentorial appearance. At this time, the acute disseminated encephalomyelitis (ADEM) secondary to previous vascular infection was the primary differential. A repeat MRI at 11 months showed diffuse T2 prolongation throughout the cerebellar hemispheres with irregular enhancement within the cerebellum that was not considered a typical appearance of a demyelinating disease, and a metabolic abnormality was included as a differential diagnosis. Physical examination at this age noted bobbing head movements and unusual eye movements, and additional MRI examination suggested Joubert syndrome, however, a unique molar-tooth sign on cranial MRI was not noted.

Another brain MRI was performed at 3 years of age indicating cerebral atrophy where most of the cerebellar white matter had been lost with stable diffuse atrophy (Figure 1, top panel). At 6 years of age, Sibling A began taking Carnitor and coenzyme Q10 for presumed mitochondrial dysfunction. By the age of 10, the parents reported that the nystagmus and ataxia of the hand was slightly improved, but there were otherwise no significant improvements.

At the age of 11, Sibling A was further evaluated by Genetics and began a series of new evaluations that included a microarray, testing of 7-dehydrocholesterol, very long chain fatty acids, phytanic and pipecolic acids, plasmologens, congenital disorders of glycosylation testing, and MECP2 sequencing for Rett syndrome. All of these tests yielded normal results. In addition,

a Complete Ataxia Evaluation Panel was performed by Athena Laboratories, which did not identify any pathogenic mutations associated with known ataxia-associated genes. A repeat MRI at age 12 y indicated that there was still evidence of cerebral atrophy (Figure 1, top panel)

Sibling B (Patient 2) is the younger sister of Sibling A. Sibling B's medical history is comparable to her sister's, but her phenotype of ataxia, developmental delay, cerebellar hypoplasia, and pontine hypoplasia was and remains to be notably milder. Developmentally, she rolled at 5 months and sat at 15 months. At 3 months of age, Sibling B developed jerky eye movements and heading shaking. She was evaluated at 10 months of age for cerebellar symptoms and was noted to have truncal hypotonia, horizontal nystagmus, a mild hand tremor, side-to-side head shaking, and no dysmorphic features. At 11 months, a brain MRI revealed diffuse increase in T2 signaling, mild diffuse hypoplasia, a mildly prominent fourth ventricle, but a preserved cerebellar peduncles and pons. At 1 year of age, she was started empirically on carnitine and coenzyme Q10 for presumed mitochondrial dysfunction. At 2 years of age (and later at 5 years of age), Sibling B had additional brain MRIs that further indicated cerebral atrophy and partial absence of the inferior portion of the cerebellar vermis (Figure 1, bottom panel).

At the ages of 12 y (Sibling A) and 3 y (Sibling B) respectively, whole exome sequencing was performed by Ambry Genetics on both siblings as well as their parents. At this time, two pathogenic mutations were identified in association with mitochondrial complex I function: a branch-site mutation (c.815-27T>C) and a missense mutation (c.311G>A) in *NUBPL*. Sibling A and B were given a diagnosis of complex I deficiency. They were placed on a mitochondrial cocktail for prevention of worsening disease progression.

Given these results from whole exome sequencing, mitochondrial respiratory chain enzyme analysis (complexes I-IV) of muscle biopsy from Sibling A was performed at Baylor College of Medicine Medical Genetics Laboratories. Interestingly, results revealed increased citrate synthase activity suggesting a possible compensatory response to decreased mitochondrial function, however no deficiencies of respiratory chain activities were detected before or after adjustment for the increased citrate synthase activity.

Per parents' report, there have been significant phenotypic differences between Sibling A and Sibling B in all global areas of development. Currently at the age of 17 y., Sibling A has been "extremely delayed in academics" and has been delayed in understanding of colors, shapes, numbers, and letters. Her ataxia has been much more severe than her younger sisters' and, today, she continues to use a walker for assistance. In comparison, at the age of 7 y., Sibling B has a milder gait disturbance and has been able run and walk much more easily without the use of a walker. She has also been doing better in academics and has been considered "mildly delayed" by her parents, however she is in several regular classes.

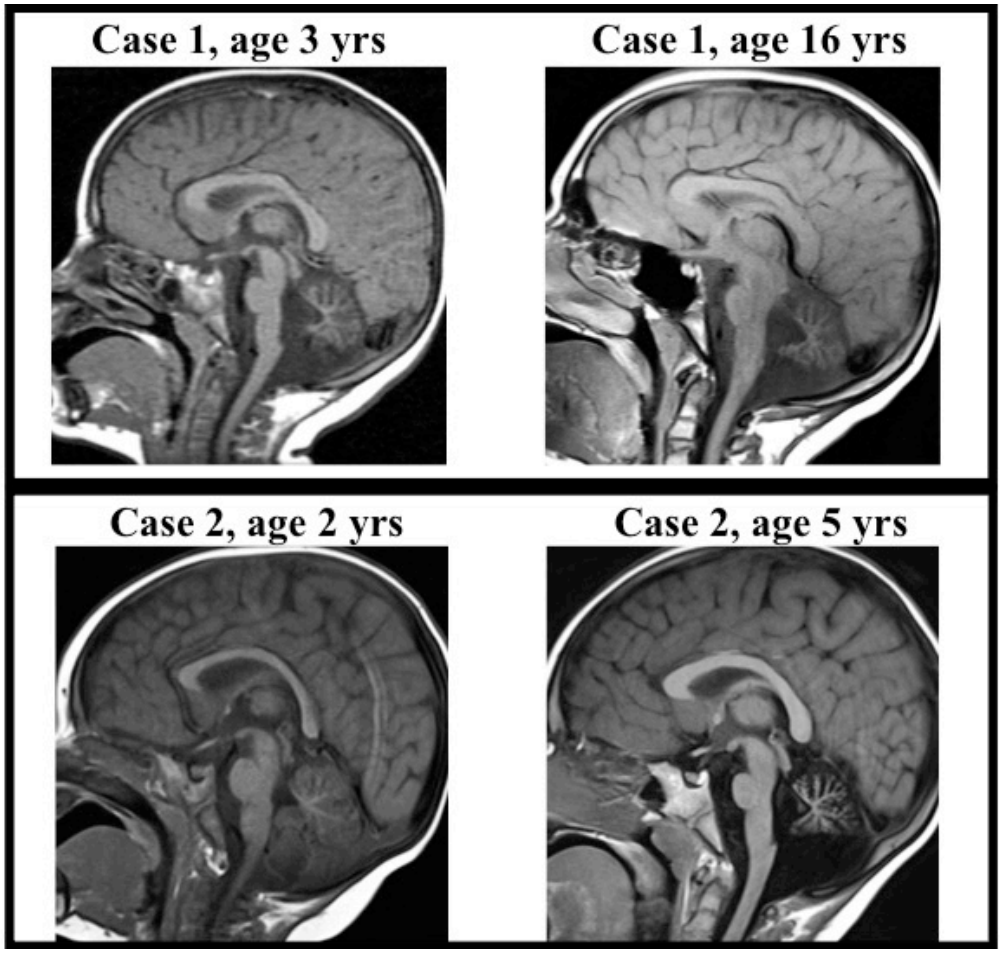

Figure 1: Progressive brain MRI images for Cases 1 and 2.
Top panel: Brain anatomy of Case 1 at ages 3 and 16 years showing cerebral atrophy. Most of the cerebellar white matter has been lost with stable diffuse atrophy. Bottom panel: Case 2 at ages 2 and 5 years also showing cerebral atrophy and partial absence of the inferior portion of the cerebellar vermis.

1.3 Summary of clinical case 3

Case 3 (Patient 3) is a now 5-year old girl with ataxia, developmental delay, and cerebellar abnormalities. Regression in motor skills were noted at age 13-14 months. At this time, she lost ability to stand without support and became clumsier. At 3.5 years of age, she was noted to have hypotonia. She could walk with a walker or support. She has a tremor present with mild dysmetria and some speech articulation difficulties. She had and still currently has normal cognition. Electron transport chain (ETC) analysis on skin fibroblasts was essentially normal.

Neuroimaging showed stable largely symmetric punctate areas of signal abnormality and restricted diffusion of the genu of the corpus callosum. MRI at the age of 3 y indicated loss of cerebellar volume with notable atrophy in the inferior vermis (Figure 2). MR spectroscopy revealed a lactate peak within the cerebellum. She was initially diagnosed with Infantile Neuroaxonal Dystrophy (INAD, MIM 256600) because of her MRI findings; however, genetic testing for mutations in the gene associated with this condition, *PLA2G6*, was negative. Clinical exome sequencing in 2015 was ordered by Dr. Parikh at the Cleveland Clinic and she was found to be a compound heterozygote: c.815-27T>C and c.166G>A (in *cis*) from her mother and c.693+1G>A from her father, a combination previously reported in a case from the US (Kevelam et al. 2013). She was diagnosed with complex I deficiency due to these mutations in *NUBPL* and was placed on a mitochondrial cocktail.

Per parents report in March 2017, she is now able to articulate her thoughts and emotions and remains to have normal cognition. She is improving in her abilities to walk independently, but still requires support. She was enrolled in the EPI-743 drug trial in August 2015, a trial involving a small molecule that is designed to treat patients with mitochondrial respiratory chain

disorders by targeting NADPH quinone oxidoreductase (NQO1). She currently receives occupational therapy and physical therapy at school. She has an individualized education plan (IEP), is fully integrated, and, with assistance, is able to participate in all activities with her peers. She is also able to write her name with little or no assistance.

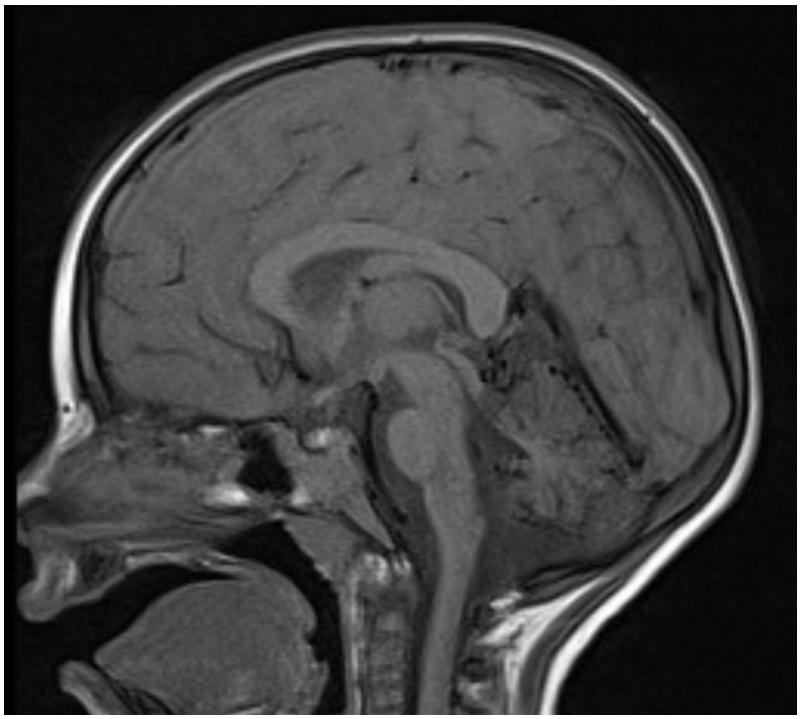

Figure 2. MRI of Patient 3 at age 3 y. MRI shows loss of cerebellar volume with notable atrophy in the inferior vermis.

1.4 Oxidative phosphorylation and complex I deficiency

In order for humans to effectively generate useable energy from proteins and sugars, individual cells utilize the process of oxidative phosphorylation in structures called mitochondria. The unit of energy that is transported in this mechanism is called adenosine triphosphate, or ATP, and is critical for all cellular function including synthesis of proteins, cellular division, and cellular movement (Berg et al 2015). In oxidative phosphorylation, ATP is produced as a result of transferred electrons from two metabolic coenzymes, NADH or $FADH_2$, to O_2 by a series of electron carriers. These electron carriers form the electron transport chain and include four complexes: NADH-Q oxidoreductase (complex I), succinate-CoQ reductase (complex II), cytochrome bc_1 (complex III), and cytochrome c oxidase (complex IV), all of which are in the mitochondrial membrane. Through these electron carriers, an electron-motive force is converted into a protein-motive force by pushing protons out of the mitochondrial matrix to create a transmembrane electrical potential. This potential is then used by ATP synthase (complex V), another mitochondrial membrane protein, to create a phosphoryl transfer potential, or the phosphorylation of ADP (adenosine diphosphate) to ATP, by pumping protons back into the mitochondrial matrix (Figure 3). Overall, oxidative phosphorylation generates a large majority of the 30 ATP molecules that are made when sugar (glucose) is fully oxidized to CO_2 and H_2O.

Genetic defects leading to altered expression of oxidative phosphorylation complexes have been implicated in a myriad of mitochondrial disorders. It is estimated that oxidative phosphorylation disorders affect approximately 1 in 5-10,000 live births, and therefore is one of the most common groups of inborn errors of metabolism. Oxidative phosphorylation disorders are extremely heterogeneous by nature as well as pleiotropic, causing an array of clinical signs

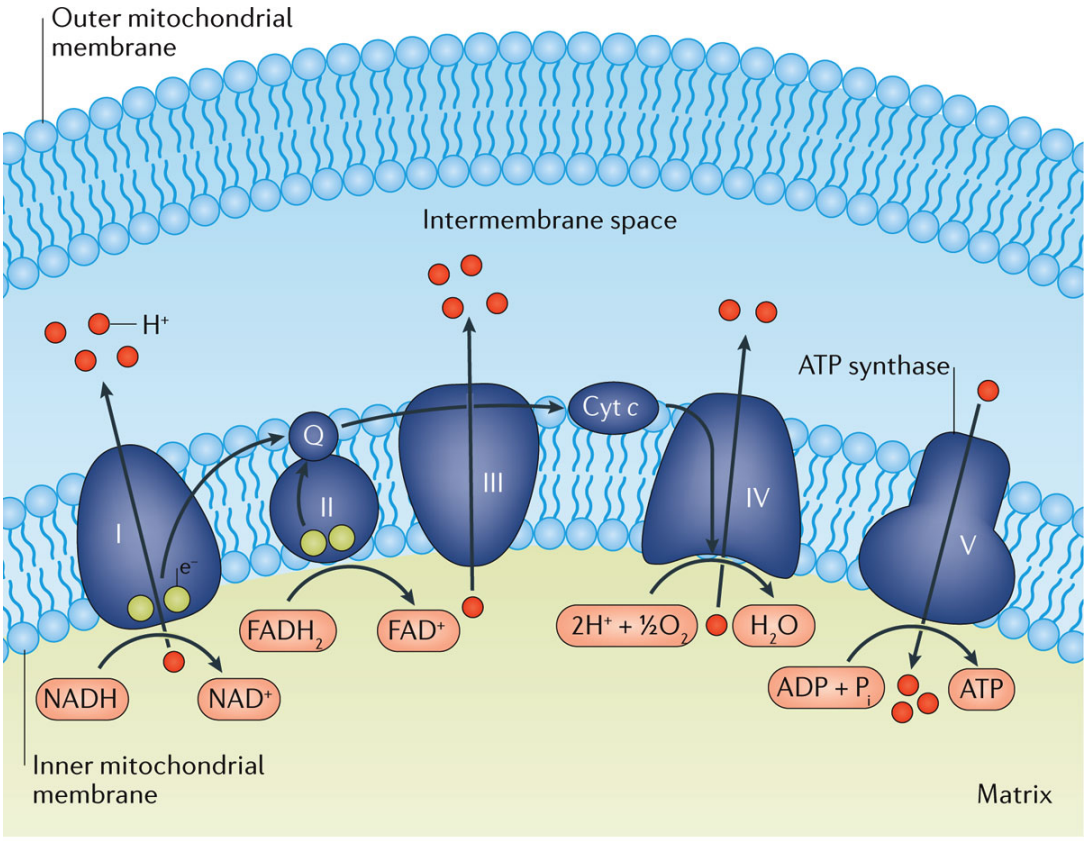

Figure 3. Diagram of oxidative phosphorylation. Oxidative phosphorylation demonstrating a coupled proton (H+) and electron (e^-) transfer across the inner mitochondrial membrane. From: Chow, J. et al. Copyright 2016 by Nature Publishing Group. Used by Permission of the publisher.

and symptoms that typically involve disruption of neurological, neuromuscular, cardiac, and/or endocrine function. Patients can present with strokes, seizures, cardiomyopathies, kidney malfunction, muscle failure, lactic acidosis, and ophthalmologic disorders (Smeitink, et al 2004). For those suspected of having an oxidative phosphorylation disorder, analysis of enzymatic activity is often one of the initial tests performed for diagnosis. Skeletal muscle biopsy is usually preferred, however cultured fibroblasts can also be used. Given analysis can be performed on either tissue, it is important to note that not all enzyme deficiencies that are observed in skeletal muscle can be observed in skin fibroblasts. In addition, almost all patients present with some residual oxidative phosphorylation enzyme complex activity (Van den heuval, et al. 2001). For patients where clinical presentation is indicative of a mitochondrial disorder but observed enzyme activity levels are within normal limits, molecular genetic analysis can also be performed.

In considering the genetic causes of these mitochondrial disorders, both mitochondrial DNA and nuclear DNA mutations are implicated in causing disease. Given 13 of the approximately 80 oxidative phosphorylation subunits are encoded by mitochondrial DNA, many of these oxidative phosphorylation disorders are maternally inherited, meaning a pathogenic mutation found in the mitochondrial DNA can only be inherited from the mother and not from the father (Distelmaier, et al. 2009). For these diseases, heteroplasmy, or the presence of more than one type of mitochondrial genome within an individual cell, can lead to variability in the severity of disease even among individuals in a family. At a cellular level, this clinical variability is a result of replicative segregation, which occurs when a heteroplasmic cell divides and the number of mutant mitochondrial genomes segregate arbitrarily leading to chance proportion of mutant organelles in each progeny cell (Pierce, 2008). In addition to this variability from

replicative segregation, threshold effects impact clinical expression when the load of mutant mitochondrial genomes within each cell impairs oxidative phosphorylation up to a certain threshold level. Heteroplasmy can also impact the severity of disease pathology within different tissues due to tissue-specific segregation of heteroplasmy (Smeitink et al., 2006). Two mitochondrial disorders that are maternally inherited and exhibit variable clinical expression due to heteroplasmy include mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) and myoclonic epilepsy with ragged red fibers (MERRF) (Smeitink et al., 2004). For patients with these disorders, clinical presentation often correlates strongly with a clinical diagnosis. Biochemical enzyme analysis from a muscle biopsy can also be performed for these disorders, however tissue-specific expression of complex activity may lead to normal enzyme activity at which point gene analysis via sequencing and/or gene deletion/duplication analysis can prove beneficial for a diagnosis.

Alternatively, oxidative phosphorylation disorders can also manifest via pathogenic mutations in the nuclear genome. In these diseases, nuclear DNA mutations can leading to defects in protein-encoding subunit genes, genes encoding factors affecting mitochondrial DNA maintenance, mitochondrial protein synthesis (ie. ribosomes), genes modifying mitochondrial tRNA post-transcriptionally, and genes encoding biosynthetic enzymes for lipids (Smeitink, JA 2006). Enzyme analysis of skeletal muscle or skin fibroblasts The majority of these diseases are typically inherited in an autosomal recessive manner (although dominant traits and X-linked inheritance have also been reported) rather than maternal inheritance; however, similar to mitochondrial-encoded mutations, nuclear-encoded mutations can also be tissue-specific due to the variable metabolic thresholds for each of the oxidative phosphorylation complexes in each tissue (Rossignol et al., 2000; Ruitenbeek et al., 1996). Ultimately, when determining possible

diagnoses of oxidative phosphorylation disorders that are due to mutations in either the mitochondrial or nuclear genome, analysis of inheritance patterns as well as gene analysis of both of these genomes are proven beneficial in diagnosing these complex mitochondrial disorders.

Oxidative phosphorylation disorders can manifest by defects in any of the five large multi-subunit complexes embedded in the inner mitochondrial membrane. For instance, mutations in the *ATP6* mitochondrial gene that encodes for complex V have been associated with a syndrome of neuropathy, ataxia and retinitis pigmentosa (NARP) (Holt, et al., 1990) as well as Leigh syndrome, a disease characterized by distinct brain lesions on MRI, progressive loss of mental and movement abilities, and eventually death within a few years of life due to respiratory failure (Tatuch, et al., 1992). Additionally, mutations in the *SDHA* gene encoding for complex II have also been associated with autosomal recessive Leigh syndrome. Interestingly, one family was found to have only one allele with a mutation in this gene, resulting in a 50% decrease in succinate dehydrogenase activity and adult-onset optic atrophy, ataxia, and myopathy. This finding may imply that late-onset neurodegenerative disorders due to partial electron transport chain deficiency may be inherited as dominant traits (Shoubridge EA, 2001). In order to understand oxidative phosphorylation deficiencies and their association with encephalomyopathies, analysis of how nuclear and mitochondrial-encoded gene mutations affect assembly and structure of each of these electron transport complexes is imperative.

Among all five of the complexes involved in oxidative phosphorylation, complex I, or NADH dehydrogenase, is the most frequently encountered defect of the mitochondrial energy metabolism and is also the largest complex, consisting of 45 subunits, 38 encoded by nuclear DNA and 7 encoded by mitochondrial DNA (Smeitink et al., 2006; Carroll et al., 2006). complex

I, contains a long membrane domain and a hydrophilic domain that contains all of the known redox centers as well as the NADH binding site (Vinothkumar et al., 2014). The main function of complex I involves accepting electrons from NADH and transferring them to ubiquinone, generating a negative membrane potential used to pump protons across the inner mitochondrial membrane. During this electron transfer process at complex I, electrons may also 'leak' to produce superoxides within the cell. Genetic alterations of complex I can therefore dramatically impact cell metabolism and regulated production of reactive oxygen species (ROS), where certain cell lines may be more or less impacted by ROS depending on a higher or lower threshold for oxidative stress (Distelmaier et al. 2009). Complex I deficiencies can be inherited in an autosomal recessive, autosomal dominant, X-linked, or maternal manner and are the most common of the electron transport chain defects, where mutations in genes encoding for complex I are associated with severe multisystem disorders that typically manifest at young ages. Nearly half of patients with deficiencies in complex I present with Leigh disease; other individuals found to have complex I deficiency are found to have brainstem lesions, respiratory abnormalities, muscular hypotonia, failure to thrive, seizures, and lactic academia. Mutations in both the mitochondrial and nuclear genomes have been reportedly associated with complex I deficiency, where nuclear-encoded mutations have been described as having a more severe impact on disease progression than mitochondrial-encoded mutations. Tissue variability of complex I activity has also been reported, where a patient with a mutation in a gene encoding for one of the complex I subunits was found to have 39% CI activity in muscle tissue, 69% activity in fibroblasts, and completely absent activity in heart muscle tissue (Distelmaier et al. 2009). Diagnosis of complex I deficiency can be especially challenging for these reasons given that post-mitotic, oxidative phosphorylation-dependent cell populations such as those in the brain,

nerves, and muscles are often difficult to obtain on a research basis. Though skin fibroblasts are not as vulnerable to complex I defects as brain, nerve, and muscle tissue, they have proven to be affected by complex I defects (Distelmaier et al., 2009). These tissue-specific differences in addition variable genetic backgrounds have lead researchers and clinicians to also utilize molecular genetic testing, especially with the onset of clinical whole exome sequencing. Once a diagnosis is obtained, treatment of complex I deficiency often consists of a variety of vitamins and co-factors, including coenzyme Q, riboflavin, vitamin B1, B12, C, E, K, L-carnitine, and dichloroacetate, however treatment plans are often challenging given clinical responses have been variable among affected patients (Parikh et al. 2009). As a whole, complex I deficiency has proven to be a heterogeneous condition with consistent challenges for diagnosis and management that clearly suggest a need for clinical trials. However, because of the rarity of the condition and the time often required for diagnosis, clinical trials have proven difficult to undertake.

1.5 *NUBPL*

As a genetically heterogeneous disease, complex I deficiency has been found to be associated with disease-causing mutations in several different genes that impact structural and mechanistic integrity of the membrane protein. Pathogenic mutations have been reported in all 7 mitochondrial DNA-encoded subunits and in at least 22 of the 38 nuclear-encoded subunits (Tucker et al., 2011). While some of these genes are associated with subunit formation of complex I, additional genes that affect assembly factors for complex I have also been reported. In fact, until 2010, gene sequencing of the known set of 44 genes encoding structural subunits of complex I only provided a genetic explanation to approximately half of individuals affected by complex I deficiency (Wydro et al., 2013). Given this, it was suggested that there were likely

additional assembly and functional factors that were required for essential function of complex I. One example of these assembly factor genes is called Nucleotide Binding Protein Like, or *NUBPL* (also referred to as *HuInd1*), and is responsible for assembly of one or more of the eight iron-sulfur (Fe-S) clusters that mediate electron transfer in complex I. The NUBPL protein contains a conserved nucleotide binding domain that is characteristic of the P-loop NTPase subclass and also contains a C-terminal domain that includes a conserved CXXC motif. This motif is hypothesized as a critical sequence for transiently binding an Fe-S cluster (Netz et al., 2007). The yeast homolog of *NUBPL*, known as *IND1*, was first described in *Yarrowia lipolytica* given its sequence similarities to scaffold proteins that are involved in cytosolic Fe-S cluster (Nbp35 and Cfd1) and its *in vitro* binding of labile Fe-S clusters (Bych et al., 2008). Based on these results, knockdown of *NUBPL* using RNA interference technology was performed in HeLa cells and results indicated a decrease in complex I activity in addition to accumulation of other abnormal sub-complexes (Sheftel et al., 2009). The decrease in complex I activity is proposed to be due to the lack of Fe-S clusters within the complex I subunits, and in fact, it is well established that proteins without their corresponding metallocofactors are not able to properly fold or assemble and, hence, are sensitive to intracellular protein breakdown (Wittung-Stafshede, P., 2002). Without the expression of NUBPL (or huInd1), one or more of the Fe-S clusters are unable to insert into the Fe-S cluster-coordinating subunits, thereby leading to failed assembly of all or parts of the peripheral arm of complex I and thus leading to complex I activity defects. These in vitro studies ultimately helped formulate the hypothesis that many patients affected by complex I deficiency may be harboring mutations in genes that encode for assembly factors such as NUBPL.

The *NUBPL* gene that encodes for NUBPL protein is located at chromosome 14 on the q arm at position 12 (14q12). The molecular location using the GRCh38 reference sequence is from base pairs 31,561,385 to 31,861,293 and the gene contains 11 exons (Calvo et al., 2010). Given the discovery of NUBPL's function as well as the onset of clinical exome sequencing within the last five years, several individuals with previously undiagnosed mitochondrial conditions were found to have pathogenic mutations in *NUBPL*. In 2010, Calvo, et al identified a patient with complex I deficiency who was a compound heterozygote for 2 alleles of the *NUBPL* gene, one carrying a c.166G>A transition in exon 2 and one carrying a c.815-27T-C transition. This patient presented at 2 years of age with developmental delay and eventually developed myopathy, nystagmus, ataxia, upper motor neuron signs, spasticity, and speech problems by the age of 8. In addition, a brain MRI of this patient showed leukodystrophy with involvement of the cerebellar cortex and deep white matter and fibroblasts showing reduced complex I activity. Another separate patient with complex I deficiency who was first reported by Calvo et al in 2010 was later found to be have inherited the c.815-27T>C branch point mutation on one allele and a heterozygous deletion of exons 1–4 and duplication of exon 7 on the other allele as a result of a translocation (Tucker et al., 2011). Furthermore, in 2013, Kevelam et al. identified 6 patients with MRI findings of cerebellar hypoplasia who were found to be compound heterozygotes for various missense, splice-site, and inversion pathogenic mutations in *NUBPL* who were later identified with complex I deficiency ranging from 23% to87%. As a whole, the combination of these clinical findings with the use of Next Generation Sequencing (NGS) further support the critical function of NUBPL in complex I activity and contribute towards the diagnosis of complex I deficiency among individuals who are affected with mitochondrial disease.

1.6 Alternative splicing

With the completion of the Human Genome Project in 2003, new insights were gathered with regards to the concept of 'one gene' to 'one protein' and the increased complexity of the proteome relative to the genome. Given that humans have roughly 90,000 proteins made from approximately 25,000 genes, the concept of alternative splicing was introduced to assist in explaining the existence of multiple protein products (Brett et al., 2002). Alternative splicing was initially described in the 1970s by Walter Gilbert who hypothesized that exons could be spliced together in different arrangements to give rise to diverse mRNA isoforms (Modrek et al., 2002). At that time, it was proposed that approximately 5% of genes undergo alternative splicing, however recent studies predict that alternative splicing events likely occur in closer to 90% of human genes (Wang et al., 2008). Single gene and genome-wide- studies have also shed light on how alternative splicing patterns can be uniquely maintained in specific tissues. In fact, analysis of human tissue and cell line transcriptomes have suggested that alternative splicing patterns are strongly correlated across tissues whereas variation among individuals were two to three fold less common (Wang et al., 2008). As a whole, the growing research on alternative splicing of human genes has proven critical for understanding how different exon patterns contribute to diversity of cells, tissues, and organs as well as to diversity in disease pathology across individuals with similar diagnoses.

1.6.1 Splicing mechanism

Alternative splicing events involve differentially regulated events across tissues and varying points in development as well as among individuals and populations. The mechanism that determines which isoforms will be preferentially spliced likely depends on the varying types of expression as well as recruitment of specific splicing factors to the alternatively spliced

transcript (Berg et al., 2015). In order to form the final mRNA strand, introns must be excised and exons must be connected together in an intricately sensitive process. For instance, a single nucleotide insertion or deletion could alter the reading frame at the 3' end and would thus give an entirely different amino acid sequence from the intended splice site sequence. In eukaryotes, there is a unique sequence motif that denotes a splice location, where the intron begins with a GU and ends with an AG. At the 3' end of an intron, the sequence consists of a polypyrimadine tract (10 U or C) followed by a base then by CAG. Within introns is an internal site called the branch point, which is located between 20 and 50 nucleotides upstream of the 3' splice site. These three sites - the 5' site, 3' site, and branch point - are fundamental for defining where splicing occurs. Mutations that occur within these sites can lead to abnormal splicing and, in most cases, contribute to disease pathology. Sequences around these sites have been shown to play important roles in designating alternative splicing, and knowledge about the effects of these specific sequences are now more frequently studied to understand alternative protein products.

Biochemical studies have revealed the complexity of RNA cleavage and ligation involved in intron removal, which occur through specialized ribonucleprotein (RNP) machinery called the spliceosome. The splicing mechanism is a complicated process involving two specific transesterification reactions, or the transfer of a hydroxyl group with an ester. The first transesterification reaction occurs at the branch site forming a lariat intermediate while the second transesterification links the 5' end of the upstream exon with the 3' end of the downstream exon, as seen in Figure 4 (Berg et al., 2015). These sites at which splicing takes place is critical for processing pre-mRNA, where alterations in these sequences can lead to different mature mRNA transcripts depending on trans-factors which can vary by cell type.

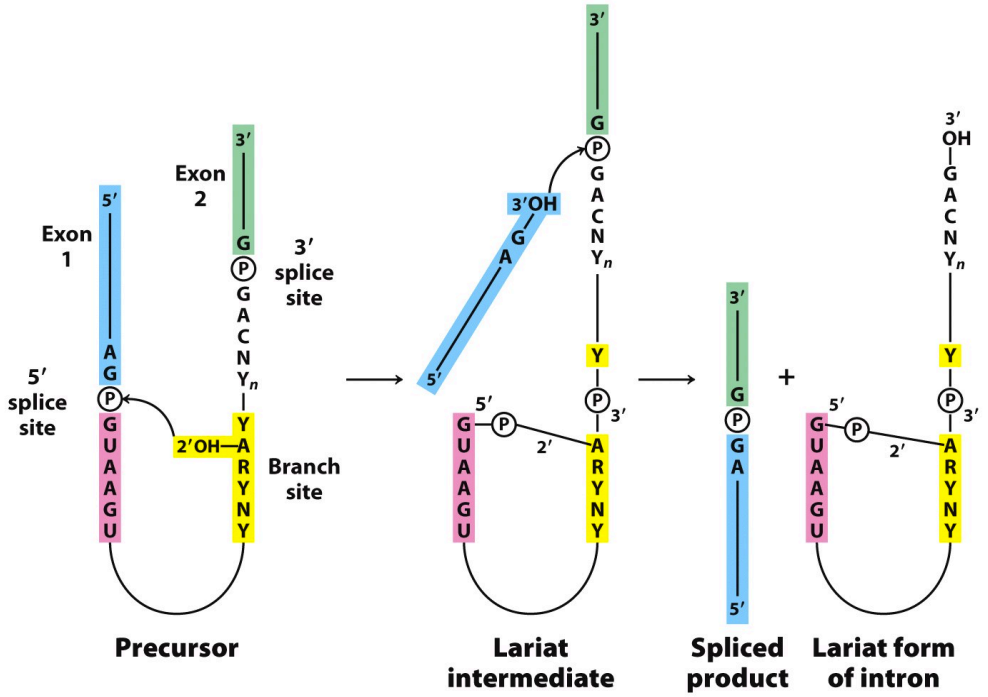

Figure 29.35
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company

Figure 4. Mechanism of splicing. Splicing mechanism of mRNA precursors demonstrating spliced exon products and excised intronic fragments forming a lariat. From: Biochemistry 8E, by Jeremy Berg, et al, Copyright 2015 by W.H. Freeman and Company. Used by Permission of the publisher.

Further understanding of splice site selection is imperative for understanding how the proteome is generated from the human genome.

1.6.2 Human diseases attributed to defects in alternative splicing

There are a myriad of genetic diseases caused by splice site mutations. Splice site mutations have been reported in the *CFTR* gene, which is associated with the mucus and sweat gland disorder known as cystic fibrosis, while splice site mutations in the *BRCA1* gene have been linked to hereditary breast and ovarian cancer. Another example of a disease resulting from splicing abnormalities is beta thalassemia, where RNA splicing mutations are the most common type of mutations found in affected individuals (Nussbaum et al., 2007). In describing different mutations that affect alternative splicing in the *HBB* gene, Nussbaum et al discussed three groups of splice defects that differ depending on their intronic or exonic location, each of which can be found in other disease-associated genes as well. The first group of splicing defects are splice junction mutations which include mutations at the 5' donor or 3' acceptor sites or in the consensus sequences surrounding these donor or acceptor sites. If the normal acceptor site is inactivated, then other acceptor-like sequences, termed cryptic splice sites, may elicit a splicing reaction. These cryptic sites can be intronic or exonic. The second group of splicing defects includes mutations in intronic cryptic sites that make the site more competitive for splicing in comparison to the normal splice site. These cryptic splice site mutations are often termed "leaky" since the normal splice site is still active and produces normal gene product, thus creating a less severe phenotype. The third group of mutations consists of mutations in the open reading frame that activate a cryptic splice site in an exon without altering the encoded amino acid (ie, synonymous mutations that become cryptic splice sites). Given the variety of mutations that can affect the alternative splicing of pre-mRNA to generate aberrant isoforms, continual

investigation of both exonic and intronic sequences alterations affect splicing is critical for gaining insight on specific genetic bases of disease pathology.

1.7 Previous studies on a branch-site mutation in *NUBPL*

As previously discussed, complex I deficiency is considered the most common oxidative phosphorylation disorder, being both phenotypically and genetically heterogeneous. In fact, there are mutations in 29 genes that have been reported to cause complex I deficiency (Tucker et al., 2011). Many of these genes directly express the 45 subunits of the complex, however, other genes have been found to express proteins that are involved in assembly and stability of complex I function, including *NUBPL*. In recent years, the NUBPL protein has been described as a critical protein required for assembly of Fe-S clusters within the complex with sequence similarity to other cytosolic iron-sulfur assembly proteins. The first patient to be reported with complex I deficiency due to mutations in *NUBPL* was initially reported to be homozygous for two missense mutations, c.166G>A (Calvo et al., 2010). Given the high degree of conservation as well as absence in healthy individuals, this missense variant was originally classified as "likely pathogenic." After further analysis by Tucker et al using array-based analysis and cDNA analysis, this same individual was found to also be a compound heterozygote for a complex gene rearrangement as well as a branch site mutation, c.815-27T>C (Tucker et al., 2012). Interestingly, this branch site mutation was found to be much more common than the c.166G>A variant where one of 60 controls are heterozygous for the c.815-27T>C variant, demonstrating possibly one of the most common autosomal recessive mutations among oxidative phosphorylation disorders. Interestingly, per the ExAC database, this mutation occurs at an allele frequency of 0.003. In addition, six reports of c.815-27T>C homozygotes have been reported in

the ExAc database without report of mitochondrial disease in the literature, suggesting the variable pathogenicity of this variant that may be dependent on other genetic and/or environmental factors.

Given mutations involved in aberrant splicing can impact disease pathology, RNA studies can help both identify intronic variants and assess whether these variants impact splicing or not. In these studies by Tucker et al, functional follow-up studies were used to further elucidate the effects of the branch site mutation. Using the patient and control fibroblasts, RT-PCR was performed and products were separated on a gel. While one band was primarily observed for control fibroblasts, three bands were observed for the patient harboring the c.815-27>C variant, supporting the fact that this mutation is causative for abnormal splicing (Figure 5). Sequencing of these three bands revealed three transcripts produced from aberrant splicing. The highest molecular band indicated a cryptic acceptor site with a higher acceptor site score than the wild type score (84.84 compared to 78.54). Given the c.815-27T>C mutation lowers the acceptor site score from 79.39 to 70.36, it may be that the alternative cryptic acceptor site is only utilized when the branch site is altered. In this instance, a portion of intron 9 is included in the alternative mRNA transcript, causing a frameshift mutation that disrupts protein expression. Sequencing of the lowest molecular weight band indicated exon 10 skipping, generating another individual transcript with a separate frameshift mutation. In addition to sequence analysis of transcripts, Tucker et al also quantified the degree of full-length transcript in the patient fibroblasts compared to those of control fibroblasts and fibroblasts from a heterozygous control who harbored the c.815-27T>C mutation but not the patients other mutation (c.166G>A). Results revealed that the heterozygous control had 74% *NUBPL* expression while the patient had only 15% *NUBPL* expression. These expression levels were also confirmed with Western blot. As a

whole, these RNA studies support the pathogenicity of the branch site mutation, which was initially missed by NGS and by the fact that it is also a much more common mutation. Furthermore, the less common c.166G>A missense mutations were found to produce stable NUBPL protein and restored complex I activity in patient fibroblasts. Given that NGS initially detected only the homozygous missense mutation and missed detection of the branch site mutation, c.815-27T>C, these studies by Tucker et al suggest the importance of including follow-up functional studies in order to best classify possibly pathogenic mutations.

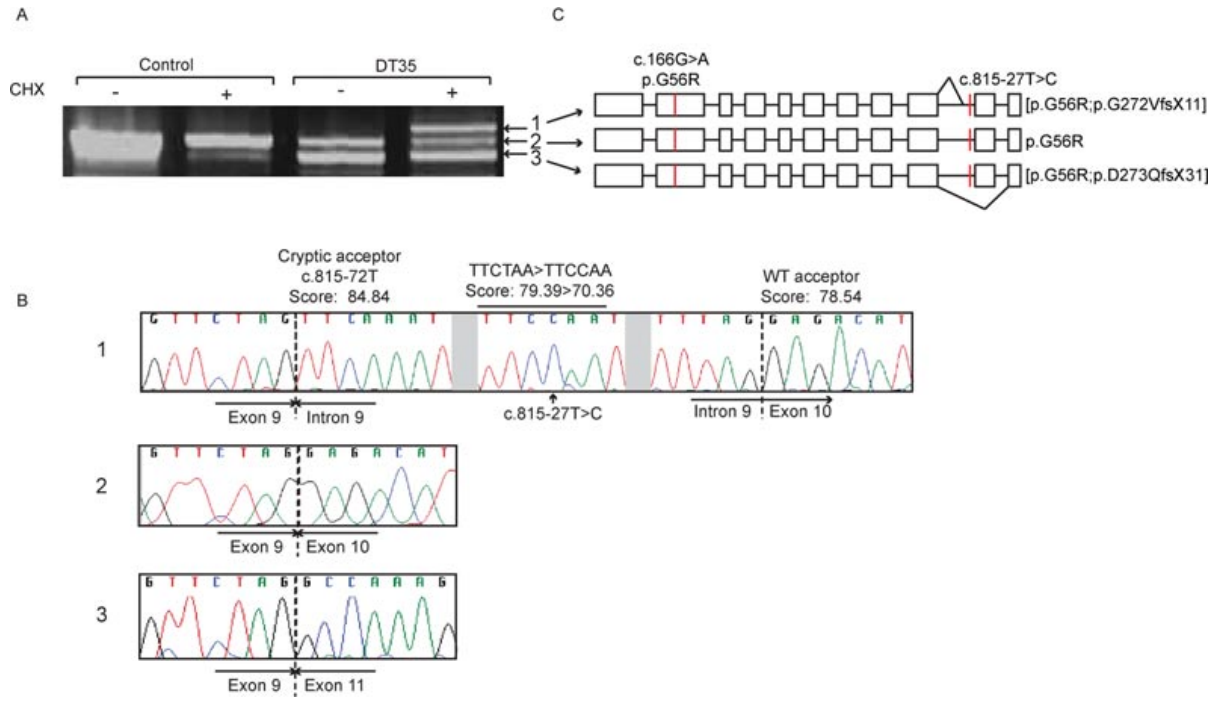

Figure 5. Published data on mRNA transcripts due to c.815-27T>C
The c.815-27T>C mutation causes aberrant NUBPL mRNA splicing. (A) Electrophoresis gel of fibroblast RT-PCR products generated using primers that amplify the entire NUBPL open reading frame. (B) Chromatograms demonstrate that the largest RT-PCR product (1) contains some intron 9 sequence. The c.815-27T>C mutation, which reduces the predicted branch-site score from 79.39 to 70.36, leads to partial use of a cryptic acceptor site that has a higher predicted score (84.84) than the wild-type acceptor site (78.54). Chromatograms of the middle band (2) show that some wild-type NUBPL transcript is generated whereas the smallest band (3) is generated by skipping of exon 10. (C) Schematic diagram showing the three transcripts generated in the presence of the c.815-27T>C mutation. From Tucker, ET, et al. Copyright 2012 by Wiley Periodicals Inc. Used by Permission of the publisher.

1.8 Aims of the current study

In previous studies by Kathy Hall, et al (2014), whole exome sequencing of two siblings with ataxia, developmental delay, and cerebral abnormalities generated a diagnosis of complex I deficiency due to compound heterozygous mutations in *NUBPL*. Though this family reached an end to the diagnostic odyssey for their two daughters, questions remained as to why these two sisters have significant phenotypic differences with regards to severity of ataxia and development of language and motor skills. In addition, a third case of a 2.5 year old female was also found to be a compound heterozygote for two slicing mutations, one of which was the same branch site mutation (c.815-27T>C) as the two previous cases while the other mutation was previously observed in a separate individual with complex I deficiency. Given these reports, further analysis of exome data in addition to RNA studies has the potential to provide insight into the role of modifier genes as well as possible differences in alternative splicing. From these RNA studies, some differences in alternative splicing were observed, particularly when interrogating the effect of the c.815-27T>C mutation between patients and control. In addition to these findings, previous analysis of whole exome data suggested the possibility of modifier genes affecting the role of NUBPL protein.

Specifically for this thesis, analyses of whole exome sequences as well as mRNA from patient and control fibroblasts will be the focus:

1. The first aim is an analysis of mRNA transcripts from fibroblasts obtained from these three cases as well as two controls. This data will together assist in providing possible reasoning for phenotypic differences among these three individuals and provide potential insight into the effects of alternative splicing on disease pathology.

2. The second aim of this study is an analysis of sequence data from two probands with mutations in *NUBPL* to assess for modifier genes that have been reported to have interaction with NUBPL protein. These genes are then assessed for clinical relevance using known population genetics resources.

2 Materials and Methods

Subjects in this study were consented for use of their biological samples and clinical information by a member of Dr. Virginia Kimonis' research team under IRB protocol 2002-2608. All research was conducted with de-identified data and research samples.

2.1 Procedures for measuring alternatively spliced transcripts

2.1.1 RT-PCR: Isolating RNA from patient and control fibroblasts

Patient and control fibroblasts were obtained by three patients affected with complex I deficiency with consent per IRB protocol 2002-2608. Control samples were selected based gender and age, being as close to the age of the patients as possible. RNA extraction and cDNA synthesis of samples were obtained with the assistance of Lan Weiss, MD, PhD. at UC Irvine. For RNA extraction from patient and control fibroblasts, the PureLink RNA Micro Kit from Thermo Fisher (Waltham, MA) was utilized. To summarize, cells were transferred to RNase-free tubes and centrifuged at 2,000xg for 5 minutes at 4ºC. Growth medium was discarded and 0.6 mL of lysis buffer with 2-mercaptoethanol was added. Cells were vortexed until fully lysed. Samples were centrifuged at 26,000xg for 5 minutes and then the lysates were transferred to new RNase free tubes. One volume of 70% ethanol (600uL) was added and then samples were vortexed and added to a spin cartridge. Samples were centrifuged at 12,000xg for 30 seconds. Spin cartridges were then washed with 450 uL of Wash Buffer and twice with 500uL of Wash Buffer II with ethanol, both of which were contained in the PureLink RNA Micro Kit. RNA was then eluted with 30uL of RNase-free water at 12000xg for 2 minutes. RNA samples were stored at -80ºC until further use. Table 1 summarizes patient samples used in RT-PCR experiments and Table 2 summarizes all sample RNA concentrations.

Table 1. Summary of RNA samples from patient fibroblasts

Case No.	**Sample Name**	**Age (yr)**	**Sex**	**Mutation 1**	**Mutation 2**
1	Patient 1 (Sibling A)	17	F	c.815-27T>C	c.311T>C
2	Patient 2 (Sibling B)	7	F		
3	Patient 3	5	F	c.815-27T>C; c.166G>A	c.693+1G>A

Table 2. Summary of RNA sample concentrations.

Sample	RNA concentration (ng/uL)	RNA amount used for cDNA synthesis (ug)
Patient 1	431.1	3.59
Patient 2	448.6	3.66
Patient 3	312.4	2.50
Control 1	244.2	2.91
Control 2	703.0	3.45

2.1.2 RT-PCR: cDNA synthesis of patient and control RNA samples

cDNA from extracted RNA was synthesized using SuperScript III First Strand Synthesis System from Thermo Fisher (Waltham, MA). To summarize, about 3 μg (8 μL) of total RNA from each sample was combined with 1 μL of 50 μM oligo(dT), 1 μL of 10 mM dNTP mix. The samples were incubated at 65ºC for 5 minutes, then placed on ice for 1 minute. A master mix of cDNA synthesis cocktail was made as follows with volumes listed as per 1 reaction: 10x RT buffer (1 μL), 25 mM $MgCl_2$ (4 μL), 0.1 MDTT (2 μL), RNaseOUT (40U/μL; 1 μL), and SuperScript III RT (22U/μL; 1 uL). A total of 10 uL of the cDNA synthesis was added to each RNA/oligo mixture. Each sample was then incubated at 50ºC for 50 minutes and terminated at 85ºC for 5 minutes. Samples were then chilled on ice followed by the addition of 1 uL of RNase and incubation at 37ºC for 20 minutes. Samples were then stored at -20ºC until further use.

2.1.3 Primer design for amplification of *NUBPL* transcripts

Primers for amplifying specific transcripts within *NUBPL* were designed with the assistance of Dr. Klemens Hertel from UC Irvine. Using UCSC Genome Browser as well as the OligoAnalyzer Tool from Integrated DNA Technologies (IDT), one set of primers was designed to assess the possible alternatively spliced products resulting from the c.311T>C missense mutation within exon 4 of *NUBPL* and another set was designed to assess possible alternative transcripts resulting from the c.815-27T>C branch site mutation within intron 9. In addition, a set of primers was designed to assess splicing differences as a result of the heterozygous c.693+1G>A splice site mutation in Patient 3 using the NCBI Primer Pick Program. Primers were ordered from IDT (San Diego, CA). Figure 6 illustrates a map of these primers relative to the mutations and Table 3 provides the sequence and melting temperature of each primer.

```
1    actccgcgcc acccgcgaca gtttcccagc agggctcaca gcagcgttcc gcgtcatggg
61   gatttggcag cgtctgctgc tttttggtgg ggtgtcgctc cgggctggtg gcggggccac
121  tgccccgctt gggggaagcc gagcgatggt ttgtgggcgc cagttgtctg gcgccgggag
181  tgagacccta aaacaaagaa gaacacaaat catgtcccga ggacttccaa agcagaaacc
241  gatagaaggt gttaaacaag ttatagttgt ggcttctgga aagggtggag tcggaaaatc
301  tactacagca gtgaatcttg cacttgcact agcagcgaac gattcgtcca aggccattgg
361  tttgctagat gtggatgtgt atggaccttc agttccaaag atgatgaatc tgaaaggaaa
421  tccggaatta tcacagagca acctaatgag gcctctcttg aattatggta ttgcttgtat
481  gtctatgggc tttctggttg aagaaagtga accagtagtt tggagaggcc ttatggtaat
541  gtcggccatt gagaaattgt tgaggcaggt agattggggt caactggact acttagttgt
601  agacatgcca ccaggaactg gagatgtgca gttatcagtc tcacagaata ttcctataac
661  aggtgctgtg attgtctcca cgccccagga catcgcattg atggatgcac acaagggtgc
721  tgagatgttt cgcagagtcc acgtgcccgt ccttggcctt gtccaaaaca tgagtgtttt
781  ccagtgtcca aaatgtaaac acaaaactca tatttttggt gctgatggtg caaggaaact
841  agcacagacc cttggtcttg aagttctagg agacattccc ttacacctta atataaggga
901  agcttcagat acaggccagc caattgtgtt ttcacagcct gaaagtgatg aggccaaagc
961  ttacttgagg attgctgtgg aagtggtaag aagattgcca tcaccttcag aatgattccc
1021 caagtgtcct ggaaatttgc ctggtactga cattaagagg acctttggaa atcagcaatg
1081 tggtgatgga acctacagaa ataatagaaa tcataactgt tttatttcta aggaaagaat
1141 gtcttcatat ttgacttgct aagctaaggt tcacaaaact ttgatgtatc aatgttaact
1201 gctatattta ggaatttttt gaaagctggt gtgtaccacc tgcaaagaac tgcattttat
1261 tttattgaat taccccttta gaaatcacga gtttatgatg ttacaagtcc attttgggag
1321 agaactgtac acttttgca ggaccacaga attagtaatt taaatgatat actaaatttg
1381 tgtatgggca taagtattac gccttcccac caggcaattt ggactgtctt tgaatcctgt
1441 ctttggtact gtcttggaca tgttcttaat atgaacctct gcttcttcat ctggatcaca
1501 ctatacccca gacttaatga atttcagtct ccaggagtgg agcctgaaca tatgtatttt
1561 tagtaagacc atgatgatgt tgatgatgat tctggtaatc atggcagtaa gaataatagc
1621 taacatttaa caagttaatt ataattgttc ttccatttgt gtgagtgggt actattatct
1681 ccattttaca gataagtaat ttgagatata atttgaaggc actcaacctt gagtgtcttg
1741 ctgagggtta ctgtatctag caggtgatgg agcttggatt tgaacttgag taattcgata
1801 ttagattcat ctttttgtac accaatgtgc ttttgttaat taaaaaaagt cggtggtggc
1861 taggcgtggt ggctcacgcc tgtaatccca gcactttggg aggttgaggt gggtggatca
1921 tgaggtcagg agttcaagac cagcctgacc aacatggtga aaccccgtct ctactaaaaa
1981 tacaaaaatt agctgggtgt ggtggcatat gcctgtaatc ccagctactc gggaggctga
2041 ggcaggagaa tcatttcaac ccgggaggct gaggttgcag tgagtggaga ttgtaccact
2101 gcactccagc ctggtcaatg gagtgcgact ccatctcaaa aaaaaaaaaa aaaaaaaaaa
2161 aagtcggggg agatgcaata tttagcattt cctcaaactt tttttggcca tgaaatccac
```

```
2221 cacccccctc ctttttcaga gcatttcatg aagttcacat ttcaaataat acattttgaa
2281 aaacatttat cttgtctaat tgaagctcaa agagatagag tcaatgactt ttcaaagtca
2341 ggtagcagat agttttaaag ctgcaagtac tgataagtgt gagccattag ttaattcgca
2401 tttataaaga agtggtctgt cacttttggc aaatgtcttg acatctgttg ctctacagtt
2461 gtcaactaga aaattaggca tagagtttca cattattata agcatgtttt cacacaaaaa
2521 catgtacatt gatgtttatt gcaattttat gtgtaataac caaaaacaac ccagattttc
2581 ttcattgggt aaatgattaa acaaaccata gtacattcat accgtgagtg ctactcagca
2641 ataaaaaaga acaaactatg ggtacaggca gcaacttata attccagaga attatgctga
2701 atgaaaaagg ccaatttaaa aaggttttgt accaaaagat tccatttata taacgttctt
2761 gaagtgacat aattatagaa atggagaaca gattagtggt ggctgtgtta aggagggggt
2821 gaggatggga gagaattggg cgtggctgtg agaaacatga gggatccttg tgatgatgga
2881 aatgttctat atcacagcta tcgatgtcag tatcccagtt gtgatattgt gctatagttt
2941 tgcaagatgt taccactggg ggaaattggg tgaagggcac atggaatctc tgtattattt
3001 caactgcatg taaatctaca attatctcaa aataaaaata aaagttaaaa aaacacaaaa
3061 aaaaaaaaaa aa
```

Figure 6. mRNA sequence of *NUBPL*. The sequence highlighted in yellow represents the forward primer designed to determine the potential splicing impact of the c.311T>C mutation within exon 4 of *NUBPL* ('c.311T>C Forward') whereas the sequence highlighted in pink is the reverse primer ('c.311T>C Reverse'). The sequence highlighted in green represents the forward primer on exon 9 ('c.815-27T>C Forward') whereas the sequence highlighted in blue represents the reverse primer on exon 11 ('c.815-27T>C Reverse'), both of which were selected in order to elucidate potential splicing impact of the c.815-27T>C mutation within intron 9. Finally, the sequence bolded and underlined beginning at nucleotide 598 is the forward primer on Exon 7 ('c.693+1G>A Forward') whereas the sequence bolded and highlighted at nucleotide 838 is the reverse primer on Exon 9 ('c.693+1G>A Reverse'). Both of these latter primers were designed to assess differences in splicing from the heterozygous c.693+1G>A splicing mutation found in Patient 3.

Table 3. List of primers used in RT-PCR of patient and control fibroblast cDNA.

Primer pair name	**Forward or Reverse**	**Sequence**	**Location**
'c.311T>C'	Forward	CGATAGAAGGTGTTAAACAAGTTATAGTTGTGGC	Exon 2
	Reverse	GGCCGACATTACCATAAGGCC	Exon 6
'c.815-27T>C'	Forward	GGTGCAAGGAAACTAGCACAGACC	Exon 9
	Reverse	GGCAATCTTCTTACCACTTCCACAGC	Exon 11
'c.693+1G>A'	Forward	TGTAGACATGCCACCAGGAAC	Exon 7
	Reverse	GACCAAGGGTCTGTGCTAGT	Exon 9

2.1.4 PCR and gel electrophoresis of patient and control transcripts

Polymerization chain reaction (PCR) was performed on cDNA from patient and control fibroblasts using primers listed in Table 2. A PCR cocktail mix was made consisting of the following, per 1 reaction: 8.5 µL of H_2O, 6 µL of 5x PCR buffer, 6 µL of 25 mM $MgCl_2$, 10 mM nucleotides, 1.5 µL 10 µM forward primer, and 1.5 µL of 10 µM reverse primer. Taq polymerase (5u/ µL, Promega, Madison, WI) was diluted separately (0.3 µL Taq polymerase with 3.7 µL of H_2O). Per reaction, 25 µL of the PCR cocktail mix and 4 µL of diluted Taq polymerase was added to 1 µL of 100 ng/ µL of each sample. PCR was performed using a thermocycler (Biorad) as summarized in Table 4. Samples were run on either a 1%, 2% or 3% agarose gel.

Table 4. Summary of PCR conditions used on cDNA of patient and control fibroblasts. The melting temperature (bolded and italicized) varied with each PCR reaction, ranging from 56ºC to 60ºC depending on which primers were used.

1 cycle	95°C	5 minutes
35 cycles	95°C	45 sec
	56°C, 58°C or 60°C	30 sec
	72°C	1 minute
1 cycle	72°C	5 minute
	4°C	indefinite

2.1.5 Calculation of mRNA transcript band intensities

Following electrophoresis, gels were exposed to UV light for visualization of mRNA transcript bands using a BioRad Gel Doc XR imaging system. Quantity One Analysis software was then used to select individual bands for quantification of pixel intensity. These band intensities were summed and each band was divided by the total sum of intensity for each control or patient sample. These values were then used as the relative intensities for comparison between patients and control samples.

2.2 Whole exome sequencing analysis: filtering for potential modifier genes in siblings with compound heterozygous mutations in *NUBPL*

In previous studies by Hall, et al (2014), assessment of potential modifier genes was performed on whole exome data of the two siblings with compound heterozygous mutations in *NUBPL* (in addition to their mother). Whole exome sequencing data was provided by Ambry Genetics and summary of the pipelines used in the analysis of data were previously described (Hall et al., 2014). Briefly, variant analysis was performed using three databases containing whole exome sequencing data: Human Gene Mutation Database (HGMD), Exome Variant Server from the Exome Sequencing Project (ESP), and 1000 Genomes. The bioinformatics pipeline used by Ambry Genetics for analysis of modifier genes allowed for the ability to assess the quality and significance of each variant by utilizing autosomal dominant and autosomal recessive inheritance patterns. Variant filtering and annotation of these variants was performed using the Varseq program from Golden Helix (Golden Helix, Inc.). For autosomal dominant inheritance, the allele frequency threshold was set to 1 percent. For autosomal recessive inheritance, the allele frequency threshold was set to 10 percent. To be included as a possible

modifier gene, Sibling A needed to be positive for the variant, Sibling B needed to be negative for the variant, and the mother could be either positive or negative for the variant. Quality of the variant was further assessed using PolyPhen (Adzhubei, 2010) and SIFT (Ng and Henikoff, 2006) scores as well as the phenotype associated with the gene.

In recent years, there have been several additional genes identified that have been shown to potentially effect complex I function via interaction with *NUBPL*. Using the open-source human gene database, GeneCards, potential interacting genes were investigated via the String Interaction Network tool that provides information regarding known and predicted gene interactions. In addition, this tool specifies gene interactions based on whether the genes are co-expressed, have protein homology, or gene co-occurrence and provide literature evidence for such interactions. The String Interaction Network provided additional genes to examine within the whole exome variant data of the two siblings in order to investigate potential modifier genes that may be affecting disease severity. By using a similar method as Hall et al for identifying potential modifier genes within the variant list provided by Ambry Genetics, additional variants were explored and assessed using gene interaction data from the String Interaction Network as well as from recent scientific publications throughout PubMed.

3 Results

3.1 Analysis of mRNA transcripts in patients with *NUBPL* branch site mutation

In order to probe possible differences in mRNA transcripts between controls and three patients with complex I deficiency due to mutations in NUBPL, three sets of primers were designed to surround three different mutations found within these patients. The first is the c.815-27T>C branch site mutation found in all three patients. The second is the c.693+1G>A splice mutation that is *in trans* with the branch site mutation in Patient 3 and not in the two siblings, Patients 1 and 2. The third mutation that was interrogated is the c.311T>C missense mutation that is found *in trans* with the c.815-27T>C mutation in Patients 1 and 2, but is not in Patient 3. Results of PCR and gel electrophoresis of these primers with cDNA from each patient are as described below.

3.1.1 Results of cDNA gel electrophoresis using primers encompassing c. 815-27T>C mutation

Control and patient cDNA extracted from fibroblasts underwent PCR using 'c.815-27T>C' primers to investigate transcripts resulting from this branch site mutation within intron 9. Gel electrophoresis of these samples resulted in a single band of approximately 200 bp in length. This observed band likely represents a transcript that includes exon 10 as well as a portion of exons 9 and 11. Interestingly, this band is partially present or absent for Patient 3. These samples were run on a 1% agarose gel as well as a 3% agarose gel (Figure 7). Upon further optimization of gel electrophoresis conditions, samples were re-run on 1.5% gel, resolving additional transcripts. Interestingly, for Patient 3, there is increase in intensity for a 150 bp band with a concomitant decrease in the approximately 200 bp transcript relative to Patients 1

and 2. While Patients 1 and 2 carry the same c.815-27T>C branch site mutation that Patient 3 carriers, they do not carry the same second *in trans* splice mutation (c.693+1G>A) nor do they carry the *in cis* missense mutation (c.166G>A).

Following further optimization of gel electrophoresis methods, PCR products using 'c.815-27T>C' primers were re-analyzed on a 2% agarose gel (Figure 8). Control cDNA was from fibroblasts of a 7-year-old healthy female. A negative control was included, which was a sample containing no DNA. Interestingly, for Patient 3, the expected 200 bp remains to have a decrease in intensity relative to Patient 1 and Patient 2 as well as the controls. In addition, a 150 bp mRNA transcript is resolved for all three patients and not the positive control. An approximately 300 bp appears in the three patients as well that is not observed in the positive control. These differences in transcripts are also well resolved on a black-to-white inverted background (Figure 9).

Band intensity (mRNA transcript intensity) for each of the control and patient samples were quantified within Quantity One gel image program as described in Materials and Methods. For mRNA transcripts in each sample, band intensities were quantified and are as summarized in Table 5. Band numbers in Figures 8 and 9 correspond to band numbers in this Table 5. These band intensities are also compared in Figure 10.

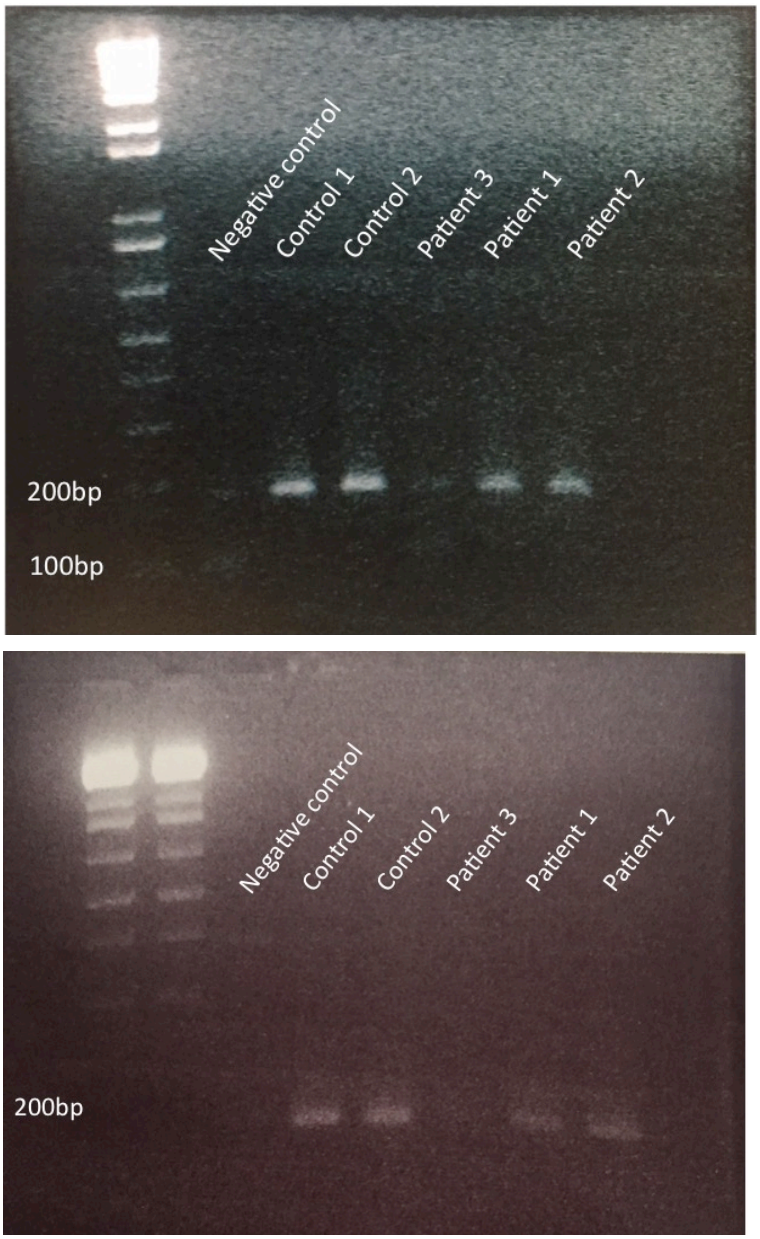

Figure 7. Gel electrophoresis of cDNA from patient and control fibroblasts. PCR was performed on cDNA using 'c.815-27T>C' primers encompassing this intron 9 splice site mutation. Control cDNA was from fibroblasts from a 7-year-old healthy female (Control 1) and 16-year-old healthy female (Control 2). A negative control was included, which was a sample using no DNA. (Top panel) Electrophoresis of PCR samples on 1% agarose gel. (Bottom panel) Electrophoresis of PCR samples on 3% agarose gel.

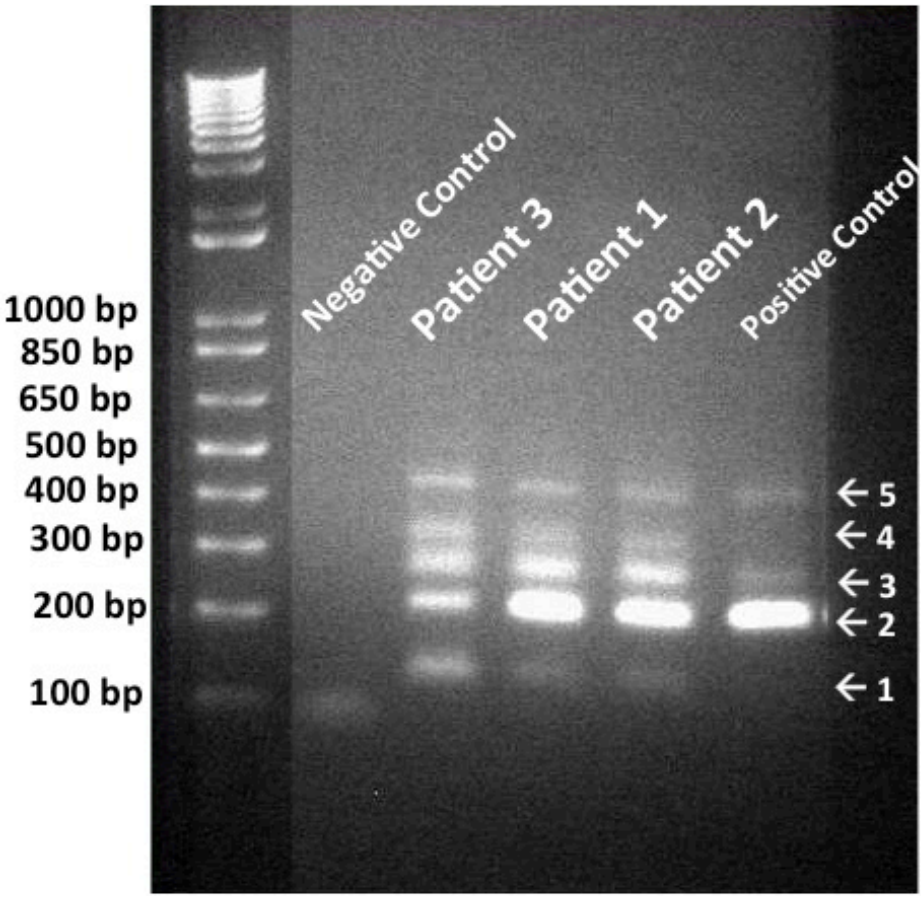

Figure 8. Gel electrophoresis of cDNA from patient and control fibroblasts. Control cDNA was from fibroblasts of a 7-year-old healthy female. A negative control was included, which was a sample containing no DNA. PCR was performed on cDNA using 'c.815-27T>C' primers encompassing this intron 9 splice site mutation. Gel electrophoresis conditions were optimized and run on 2% gel.

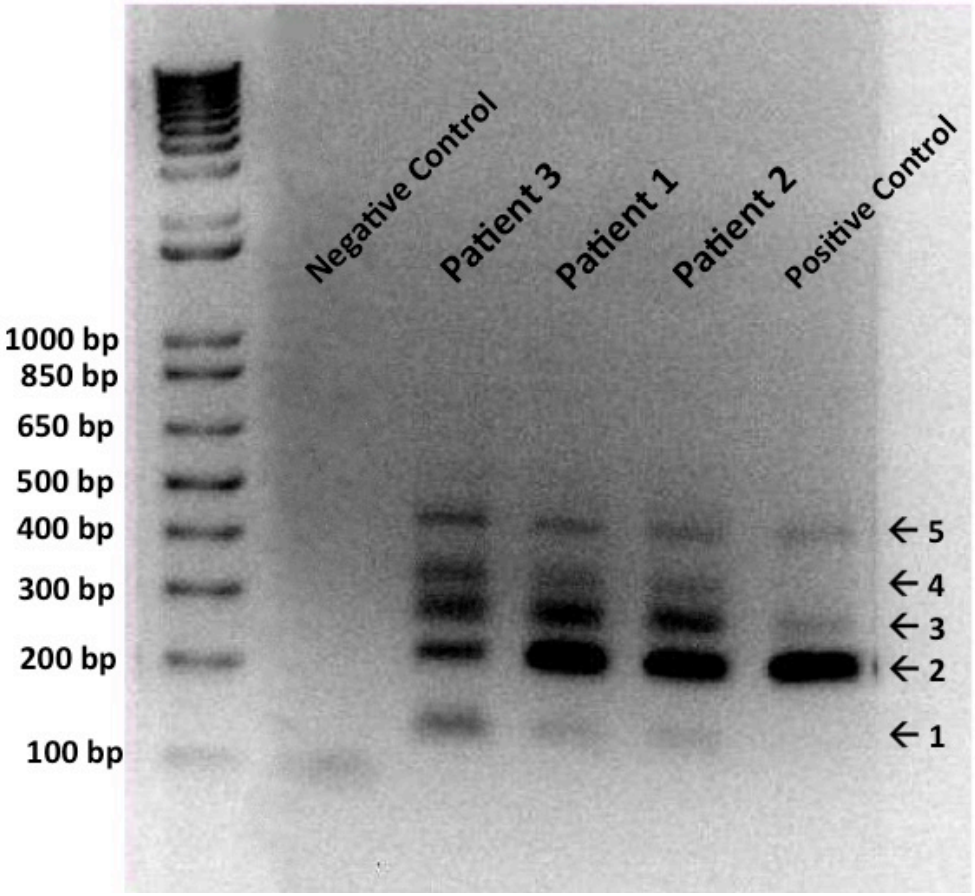

Figure 9. Gel electrophoresis of cDNA from patient and control fibroblasts. This figure is the same as Figure 8 but inverted to show alternative view of bands at approximately 150 bp. Control cDNA was from fibroblasts of a 7-year-old healthy female. A negative control was included, which was a sample containing no DNA. PCR was performed on cDNA using 'c.815-27T>C' primers encompassing this intron 9 splice site mutation. Gel electrophoresis conditions were optimized and run on 2% gel.

Table 5. Summary of quantified band intensities for mRNA transcripts obtained using 'c.815-27T>C' primers. Band number corresponds to the band number labels in both Figures 8 and 9.

Sample	Band	Band Intensity	Sum of intensities for all bands per sample	Ratio of each band intensity/Sum of all band intensities	Relative band intensity
Negative Control	1	3665.112141	110.7446731	-	-
	2	-111.9091809		-	-
	3	22.64147202		-	-
	4	83.49585028		-	-
	5	116.5165317		-	-
Positive Control	1	-390.9	51236.582	-0.008	-0.8
	2	44829.0		0.875	87.5
	3	2393.9		0.047	4.7
	4	256.3		0.005	0.5
	5	4148.2		0.081	8.1
Patient 3	1	6665.4	27744.3	0.240	24.0
	2	9477.8		0.342	34.2
	3	5970.4		0.215	21.5
	4	2902.2		0.105	10.5
	5	2728.5		0.098	9.8
Patient 1	1	1916.2	48686.9	0.039	3.9
	2	35925.9		0.738	73.8
	3	6873.8		0.141	14.1
	4	1767.0		0.036	3.6
	5	2203.9		0.045	4.5
Patient 2	1	279.0	44444.0	0.006	0.6
	2	32521.9		0.732	73.2
	3	7224.6		0.163	16.3
	4	1834.2		0.041	4.1
	5	2584.330242		0.058	5.8

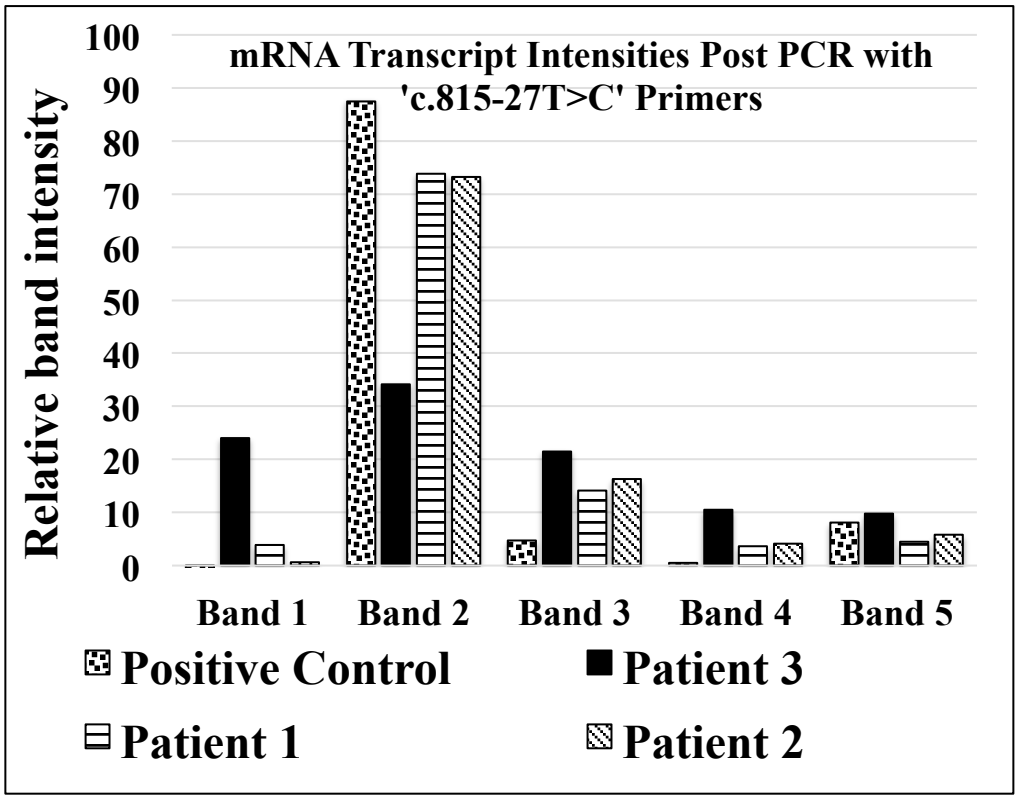

Figure 10. Summary of relative band intensities corresponding to bands labeled in Figure 8 and 9 as well as Table 5. Bars represent the intensity of each band following PCR using 'c.815-27T>C' primers, which corresponds to individual mRNA transcripts for each control and patient sample.

3.1.2 Results of cDNA gel electrophoresis using primers encompassing c.693+1G>A mutation

Given Patient 3 carries a separate splice-site mutation (c.693+1G>A) *in trans* with the c.815-27T>C branch site mutations, primers were designed to probe possible differences in mRNA transcripts resulting from this mutation. The same control and patient cDNA extracted from fibroblasts underwent PCR using these 'c.693+1G>A' using the same DNA sample volumes (1uL). PCR was performed at 56^{o}C (the melting temperature of both of these forward and reverse primers) as well as 58^{o}C and 60^{o}C as a show in Figures 11 and 13, respectively. Gel electrophoresis of these samples resulted in approximately 260 bp, mRNA transcripts for the positive controls and all three patient samples when PCR was run with a melting temperature of 56^{o}C. Interestingly, approximately 700 bp and 1000 bp mRNA transcripts were observed for both positive controls as well as Patient 3. When a melting temperature of 58 ^{o}C and 60 ^{o}C were implemented into the PCR protocol, only the approximately 260 bp and 700 bp mRNA transcripts were resolved for both positive controls and Patients 1, 2, and 3. The 1000 bp transcript may be a non-specific transcript outside of the NUBPL gene. The 700 bp transcript may also be a non-specific transcript, however this band is either not observed (Figure 11) or less intense (Figure 13) for Patients 1 and 2.

Similar to Table 4, band intensity for each of the control and patient samples were quantified within Quantity One gel image program as described in Materials and Methods. For mRNA transcripts in each sample for Figures 11 and 13, intensities were quantified and are as summarized in Tables 6 and 7, respectively. These band intensities are also compared in Figures 12 and 14, respectively.

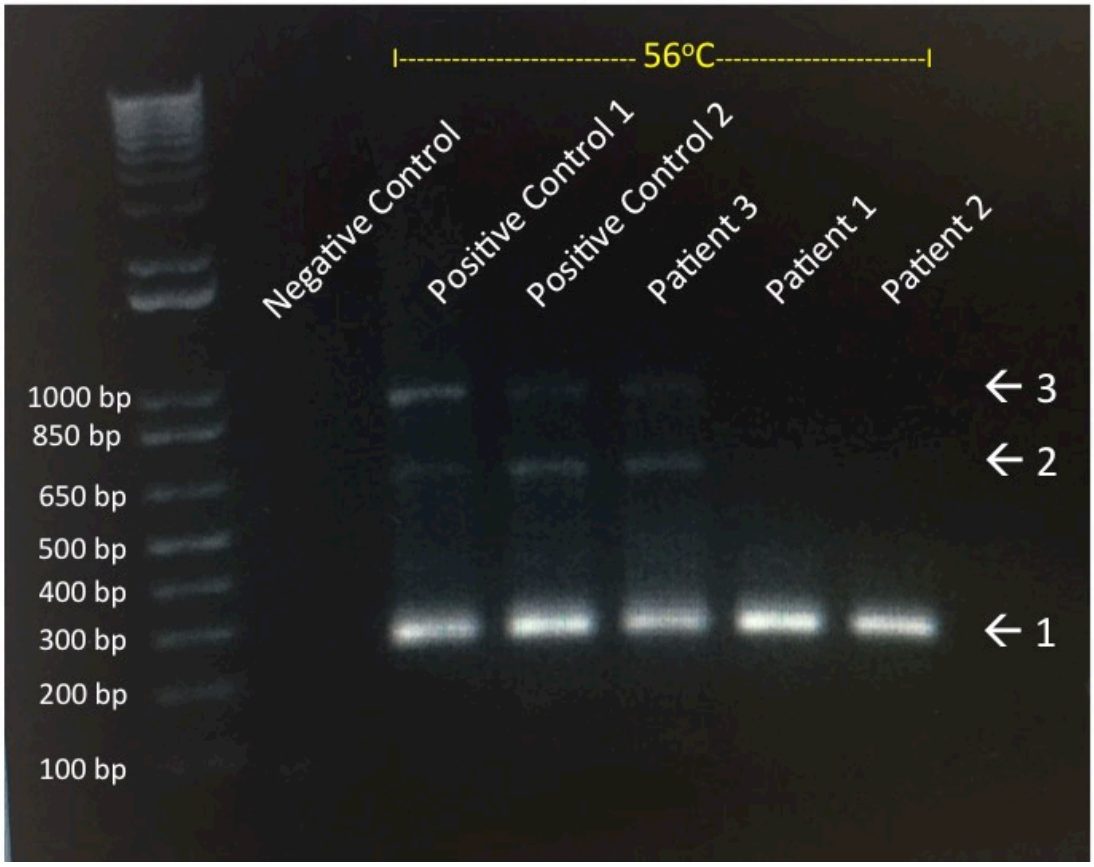

Figure 11. Gel electrophoresis of cDNA from patient and control fibroblasts. Control cDNA was from fibroblasts of a 7-year-old healthy female (Control 1) and a 16 year-old healthy female (Control 2). A negative control was included, which was a sample using no DNA. PCR was performed on cDNA using 'c.693+1G>A' primers encompassing the c.693+1G>A heterozygous splice site mutation found in Patient 3 but not in Patients 1 and 2. This PCR was performed at 56^{o}C, the melting temperature for theses forward and reverse primers. Gel electrophoresis conditions were optimized and run on 2% gel.

Table 6 Summary of quantified band intensities for mRNA transcripts obtained using 'c.693+1G>A' primers at 56^{o}C melting temperature. Band number corresponds to the band number labels in Figures 11 and 12.

Sample	Band	Band Intensity	Sum of intensities for all bands per sample	Ratio of each band intensity/Sum of all band intensities	Relative band intensity
Negative Control	1	-664.8	-632.3	-	-
	2	122.6		-	-
	3	-90.1		-	-
Positive Control 1	1	37283.1	47066.1	0.792	79.2
	2	3170.7		0.067	6.7
	3	6612.3		0.140	14.0
Positive Control 2	1	45613.1	53765.4	0.848	84.8
	2	5125.7		0.095	9.5
	3	3026.5		0.056	5.6
Patient 3	1	30430.8	37386.2	0.814	81.4
	2	4095.5		0.110	11.0
	3	2859.9		0.076	7.6
Patient 1	1	48784.2	48228.3	1.012	101.2
	2	-261.4		-0.005	-0.5
	3	-294.6		-0.006	-0.6
Patient 2	1	40844.0	40802.7	1.001	100.1
	2	238.8		0.006	0.6
	3	-280.1		-0.007	-0.7

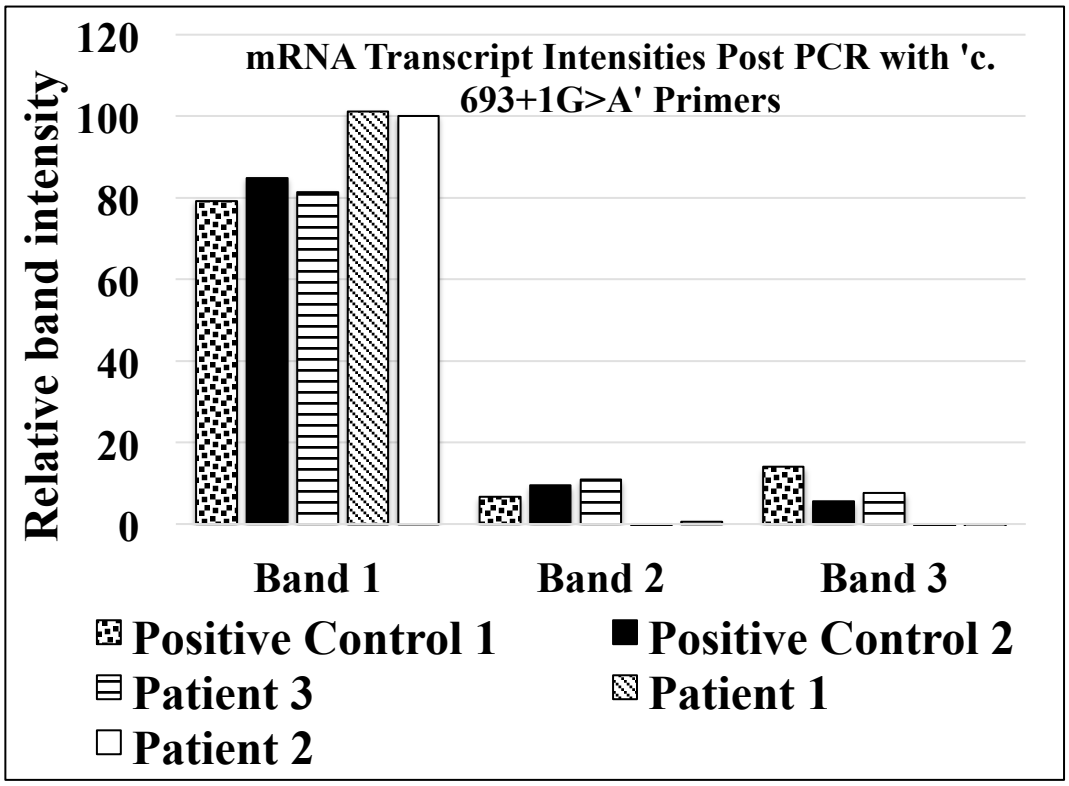

Figure 12. Summary of relative band intensities corresponding to bands labeled in Figure 11 as well as Table 6. Bars represent the intensity of each band following PCR using 'c.693+1G>A' primers at 56^{o}C melting temperature, which corresponds to individual mRNA transcripts for each control and patient sample.

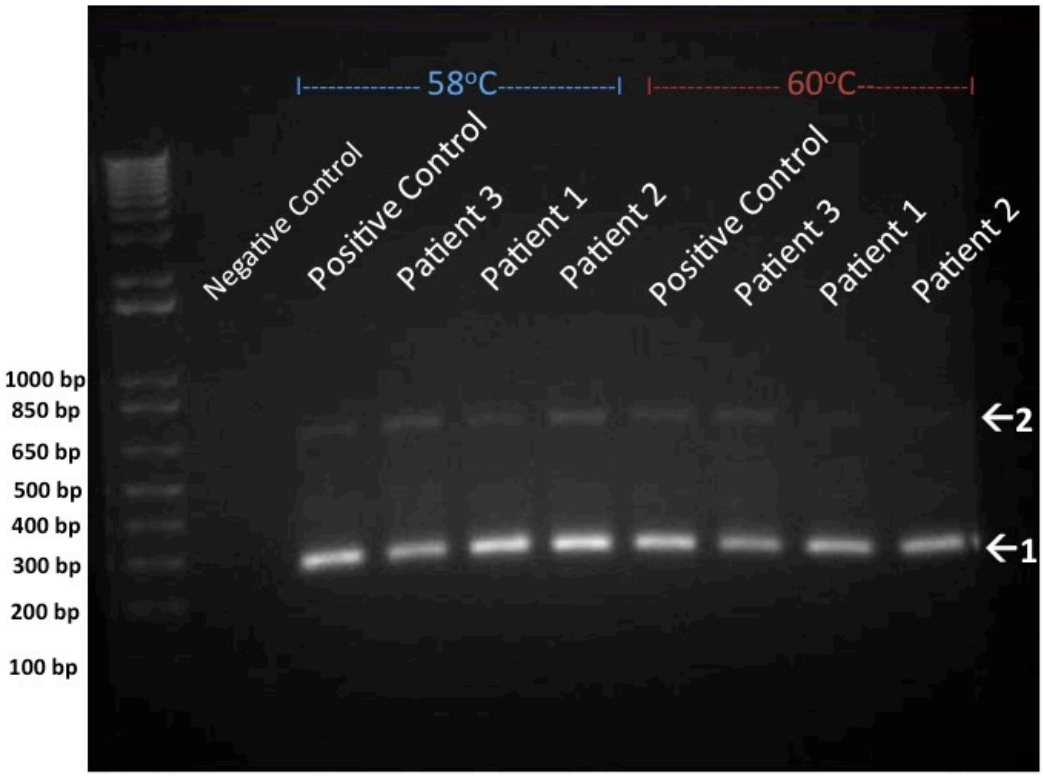

Figure 13. Gel electrophoresis of cDNA of patient and control fibroblasts. Control cDNA was from fibroblasts of a 7-year-old female. A negative control was included, which was a sample using no DNA. PCR was performed on cDNA using 'c.693+1G>A' primers encompassing the c.693+1G>A heterozygous splice site mutation in Patient 3 but not in Patients 1 and 2. This PCR was performed at 58°C and 60°C. Gel electrophoresis conditions were optimized and run on 2% gel.

Table 7. Summary of quantified band intensities for mRNA transcripts obtained using 'c.693+1G>A' primers at 58°C and 60°C melting temperature. Band number corresponds to the band number labels in Figures 13 and 14.

Sample	Band	Band Intensity	Sum of intensities for all bands per sample	Ratio of each band intensity/Sum of all band intensities	Relative band intensity
Negative Control	1	-998.4	-922.6	-	-
	2	75.8		-	-
Positive Control, 58 °C	1	51516.8	55187.7	0.933	93.3
	2	3670.8		0.067	6.7
Patient 3, 58°C	1	37334.9	44908.2	0.831	83.1
	2	7573.4		0.169	16.9
Patient 1, 58 °C	1	55661.4	59296.6	0.939	93.9
	2	3635.2		0.061	6.1
Patient 2, 58 °C	1	58042.5	65945.2	0.880	88.0
	2	7902.7		0.120	12.0
Positive Control, 60 °C	1	48836.9	53487.7	0.913	91.3
	2	4650.9		0.087	8.7
Patient 3, 60 °C	1	34364.2	39314.4	0.874	87.4
	2	4950.2		0.126	12.6
Patient 1, 60 °C	1	42475.6	43900.7	0.968	96.8
	2	1425.1		0.032	3.2
Patient 2, 60 °C	1	38943.3	41315.0	0.943	94.3
	2	2371.8		0.057	5.7

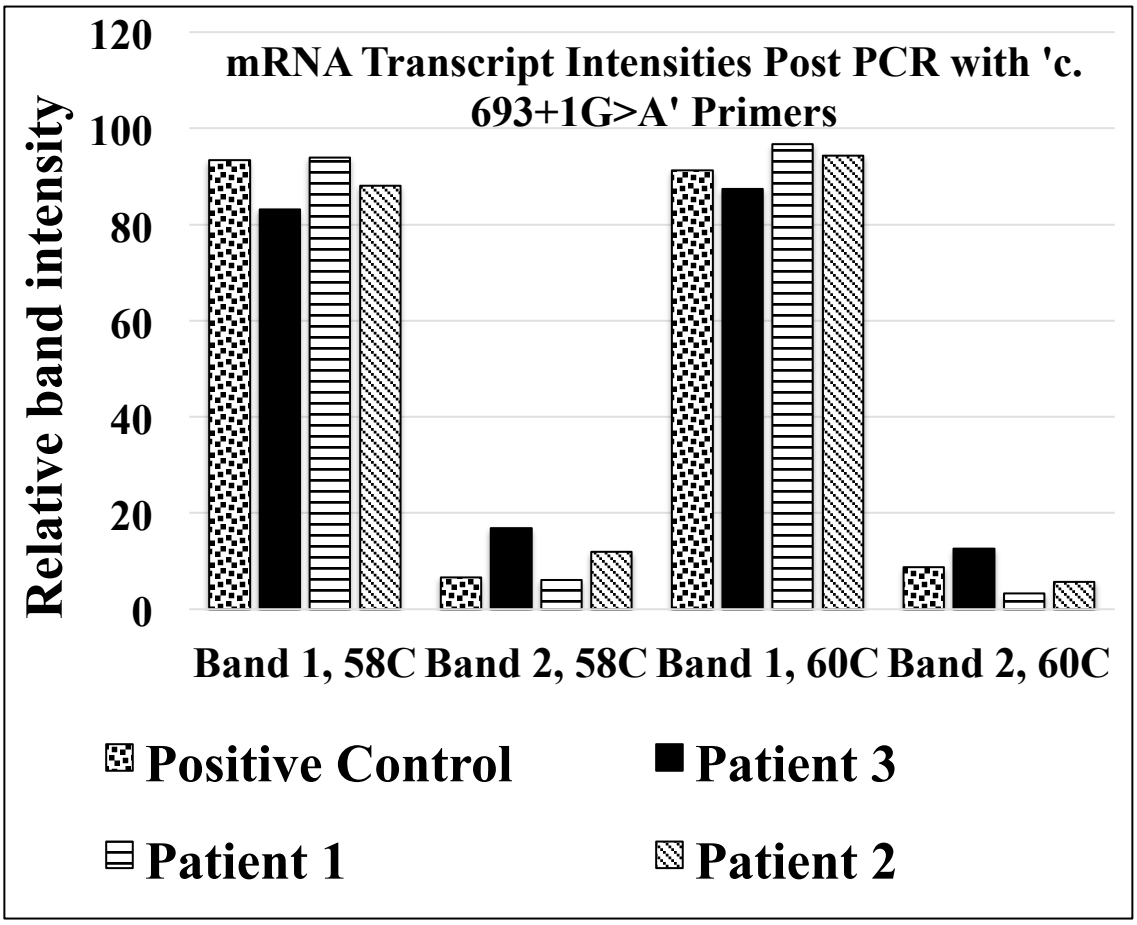

Figure 14. Summary of relative band intensities corresponding to bands labeled in Figure 13 as well as Table 7. Bars represent the intensity of each band following PCR using 'c.693+1G>A' primers at 58^{o}C and 60^{o}C melting temperatures, which corresponds to individual mRNA transcripts for each control and patient sample.

3.1.3 Results of cDNA gel electrophoresis using primers encompassing c.311T>C mutation

Control and patient cDNA extracted from fibroblasts underwent PCR using 'c.311T>C' primers to investigate potential transcripts resulting from this missense mutation within exon 4. Gel electrophoresis of these samples resulted in 2 bands of approximately 100 bp and 300 bp in length. The 300 bp band corresponds to the expected transcript that includes exon 3, 4 and 5 as well as a portion of exons 2 and 6. These samples were run on a 1% agarose gel (Figure 15).

Band intensities for each of the control and patient samples were quantified previously as described. For mRNA transcripts in each sample for Figure 15, band intensities were quantified and are as summarized in Table 8. These band intensities are also compared in Figure 16.

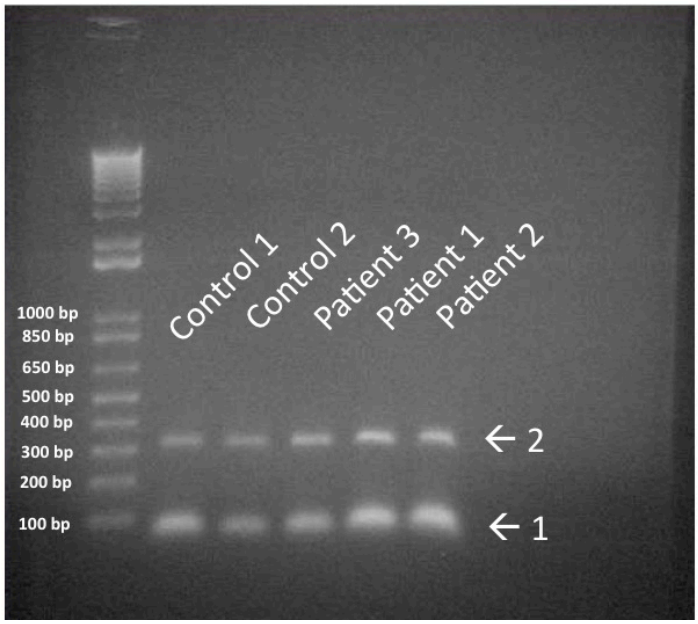

Figure 15. Gel electrophoresis of cDNA of patient and control fibroblasts. PCR was performed on cDNA using 'c.311T>C' primers encompassing this exon 4 mutation. Control cDNA was from fibroblasts from a 7-year-old female (Control 1) and 16-year-old-female (Control 2).

Table 8. Summary of quantified band intensities for mRNA transcripts obtained using 'c.311T>C' primers. Band number corresponds to the band number labels in Figure 15.

Sample	Band	Band Intensity	Sum of intensities for all bands per sample	Ratio of each band intensity/Sum of all Band intensities	Relative band intensity
Negative	1	-103.5135619	-560.5	-	-
Control	2	-456.9395856		-	-
Positive	1	20181.8	27433.4	0.736	73.6
Control 1	2	7251.6		0.264	26.4
Positive	1	11083.7	18327.7	0.605	60.5
Control 2	2	7243.9		0.395	39.5
Patient 3	1	15900.3	26843.2	0.592	59.2
	2	10942.9		0.408	40.8
Patient 1	1	26567.4	41243.2	0.644	64.4
	2	14675.8		0.356	35.6
Patient 2	1	27950.3	41387.2	0.675	67.5
	2	13436.9		0.325	32.5

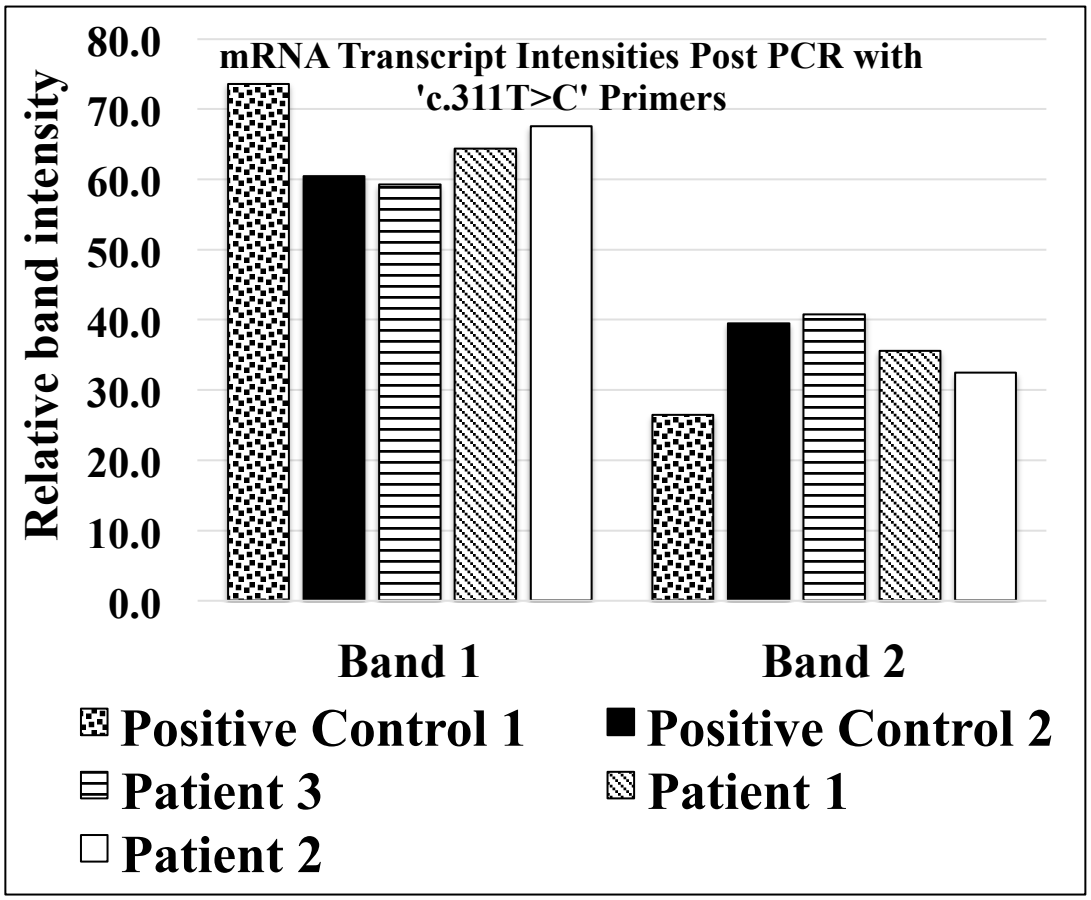

Figure 16. Summary of relative band intensities corresponding to bands labeled in Figure 15 as well as Table 8. Bars represent the intensity of each band following PCR using 'c.311T>C', which corresponds to individual mRNA transcripts for each control and patient sample.

3.2 Analysis of exome data for modifier genes in two siblings with *NUBPL* branch site mutation

In previous studies by Hall, et al (2014), investigation of potential modifier genes was performed on whole exome data of the two siblings with compound heterozygous mutations in *NUBPL* (c.815-27T>C and c.311T>C). Given the differences in phenotypic severity between these two siblings, Hall et al assessed whether there were variants that were present in the more affected sibling (Sibling A) that would negatively affect complex I activity and impact disease progression. In the past two years, several genes have been shown to interact with NUBPL and complex I activity. Using the String Interaction Network from GeneCards, additional genes were investigated and assessed within the whole exome variant data of the two siblings. In total, 5 genes were listed in close proximity within the String Interaction Network, interacting with NUBPL by correlation of co-expression, protein homology, or gene co-occurrence. An additional 2 genes were assessed based on literature analysis of genes interacting with NUBPL, including those in the iron sulfur cluster biosynthesis pathway.

The 4 genes that were found within the String Interaction Network were assessed for in the whole exome sequencing data, which was filtered as described in the Materials and Methods.

3.2.1 *NARF* and *NARFL*

The first gene assessed was *NARF*, a gene that produces a nuclear protein suggested to interact with prenylated prelamin A. By impacting prenylated prelamin A, NARF may affect lamin A function, which is thought to play a role in the organization of DNA and transcription in cells (Barton et al., 1999). *NARFL* (also referred to as *IPO1*) was also assessed in the whole exome sequencing data, which encodes a protein involved in the cytosolic iron-sulfur protein assembly (CIA) complex that mediates the incorporation of iron-sulfur cluster into

extramitochondrial iron sulfur proteins (Huang et al., 2007). No variants in this gene were detected for either sibling.

3.2.2 *CIAPIN1* and *NDOR1*

Additional interacting genes listed in the String Interaction Network were *CIAPIN1*, *NDOR1*, and *ACP1*. These genes have been implicated in cytosolic iron-sulfur protein biogenesis. *NDOR1* encodes for the protein NDOR1 or NADPH-dependent diflavin oxidoreductase 1. *CIAPIN1* encodes for the protein cytokine-induced apoptosis inhibitor 1, which is one of the domains of the larger protein, anamorsin. In studies by Banci et al, the electron transfer process between Ndor1 and anamorsin has also been implicated in the assembly of iron-sulfur clusters aiding in the regulation of cell survival/death mechanisms (Banci et al.. 2011). After examination of these genes in the whole exome data of the two siblings, there were no variants in either of these genes.

3.2.3 *C8ORF38 (NDUFAF6)*

The *C8ORF38* gene encodes for the protein, C8ORF38, which is also one of the assembly factors that aid complex I biogenesis. The protein has evolutionary conservation and has been shown to be involved specifically in the integration of ND1, a complex I subunit (McKenzie et al., 2011). Maturation of this mtDNA-encoded subunit requires C8orf38, as defects in this protein results in the loss of ND1 and therefore a complex I assembly defect. In vitro analyses were also performed where patient fibroblasts harboring a mutation in the *C8ORF38* gene contained reduced steady-state complex I levels and a defect in the translation of this ND1 subunit. Mutations in *C8ORF38* have been associated with the mitochondrial disease, Leigh syndrome, a disorder that is associated with psychomotor regression and progressive neurologic decline that causes spasticity, chorea, cerebellar ataxia, and peripheral neuropathy

(Rahman et al., 2015). Seven such cases of Leigh disease caused by allelic variants in *C8ORF38* are described in OMIM. In studying this gene within whole exome data of the two siblings and their mother, the mother and Sibling B did not have any variants, however Sibling A was found to be heterozygous for c/t alteration at 8:96037352. This variant had a read depth of 18 with a quality of 99% and an alternate allele frequency of 50%. This change was not present in the ExAC database and does not have an associated reference SNP (rs) number. This variant is also not listed in ClinVar, implying this change is likely rare. However, this variant is predicted to be neutral by modeling software Provean (provean.jcvi.org; Choi et al., 2012). Despite the functional connection of *C8ORF38* that suggests a role in processes connected to mitochondrial complex I function, it is uncertain, yet possible, that this particular variant would affect the role of NUBPL protein function in it's separate assembly of iron-sulfur clusters in complex I.

3.2.4 *ISCA1*

Other potential modifier genes that could impact the NUBPL protein function are other genes involved in expressing genes in the iron sulfur cluster pathway. These genes include *ISCU, NFS1, ISD11,* Frataxin *(FXN),* Ferredoxin 2 *(FDX2),* Ferredoxin reductase *(FDXR),* Mortalin *(GRP75, HSPA9), HSCB, GRPEL1/2, GLRX5, ISCA1, ISCA2, IBA57, NFU1,* and *BOLA3* (Stehling et al., 2014). Exome data was searched for variants in these genes, and heterozygous variants were found in *ISCA1* in both siblings, and not in the mother. The variant is c.241+1G>T splice mutation at position 9: 88886921 and listed as a loss of function mutation in the VarSeq variant analysis software. The variant is not listed in the ExAc, indicating it is a likely rare variant in the general population.

Like NUBPL, ISCA1 is one of the late-acting targeting factors in the iron sulfur cluster pathway. In a study by Sheftel et al, ISCA1, ISCA2, and IBA57 were found to be crucial

proteins involved specifically in the generation of [4Fe-4S] clusters and their insertion into mitochondrial apoproteins (Sheftel et al., 2012). One mitochondrial disease arising from homozygous mutations in the *ISCA1* gene has been recently described in two unrelated families (Shukla et al., 2017). Two affected children in each of these two families were described as having early onset neurological deterioration, seizures, extensive white matter abnormalities, cortical migrational abnormalities, lactic acidosis and early demise. Exome sequencing of one individual from each of these families revealed homozygous c.295G>A variants in *ISCA1*. This phenotype has been observed in other individuals diagnosed with multiple mitochondrial dysfunction syndrome (MMDS) due to mutations in other iron sulfur biogenesis genes that included *NFU1*, *BOLA3*, *IBA57*, and *ISCA2*. Given the phenotypic overlap and that ISCA1 is also involved in iron sulfur cluster synthesis, Shukla et al suggest that pathogenic variants in ISCA1 are also related to MMDS.

Although the two siblings in our study are heterozygous for c.241+1G>T variant in *ISCA1*, this variant is also likely rare given it is not listed in ExAc database and it's involvement in the late-stage iron sulfur cluster synthesis pathway suggests that this alteration in conjunction with the *NUBPL* alterations may as a whole impact iron sulfur synthesis.

4 Discussion

Previous studies on two siblings with a rare diagnosis of complex I deficiency caused by mutations in *NUBPL* highlighted the important and effective use of clinical whole exome sequencing for diagnosis of rare diseases (Hall et al., 2014). By investigating variants within the exome of these two siblings and their parents, compound heterozygous mutations were identified, one of which was a branch site mutation within intron 9. Branch site mutations can often alter mRNA products and affect protein expression, and therefore protein function, within the organism. One method in which these specific mutations can be assessed for pathogenicity is by studying the mRNA transcripts after intronic DNA is spliced from exonic DNA. Given that these splicing events can vary between individuals and thereby produce different mRNA transcripts, the first aim of this thesis investigated mRNA transcripts from these two siblings in addition to a third patient whom have the same intronic branch site mutation (c.815-27T>C). In addition, this third patient was also found to have a second splice mutation *in trans* with the c.815-27T>C branch site mutation. Differences in this splice site mutation, c.693+1G>A, were also explored by comparing mRNA transcripts between all three patients and controls. Other factors that can influence disease pathology include the presence of modifier genes which impact the expression of a gene or the role of the gene product. Given the existence of other genes found to interact with *NUBPL* through an open-source database of genomic and proteomic information on human gene, the second aim of this thesis used whole exome sequencing data to investigate the presence of variants within these and other potential interacting genes found in the literature (Rebhan et al., 1997).

4.1 Investigation of alternative splicing in *NUBPL*

The mechanism by which human genes are expressed is remarkably intricate, involving the separation of pre-mRNA introns from coding exons. Throughout the least two to three decades, the machinery involved in this process has become better understood along with the diversity of isoforms that are generated from alternative splicing events. Specifically, the mechanism of splicing is made possible by the precision of the spliceosome, and genomic alterations impacting this machinery have the potential to cause disease. In order to appropriately regulate alternative splicing and meet the functional requirements within the cells of different tissues, the splicesome needs to effectively identifying exons and ignoring pre-mRNA sequences (Wang et al., 2007). This process is achieved through the specific *cis*-acting elements that direct the spliceosome to the correct nucleotides in order to form the final mRNA product. Alternative splicing events take place in approximately 92-94% of human genes, resulting in distinct isoforms versus what would otherwise be identical mRNA products in each tissue. (Wang et al., 2008). Most of this variability is within open reading frames (80%), while the other variability is contained within the untranslated region (20%). As a whole, human gene mutations have the potential to impact alternative splicing events through various methods, such as altering the timing of expression of functionally diverse isoforms, modifying regulation of nonsense-mediated decay, and impacting other regulatory processes.

With the onset of whole exome and whole genome sequencing, our understanding of the diverse mechanisms by which these alternative splicing events are modified is growing rapidly. Alternative splicing is now understood to be one of the molecular mechanisms responsible for variable phenotypes among patients with similar molecular diagnoses. Many pathogenic mutations that are found through sequencing technologies have been shown to affect the

sequence regions necessary for splicing rather than the protein sequence that directly codes for structure. Mutations that are *cis*-acting represent greater than 15% of all disease-causing mutations and often occur by introducing a cryptic splice site. Interestingly, there have been many disease-causing mutations that have been shown to impact *trans*-acting splicing regions, potentially altering the expression of multiple genes (Wang, et al., 2007). As a whole, the growing understanding of how these splice mutations impact disease also demands further understanding of how these splice mutations modify alternative splicing in different individuals with more or less severe phenotypes.

In our study, potential alternative splicing events from a common branch-site mutation as well as a more rare splice mutation in the gene, *NUBPL*, are analyzed using RT-PCR and gel electrophoresis. These results are discussed below.

4.1.1 Differences in mRNA transcripts in three patients with c.815-27T>C branch site mutation

In previously discussed studies by Tucker et al. (see Introduction), the c.815-27T>C branch site mutation was studied via RT-PCR with gel electrophoresis and sequencing of PCR products (Tucker et al., 2012). Three distinct bands were observed from patient fibroblasts compared to one band that was observed for the control. After sequencing of these bands, results indicated that the c.815-27T>C leads to partial use of a cryptic acceptor site that has a higher predicted score, causing partial inclusion of intron 9. The other two bands corresponded to wild-type NUBPL transcript as well as transcript as a result of exon 10 skipping.

In our studies differences in mRNA transcripts were observed, particularly for an approximately 100 bp band that is seen in all three patients and not the control. In Figure 5, the brightest band in both the patient and control samples is expected to be full length transcript,

which would correspond to 180 bp. Due to errors in band migration from possibly uneven charge distribution in the gel box, this band appears to be slightly larger (200 bp) in comparison to the ladder. Given this, the smaller band that appears to be 100 bp band is likely shorter (approximately 85 bp) than what is observed. These results would therefore be consistent with the brightest, 180 bp band being the transcript containing exons 9, 10, and 11, while the shorter 85 bp band represents a transcript that has skipped exon 10. Band 3, which is approximately 200-250 bp, is brighter in the patient samples than in the control. In addition, band 4, which is approximately 300bp, is present in all three patient samples and not in the control. These bands may represent alternative transcripts that include portions of intron 9, however sequencing of these bands would be able to more accurately confirm this. Band 2 also has greater band intensity for the control and Patients 1 and 2 relative to Patient 3. Though band intensity cannot precisely quantify transcript levels, these differences in band intensity may represent preferential expression of this transcript over others. The other *in cis* gene mutation, c.166G>A, in Patient 3 is likely not impacting these differences given the mutation is significantly upstream of the branch site mutation. At this point without sequencing these transcripts, it remains unclear as to why there are differences in relative band intensity for Patient 3. From this gel electrophoresis of mRNA transcripts using primers encompassing the c.815-27T>C mutation, we confirm that additional transcripts are being observed in these patients relative to the control, one of which is likely the transcript previously published that involves exon 10 skipping (Tucker et al., 2012) Whether or not this specific isoform results in a nonfunctional NUBPL protein remains to be elucidated. Despite these differences, an important point to consider is that the control used in this analysis has not been genotyped. Given that 1 in 60 individuals in the population are heterozygous for the c.815-27T>C variant, excluding this variant in controls would be an

important step for future RNA studies involving this mutation. The same number of mRNA transcripts were seen for all three patients with different relative intensities, suggesting that alternative splicing from this mutation may contribute to the variation in complex I disease severity.

Interestingly, per the ExAc database, the affected base pair in the c.815-27T>C variant is included in 7 of the 9 transcripts in the *NUBPL* gene, implying that there are 2 isoforms that do not carry this variant. In the case of these three patients harboring this mutation, one reason that fibroblasts may not be exhibiting complex I deficiency is due to expression of these two isoforms that do not carry the mutation. In fact, in a study on alternative isoform regulation in human tissue, Wang et al. (2008) that among eight different types of alternative transcript events, 52% to 80% are regulated between tissues, supporting the hypothesis that alternative splicing is a fundamental contributor to phenotypic diversity and complexity in humans. In addition, this study and analyses performed by others on isoform variation have demonstrated that approximately 21% of alternatively spliced genes are affected by polymorphisms that modify the relative abundances of alternative isoforms (Nembaware et al., 2004). RNA functional studies across various tissues may provide insight on how variations in isoform regulation can affect disease pathology for those with constitutional splicing mutations.

4.1.2 Differences in mRNA transcripts in Patient 3 who is compound heterozygous for c.815-27T>C and a separate splice mutation, c.693+1G>A

While all three patients have the c. 815-27T>C mutation, only Patient 3 has a splice site mutation, c.693+1G>A, *in trans*. This mutation in combination with the c.815-27T>C and c.166G>A mutations *in trans* have been previously seen in an individual with complex I deficiency (Kevelam et al., 2013). This individual was among a study of several individuals

presenting with MRI abnormalities of the cerebellar cortex, deep cerebral white matter, and corpus callosum. These individuals had developmental delay and varied cognitive abilities from normal to severely affected. All of these patients demonstrated complex I deficiency in enzyme assays of muscle and fibroblast biopsy, ranging between 23% and 87% of the lowest reference value. Interestingly, results of muscle biopsy electron transport analysis on Patient 3 were "within normal limits," despite having a similar phenotype and the same *NUBPL* genetic variants. Similarly to the hypothesis for Patients 1 and 2, tissue-dependent alternative splicing may be responsible for the absence of complex I deficiency in Patient 3's muscle.

Using primers designed to encompass the c.693+1G>A mutation, differences in alternative splicing were interrogated by measuring mRNA transcripts on gel electrophoresis. Thermocycler settings were adjusted for varying annealing temperatures, showing an absence of a 1000 bp mRNA transcript for Patient 1 and 2 at an annealing temperature of 56^{o}C (Figure 11). At 58^{o}C and 60^{o}C, these 1000 bp transcripts were no longer present in the any of the patients or controls, indicating that this transcript may not have been specific for *NUBPL*. At 58^{o}C and 60^{o}C, no differences in mRNA transcripts were observed between patient and controls, however there were minor differences in relative band intensity (Table 7). Patient 3 in particular had a decrease in the relative intensity of the 260 bp transcript at all three temperatures in comparison to the control and Patients 1 and 2. In addition, an increase in the relative intensity of the approximately 700 bp transcript is observed in Patient 3 relative to control and patients. At this point, these minor differences in relative intensity for Patient 3 may be associated with preferential splicing for one transcript over another. However, quantification of these mRNA transcripts is difficult based on gel electrophoresis alone.

The c.693+1G>A mutation has been previously classified in ClinVar database as pathogenic. The variant is rare; within the ExAc database it was only observed in 1 of 120, 576 individuals. As this variant falls on 6 of the 9 transcripts in the *NUBPL* gene, tissue-specific isoform expression may explain why complex I deficiency was not seen in the patient's fibroblasts. With regards to gel electrophoresis results, there were no detectable differences in the number of fibroblast mRNA transcripts between the patients and control. This result may be due to tissue-specific isoform regulation, in which RNA functional assays in different tissues would further elucidate the possibility of alternative splicing in other tissues such as lymphocytes or cerebellar tissue.

4.1.3 Investigation of mRNA transcript differences due to a missense mutation, c.311T>C, in Patients 1 and 2

RT-PCR was performed on these patient and control samples using primers encompassing a c.311T>C mutation. This mutation was present *in trans* with the c.815-27T>C mutation in Patients 1 and 2 and was not present in Patient 3. In previous studies, an estimated 1.6% of disease-causing missense mutations have been predicted to affect splicing; in more recent publications, it has been suggested that approximately 7% of exonic variants in the general population may disrupt splicing, many of which are due to introduction of a cryptic splice mutation (Krawczak et al., 2007; Mort et al., 2014). These mutations have the ability to alter the strength of the natural splice site, shifting the balance of isoforms towards non-constitutive splice isoforms from normal mRNA transcripts. Given this, we investigated the c.311T>C mutation present in Patients 1 and 2 in order to interrogate possible differences in mRNA transcripts between these two patients, Patient 3, and controls.

Results from the RT-PCR using primers at exons 2 and 6 resulted in no differences in mRNA transcript sizes between all three patients and two controls. Relative band intensity was lower for the 300 bp transcript in Patients 1 and 2, which is the expected wild-type transcript that is also seen in Patient 3 and the controls. Similar to the observation of varying band intensities following RT-PCR using the 'c.693+1G>A' primers, these minor differences in relative intensity for Patients 1 and 2 may be associated with preferential splicing for one transcript over another. However, given that only a small portion of missense mutations impact splicing, we predict that this missense mutation does not likely cause alternative isoforms in these patients relative to the controls. To more accurately confirm this, sequencing of these NUBPL-derived transcripts from this c.311T>C mutation would be necessary.

4.2 Investigation of modifier genes in whole exome sequencing data

For several decades, modifier genes have been known to impact disease outcome. In the setting of a disease caused by a primary gene defect, it is often apparent that the range of clinical presentation associated with mutation in a single gene is broader than previously anticipated. Possible reasons behind clinical variation include environmental factors as well as modifier genes that can affect age of onset, severity of disease, or duration of the disease. Some examples of modifier genes include that *EPHA4* and *SMN2* genes which are modifiers of amyotrophic lateral sclerosis and spinal muscular atrophy (Van Hoeke et al., 2012; Kariya et al., 2012).

With the creation of clinical whole exome sequencing, identification of genetic modifiers has become more feasible through assessment of potential pathogenic variants in other genes that interact with the gene of interest. In the case of the three patients in this study, whole exome data was available for the two patients who are siblings. Since their diagnosis of complex I deficiency

in 2014, several genes have been studied to interact with NUBPL, particularly in the iron sulfur cluster synthesis pathway. Of at least 16 genes that have been studied in the iron sulfur cluster pathway, a heterozygous loss of function variant in one of these genes, *ISCA1*, was identified in both of the siblings but not the mother. As previously mentioned, homozygous pathogenic mutations have recently been reported in two families with multiple mitochondrial dysfunction syndrome, suggesting that the mutations in this gene can significantly impact iron sulfur cluster assembly critical for respiratory chain function. Given that the two siblings are heterozygous for a loss of function splice mutation in this gene, one possibility is that the *ISCA1* protein function, in combination with the *NUBPL* proteins defect, is partly impacted enough to down regulate or significantly inhibit iron sulfur cluster assembly.

In considering potential modifier genes for *NUBPL*, an important consideration to recall is that all three of the patients in this study showed no evidence of complex I deficiency on skeletal muscle biopsy. As previously discussed, differences in isoform regulation based on tissue specificity may be one reason for this observation, however an additional possibility includes the role of iron sulfur clusters in many critical enzymes such as DNA helicases, DNA polymerases, an ATP synthase (ABCE1). Given this involvement of iron sulfur clusters in key biological enzymes, it is postulated that mitochondrial function is therefore crucial for extra-mitochondrial mechanisms such as nucleotide and ribosomal protein metabolism (Stehling et al., 2014).

Given that the siblings have compound heterozygous mutations in *NUBPL* in conjunction with a heterozygous loss of function splice mutation in *ISCA1*, one may hypothesize that these genes may be impacting iron sulfur cluster assembly together, leading to aberrant uptake of these iron sulfur clusters in other apoproteins other than those that directly assemble

complex I. In fact, in the study by Sheftel et al (2012), depletion of ISCA and IBA57 proteins in HeLa cells impacted the citric acid cycle in four different enzymes, including the two other irons sulfur proteins, aconitase and succinate dehydrogenase (Sheftel et al., 2012). In addition, they observed a decrease in COX (cytochrome c oxidase) activity when ISCA and IBA57 protein were depleted in both human and yeast cells despite the fact that COX does not contain iron sulfur clusters. One argument as to why this occurred was due to the failure of complex I to assemble, eliciting an effect on COX and the electron transport chain (Mühlenhoff et al., 2011; Sheftel et al., 2012). These studies support the idea that the failure of iron sulfur clusters to incorporate into their specific aproproteins may have downstream regulatory effects on mitochondrial function, potentially impacting disease through alternative pathways that are not fully understood at this time.

4.3 Limitations to the Study

While performing experiments to study potential differences in mRNA transcripts, it is necessary to consider several limitations in order to qualify result implications. Firstly, in studies by Tucker et al, patient and control samples were treated with cyclohexamide in cell culture in order to inhibit nonsense mediated decay. (Tucker et al, 2012) As was shown in their results, additional mRNA transcripts could be differentiated as opposed to when the fibroblasts were not treated with cyclohexamide. For the experiments performed for this study, cyclohexamide was not used, which may have resulted in better resolution of bands, particularly those that appeared more dense on electrophoresis.

Secondly, fibroblasts from two healthy female controls were used in these studies, one 7 year-old female and one 16-year-old female. Both of these individuals were not genotyped,

which in the case of the common c.815-27T>C would be a necessary validation given 1 in 60 individuals in the population are heterozygous for this variant. In addition, several additional controls of individuals with and without this branch site variant would provide a better comparison for variation in mRNA transcripts.

Lastly, given that some of the mRNA transcripts seen on electrophoresis have a base pair length that does not correspond to a known transcript, isolation and sequencing of these transcripts would better differentiate splicing events that are occurring. In studies by Tucker et al, sequencing of individual bands lead to identification of specific splice sites and determination of a cryptic acceptor site score relative to the wild-type acceptor site score. For the mRNA transcripts in this present study, sequencing of these bands would be of particular interest given there were some transcripts corresponding to a longer length of unknown origin that were present in the patients and not the controls.

4.4 Future Studies

In light of these recent studies performed on three patients with a current diagnosis of complex I deficiency due to mutations in *NUBPL*, additional studies are warranted in order to further investigate the potential effects of alternative splicing. By performing RNA expression studies in different tissues, the question of tissue specific isoform regulation can be investigated, particularly if in the setting of multiple, genotyped controls. In collaboration with a team at Ambry Genetics, further RNA studies and genotyping of Patients 1 and 2 (siblings) and their family members will be implemented. Firstly, site specific testing will be performed on the unaffected sister in this family to determine if she carries either of the *NUBPL* mutations, contributing to the argument that these mutations together are more likely pathogenic if both are

not found in this unaffected sibling. In addition, site-specific testing for the *ISCA1* splice mutation (c.241+1G>T) will be performed on this unaffected sibling as well as the unaffected father to further elucidate the potential role of this gene in combination with alterations in *NUBPL*. Secondly, RNA isolated from fibroblasts and lymphocytes from the two siblings and their unaffected family members as well as cerebellar tissue from the two siblings will be studied using RNA sequencing (RNAseq), a next generation sequencing technique used to analyze the transcriptome. By using RNAseq, transcript assembly, alternatively splice transcripts, novel transcripts, and transcript quantification can be more precisely determined across different tissues in these patients and their family members. In addition, Ambry Genetics maintains several internal, genotyped controls for comparison of individuals who do not carry either the c.815-27T>C branch site mutation or c.311T>C missense mutation. These studies will further our understanding of how this branch site mutation impacts splicing activity within *NUBPL* and potentially characterize how these mutations affect tissue-specific alternative splicing.

4.5 Conclusion

In previous studies by Hall et al (2014), clinical whole exome sequencing provided a molecular diagnosis for two siblings with developmental delay, ataxia, and cerebellar abnormalities (Hall et al., 2014). Two compound heterozygous variants in the *NUBPL* gene were identified and were found to be associated with complex I deficiency based on previous reports of individuals with the same mutations and a similar phenotype. One of these mutations was found to be a missense mutation, c.311T>C, while the other was a common branch site mutation, c.815-27T>C. In addition to these two siblings, an additional unrelated individual was also found to have complex I deficiency; this patient was found to carry the same branch site mutation *in*

trans with two other mutations, a missense mutation (c.166G>A) and a splice mutation (c.693+1G>A). Despite having similar mutations and an overall similar phenotype, disease severity varies between each patient, where the younger of the two siblings appears to be less severely affected than the older sibling with regards to mobility and the unrelated patient is cognitively normal unlike the two siblings.

Using RT-PCR and primers designed to surround the branch site mutation, gel electrophoresis of transcripts revealed an additional transcript in all three patients relative to the control that likely represented a previously reported transcript that skips exon 10 in *NUBPL*. Additional longer transcripts were also observed in all three patients relative to the control, indicating that other mRNA transcripts formed possibly due to this branch site mutation. Implications of these alternatively spliced transcripts on NUBPL protein function remain to be elucidated. Investigation of the splice mutation found only in the unrelated patient (c.693+1G>A) resulted in little to no differences in mRNA transcripts, indicating that this mutation is not impacting splicing in fibroblasts, or at least not at the resolution that is obtained using gel electrophoresis. The missense mutation, c.311T>C, found *in trans* with the branch site mutation in the two siblings was also found to have little to no differences in mRNA transcripts relative to controls. Given that two of the nine NUBPL isoforms do not contain the c.815-27T>C branch site mutation and three NUBPL isoforms do not contain the c.693+1G>A splice mutation, there remains the possibility that alternative splicing is generating different isoforms in various tissues. This argument is also important when considering that each of these patients had normal complex I activity in muscle tissue, implying that different tissues may have distinctive isoforms due to alternative splicing events. Similarly, the effect of each mutation on tissue-specific alternative splicing may explain the variance in complex I disease severity between our patients.

In addition to these studies on mRNA transcripts, whole exome data from the two siblings (Patients 1 and 2) was re-analyzed and potential modifier genes affecting *NUBPL* were studied. A heterozygous loss of function variant in the gene, *ISCA1*, was identified in both of the siblings but not the mother. Similar to *NUBPL*, *ISCA1* is involved in late stage biosynthesis of iron sulfur clusters, suggesting genetic alterations in both of these genes may be jointly detrimental to this iron sulfur synthesis pathway that is critical for complex I assembly.

Overall, this study supports the importance of investigating how alternative splicing affects phenotypic variability among individuals with similar molecular diagnoses. Not only is alternative splicing a fundamental contributor to phenotypic diversity and complexity in humans, variable expression of isoforms can cause disease directly, modify the severity of the disease phenotype, or be linked with disease susceptibility. Studying these isoforms in different tissues of patients with similar molecular diagnoses would provide potential insight into more specific causes of variable expression of particular diseases.

Resources

1. Banci L, Bertini I, Ciofi-baffoni S, et al. Anamorsin is a [2Fe-2S] cluster-containing substrate of the Mia40-dependent mitochondrial protein trapping machinery. Chem Biol. 2011;18(6):794-804.

2. Barton RM, Worman HJ. Prenylated prelamin A interacts with Narf, a novel nuclear protein. J Biol Chem. 1999;274(42):30008-18.

3. Berg JM, et al. Biochemistry. 8th edition. New York: W H Freeman; 2015.

4. Brett, D., et al. Alternative splicing and genome complexity, Nat Genet, 30 (2002) 29-30.

5. Bych K, Kerscher S, Netz DJ, et al. The iron-sulphur protein Ind1 is required for effective complex I assembly. EMBO J. 2008;27(12):1736-46.

6. Calvo SE, Tucker EJ, Compton AG, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. Nat Genet. 2010;42(10):851-8.

7. Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J, Walker JE. Bovine complex I is a complex of 45 different subunits. J Biol Chem 2006;**281**:32724-7.

8. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP: Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; 7: e46688.

9. Chow, J. *et al.* Mitochondrial disease and endocrine dysfunction *Nat. Rev. Endocrinol,* 2016. [Diagram]

10. Distelmaier F, Koopman WJ, Van den heuvel LP, et al. Mitochondrial complex I deficiency: from organelle dysfunction to clinical disease. Brain. 2009;132(Pt 4):833-42.

11. Fassone E, Rahman S. complex I deficiency: clinical features, biochemistry and molecular genetics. J Med Genet. 2012;49(9):578-90.

12. Golden Helix VarSeq ® visualization tool [Software]. Bozeman, MT: Golden Helix, Inc. Available from http://www.goldenhelix.com.

13. Gonzalez-Perez A, Lopez-Bigas N: Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011; 88: 440–449.

14. Gregory BL, Cheung VG. Natural variation in the histone demethylase, KDM4C, influences expression levels of specific genes including those that affect cell growth. Genome Res. 2014;24(1):52-63.

15. Hall, Katherine. (2014). One Family's Story: How exome sequencing opened the door to understanding and research. UC Irvine: Genetic Counseling.

16. Holt, I.J., Harding, A.E., Petty, R.K. and Morgan-Hughes, J.A. (1990) A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. Am. J. Hum. Genet., 46, 428–433.

17. Huang J, Song D, Flores A, et al. IOP1, a novel hydrogenase-like protein that modulates hypoxia-inducible factor-1alpha activity. Biochem J. 2007;401(1):341-52.

18. Kariya S, Re DB, Jacquier A, Nelson K, Przedborski S, Monani UR. Mutant superoxide dismutase 1 (SOD1), a cause of amyotrophic lateral sclerosis, disrupts the recruitment of SMN, the spinal muscular atrophy protein to nuclear Cajal bodies. Hum Mol Genet. 2012;21(15):3421–3434.

19. Krawczak M, Thomas NS, Hundrieser B, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat.* 2007;28(2):150–158.

20. Mckenzie M, Tucker EJ, Compton AG, et al. Mutations in the gene encoding C8orf38 block complex I assembly by inhibiting production of the mitochondria-encoded subunit ND1. J Mol Biol. 2011;414(3):413-26.

21. Merlo CA, Boyle MP. Modifier genes in cystic fibrosis lung disease. J Lab Clin Med. 2003;141(4):237-41.

23. Modrek, B., Lee, C. A genomic view of alternative splicing, Nat Genet, 30 (**2002**) 13-19.

24. Mort M, Sterne-Weiler T, Li B, et al. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol.* 2014;15(1):R19.

25. Nagalakshmi, U., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320, 1344–1349.

26. Nembaware, V., Wolfe, K. H., Bettoni, F., Kelso, J. & Seoighe, C. Allele-specific transcript isoforms in human. *FEBS Lett.* 577, 233–238 (2004).

27. Netz, D. J., A. J. Pierik, M. Stümpfig, U. Mühlenhoff, and R. Lill. 2007. The Cfd1-Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol. Nat. Chem. Biol. 3:278-286.

28. Nussbaum, R. L., McInnes, R. R., Willard, H. F., Hamosh, A., & Thompson, M. W. (2007). *Thompson & Thompson genetics in medicine*. 7th Edition. Philadelphia: Saunders/Elsevier.

29. Parikh S, Saneto R, Falk MJ, et al. A modern approach to the treatment of mitochondrial disease. Curr Treat Options Neurol. 2009;11(6):414-30.

30. Pierce, Benjamin. *Genetics: A Conceptual Approach*. Vol. 1. W.H. Freeman and Company, 2008.

31. Rahman S, Thorburn D. Nuclear Gene-Encoded Leigh Syndrome Overview. 2015 Oct 1. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews. Seattle (WA): University of Washington, Seattle; 1993-2017.

32. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (April 1997). "GeneCards: integrating information about genes, proteins and diseases". *Trends Genet*. 13 (4): 163.

33. Rossignol, R., Letellier, T., Malgat, M., Rocher, C., and Mazat, J.-P. (2000). Tissue variation in the control of oxidative phosphorylation: implication for mitochondrial diseases. Biochem. J. 347, 45–53.

34. Ruitenbeek W, Wendel U, Hamel BC, Trijbels JM. Genetic counselling and prenatal diagnosis in disorders of the mitochondrial energy metabolism. J Inherit Metab Dis 1996;19:581 - 587.

35. Sheftel AD, Stehling O, Pierik AJ, et al. Human ind1, an iron-sulfur cluster assembly factor for respiratory complex I. Mol Cell Biol. 2009;29(22):6059-73.

36. Sheftel AD, Wilbrecht C, Stehling O, et al. The human mitochondrial ISCA1, ISCA2, and IBA57 proteins are required for [4Fe-4S] protein maturation. Mol Biol Cell. 2012;23(7):1157-66.

37. Shoubridge EA. Nuclear genetic defects of oxidative phosphorylation. Hum Mol Genet. 2001;10(20):2277-84.

38. Shukla A, Hebbar M, Srivastava A, et al. Homozygous p.(Glu87Lys) variant in ISCA1 is associated with a multiple mitochondrial dysfunctions syndrome. J Hum Genet. 30 March 2017.

39. Smeitink JA, Zeviani M, Turnbull DM, Jacobs HT. Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders. Cell Metab. 2006;3(1):9-13.

40. Smeitink, JA, et al. *Oxidative Phosphorylation in Health and Disease*. Kluwer Academic/Plenum Publishers, 2004.

41. Stehling O, Wilbrecht C, Lill R. Mitochondrial iron-sulfur protein biogenesis and human disease. Biochimie. 2014;100:61-77.

42. Tatuch, Y., Christodoulou, J., Feigenbaum, A., Clarke, J.T., Wherret, J., Smith, C., Rudd, N., Petrova-Benedict, R. and Robinson, B.H. (1992) Heteroplasmic mtDNA mutation (T---G) at 8993 can cause Leigh disease when the percentage of abnormal mtDNA is high. Am. J. Hum. Genet., 50, 852–858.

43. Tucker EJ, Mimaki M, Compton AG, Mckenzie M, Ryan MT, Thorburn DR. Next-generation sequencing in molecular diagnosis: NUBPL mutations highlight the challenges of variant detection and interpretation. Hum Mutat. 2012;33(2):411-8.

44. Tucker EJ, Compton AG, Calvo SE, Thorburn DR. The molecular basis of human complex I deficiency. IUBMB Life. 2011;63(9):669-77.

45. U.S. National Library of Medicine, National Institutes of Health website. "Rare Diseases." http:// www.nlm.nih.gov/medlineplus/rarediseases.html. Accessed January 4, 2016. 2. World Health Organization website.

46. Van Hoecke A, Schoonaert L, Lemmens R, Timmers M, Staats KA, Laird AS, Peeters E, Philips T, Goris A, Dubois B, Andersen PM, Al-Chalabi A, Thijs V, Turnley AM, van Vught PW, Veldink JH, Hardiman O, Van Den Bosch L, Gonzalez-Perez P, Van Damme P, Brown RH, Jr, van den Berg LH, Robberecht W. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models in humans. Nat Med. 2012;18(9):1418–1422

47. Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* 515, 80–84 (2014)

48. Wang, ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470-6.

49. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. 2007;8(10):749-61.

50. Wilhelm, B.T., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. 453, 1239–1243.

51. Wittung-Stafshede, P. 2002. Role of cofactors in protein folding. Acc. Chem. Res. 35:201–208.

52. Wydro MM, Balk J. Insights into the pathogenic character of a common NUBPL branch-site mutation associated with mitochondrial disease and complex I deficiency using a yeast model. Dis Model Mech. 2013;6(5):1279-84.