UNIVERSITY OF CALIFORNIA,
IRVINE

One Family's Story:
How Exome Sequencing Opened the Door to Understanding and Research

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Genetic Counseling

by

Katherine Anne Hall

Thesis Committee:
Professor Moyra Smith
Professor Virginia Kimonis
Assistant Adjunct Professor Elizabeth Chao

2014

© 2014 Katherine Anne Hall

DEDICATION

For Bryce and Parker,

who make my heart happy every single day.

I love you.

TABLE OF CONTENTS

	Page
LIST OF SYMBOLS	iv
LIST OF FIGURES	v
LIST OF TABLES	vi
ACKNOWLEDGEMENTS	vii
ABSTRACT OF THE THESIS	viii
INTRODUCTION	1
CHAPTER 1: Materials and Methods	7
Gathering Clinical and Testing Records	7
Case Reports	7
Medical History of Sibling B	12
Differences between the Siblings	13
Family History	14
Whole Exome Sequencing	15
Mitochondrial Respiratory Chain Enzyme Analysis	16
Pipelines used in the Analysis of the Data to Explain Phenotypic Differences in the Siblings	16
Analysis of Modifier Genes for Assessment of Increased Severity in Sibling A	19
CHAPTER 2: Results	21
Whole Exome Sequencing	21
Mitochondrial Enzyme Testing	22
Further Analysis to Determine Clinical Variations in Siblings	23
CHAPTER 3: Discussion	26
Complex I (CI) and *NUBPL* Mutation Basics	26
Tools for *NUBPL* Mutation Identification	29
Bioinformatic Pipelines	30
Genetic Modifiers Can Change Phenotypes	32
Implications	35
REFERENCES	37
APPENDIX A: Sibling A Testing by Year and the Results of This Testing	39
APPENDIX B: Sibling B Testing by Year and the Results of This Testing	42

LIST OF SYMBOLS

ADEM	acute disseminated encephalomyelitis
AFO	ankle foot orthosis
AVA	Ambry Variant Analyzer
CBC	complete blood count
CI	Complex I
CLIA	Clinical Laboratory Improvement Amendments
CSF	cerebrospinal fluid
DNA	deoxyribonucleic acid
EEG	electroencephalography
EMR	electronic medical record
ESP	Exome Sequencing Project
ETC	electron transport chain
HGMD	Human Gene Mutation Database
IRB	Institutional Review Board
MPS IIID	mucopolysaccharidosis IIID
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
NICU	neonatal intensive care unit
VLDLR	very low density lipoprotein receptor

LIST OF FIGURES

		Page
Figure 1.	Structure of a Chromosome, including Introns and Exons within a Gene	3
Figure 2.	Processing of DNA Encoded Information into Protein via an mRNA Transcript	4
Figure 3.	Sagittal Brain MRI Images of Siblings A and B	10
Figure 4.	Pedigree of the Family	14
Figure 5.	Sibling A and B: Chromatogram of Mutation Found in *NUBPL* Gene	22

LIST OF TABLES

Page

Table 1. Candidate Gene List for Focused Review by Seven Bridges Genomics Pipeline 17

Table 2. Genetic Models for Modifier Genes 20

Table 3. *NUBPL* Mutation Data in Sisters 22

Table 4. Mitochondrial Respiratory Chain Enzyme Analysis (ETC) – Skeletal Muscle 23

Table 5. Inheritance Model Interesting Mutations 24

Table 6. Documented Cases of Complex I Deficiency Due to *NUBPL* Mutations 28

ACKNOWLEDGEMENTS

A special thank you to Dr. Moyra Smith, Dr. Virginia Kimonis, and Dr. Elizabeth Chao for their help with this thesis. In addition, thank you to Pamela Flodman, Kathryn Steinhaus French, and Kathryn Singh for their guidance and support. Thank you to Grace Kim and Elizabeth Koos for their technical help and encouragement. And thank you to my boys and the rest of my family and friends without whom this would not have been possible.

ABSTRACT OF THE THESIS

One Family's Story:
How Exome Sequencing Opened the Door to Understanding and Research

By

Katherine Anne Hall

Master of Science in Genetic Counseling

University of California, Irvine, 2014

Professor Moyra Smith, Chair

Often with rare genetic diseases, families must endure a frustrating, expensive, and exceptionally lengthy course to find an etiology for the clinical symptoms found in a patient – the diagnostic odyssey. This study examines the impact of exome sequencing on finding the cause of two siblings' global delay and ataxia and the treatments and research resulting from the diagnosis. Included is a glimpse at how pipelines can influence diagnosis.

For this study, the medical history of two affected siblings was used to consider the exome sequence results with regard to clinical symptoms and mitochondrial respiratory chain enzyme analysis. Due to the differing phenotypes of the siblings, the exome data was scoured for possible modifier genes to explain this variation. Finally, the FASTQ files were processed through an alternate pipeline to determine if filtering can play a role in discordant variant calling.

Exome testing diagnosed Complex I deficiency. Each girl had one allele with a common mutation in the *NUBPL* gene and a novel mutation in trans. Neither sibling had a polymorphism in cis with the common mutation. These are the first reported symptomatic patients with

compound heterozygous mutations lacking the polymorphism in cis with the common mutation. The siblings were started on treatment, and clinical trials have been appropriately examined for qualifications. In the more affected sibling, both pipelines found two mutations in the *SYNE1* gene, a gene associated with spinocerebellar ataxia. A second pipeline found only one mutation in each sibling's *NUBPL* gene, therefore missing this diagnosis.

INTRODUCTION

As technology races forward, physicians have access to increasingly technical and expensive tests to utilize in diagnosing patients with rare or atypical clinical presentations of conditions. In genetics, we hope to spare patients and families the dreaded "diagnostic odyssey" by employing screening and then diagnostic tests in a logical order. Each test in the pathway to diagnosis requires money, time, and interpretation, leading to weeks, months, and even years of time spent by lab technicians puzzling over data, doctors contemplating next steps, and families hoping and then being disappointed in finding no resolution. Thus, the "diagnostic odyssey" is the situation we challenge ourselves, and our field, to render obsolete.

This thesis focuses on one family's quest to explain what is happening to their children and why. The story is unique in many ways, but it allows exploration of familiar questions from a technical and genetic standpoint. In this thesis, I investigate the utility of whole exome sequencing as a diagnostic tool when other tests have failed to give an answer, examine results from testing that do not appear as predicted with an exome diagnosis *a priori,* and compare bioinformatics pipelines in variant calling. I also discuss how the affected siblings differ in presentation and use exome sequencing to identify modifier genes.

Exome testing has found a comfortable niche with clinicians who find themselves faced with a dilemma. Often, patients present clinically as though they have underlying genetic conditions. In these situations, the indications are somewhat general and not well reported in the literature, leaving clinicians to wonder what testing should be offered. Although many patients are now covered by insurance, the insurance providers may decline many forms of genetic testing without a convincing appeal letter full of previous negative test results. In requesting

testing, it must be clear what the health care provider is thinking and how the testing will rule in or out many differential diagnoses.

Clinicians must then evaluate the various tests available to them, considering the information that can be gleaned from the results and whether they are financially viable options. Karyotypes assume that the etiology of conditions stem from a visible rearrangement of portions of chromosomes. Such testing will cover a wide region of chromosomes with many megabases of genetic material, usually including many genes. Microarray testing also covers larger portions of chromosomes, but at a finer scale than the karyotype (i.e., kilobases). This type of test relies on differences in the copy number of regions of chromosomes. In other words, these changes in chromosomes will only be detected if portions of the chromosomes are missing or duplicated. When there is a family history of the condition, linkage analysis is a possibility to narrow the field of vision to find an area in a chromosome upon which to look for abnormalities. As we have come to understand single nucleotide polymorphisms, this ability to focus on a specific region has become better, but it still likely includes many genes. When consanguinity is suspected, looking for regions of homozygosity may give a clue to a sector of a chromosome where a mutation may be located. If the condition is due to a monogenic point mutation, the gene is known, and the phenotype is well documented and correlated with the patient's clinical presentation, then sequencing that gene is a reasonable practice. (Gilissen, 2011)

However, it is common to be faced with any one of these pieces of information missing, and running all possible tests can easily become draining in terms of time and money. To find a gene as efficiently as possible, whole exome sequencing can provide a way to scan the genetic material without necessarily having a particular target gene in mind. Not all deoxyribonucleic acid (DNA) of a cell is actually transcribed and translated into functional proteins. The functions

of many regions are not yet well understood, and it is unclear if they are used in forming proteins. Whole exome sequencing attempts to "read" the genetic code of the protein coding regions of the whole genome, known as exons (Figure 1). It focuses on areas known to make proteins, with the assumption that most conditions with clinical phenotypes are due to gain or loss of function of the proteins made by these portions of DNA sequence.

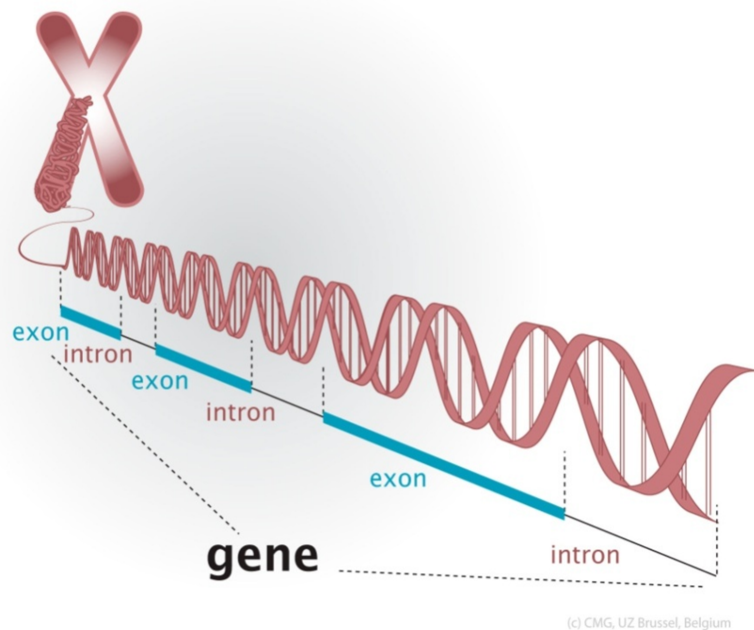

Figure 1. Structure of a Chromosome, including Introns and Exons within a Gene
(Centrum Medische Geneticz UZ Brussel, 2012)

Within each gene are stretches of DNA used by the cell for transcription. The transcript is a type of nucleic acid, known as messenger ribonucleic acid (mRNA), which makes it possible for the code stored within the DNA to migrate outside of the nucleus and be processed further to provide instructions for forming a protein. However, not all of the nucleotides in the gene will be used to make these transcripts or be contained in the final transcript. Alternating sequences known as introns and exons play an important role in the transcription process, but only the information coded in the exons will exit the nucleus as part of an mRNA transcript. Although the

complete function of introns is not always clear, it is well known that these stretches of sequence can play an important part in splicing together the exons from which the final mature transcripts will be formed. The introns provide the scaffolding for the exonic regions to be cut and put back together. Because these intronic regions play such an integral role in the formation of mRNA transcripts and, therefore, the formation of proteins, whole exome sequencing will often capture and read small stretches of sequence at the junctions of introns and exons (Figure 2).

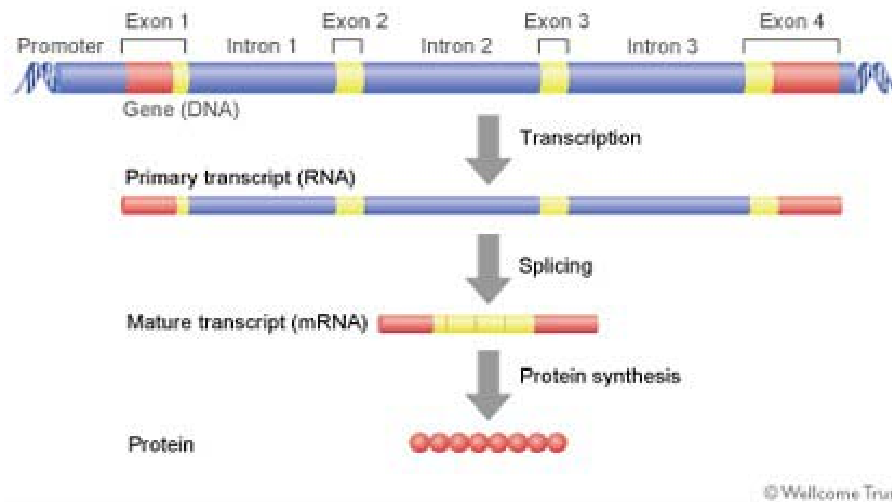

Figure 2. Processing of DNA Encoded Information into Protein via an mRNA Transcript
(Wellcome Trust, 2003)

With approximately 20,000 genes within the whole genome, each consisting of multiple exons, the resources to undertake such a project sound quite daunting; however, with new high-throughput technology, it is possible to sequence large amounts of exome data in short periods of time with costs lower than ever. Many companies are now offering whole exome sequencing for patients, and the field is quickly becoming competitive with companies trying to return the most accurate reports possible.

Sequencing thousands of genes returns multitudes of raw data. This data in and of itself is not usually helpful to the clinician. The data must be narrowed down to a manageable amount so

the time spent in analysis is useful to the patient and the physician; however, one does not want to overlook possible data that may not be well understood to cause disease but could contain a candidate diagnosis for the patient. This process requires a fine balance of careful thought, research, and planning. Laboratories often design a unique bioinformatics pipeline through which to run raw sequence data files. This pipeline creates minimum parameters the data must meet to continue through the data analysis process. These qualifications can be set in any order and chosen based on how inclusive or exclusive the candidate gene list should be. The level of inclusivity may be based on whether the sequencing is done for a clinical or research focus. Differences in these parameters can affect the candidate gene output at the end of the pipeline process.

The possibilities of chosen criteria are endless, and the order of each restriction in the pipeline can matter. First, the smaller reads of sequence must be assembled into longer strings that can overlap and give an overall read of the entire stretch of each of the many exons. To do this, small reads must be aligned to a reference sequence to make sure the small reads are in order in the right regions of the chromosomes. Many labs may choose to use the most updated human genome reference build. Some labs, such as those that sequence exomes frequently, may choose to assemble their own reference sequence. The choice of reference sequence will affect whether each individual read will be kept and analyzed based on the degree of alignment of the read. This, in turn, will affect how any variant will be called later. For instance, if many of the reads are thrown out because they do not align well enough with the reference sequence, then the particular gene with the variant call may not make the candidate gene list to be later scrutinized by the lab director. This is one way a possible diagnosis could be missed. Another case in point is order of constraints. Some examples of this could include whether some intronic sequence is

read to catch the ends of exons, when the intronic region is filtered out, and minimum limits for read depth or even maximum limits for allele frequencies in the patient's ethnic population, as well as the overall general population.

CHAPTER 1: MATERIALS AND METHODS

Gathering Clinical and Testing Records

Medical records were obtained with permission of the children represented by their parents. The electronic medical record (EMR) of the various institutions where the siblings were seen was accessed for the most recent information. The children's charts were read and detailed within the scope of the Institutional Review Board (IRB). The family was seen in clinic, and some communication occurred via e-mail regarding the parents' recollections of the history of the girls. All of this data was compiled to understand the diagnostic journey of the family.

Case Reports

The summary of the diagnostic odyssey is quite long (Appendix A). Sibling A is a 15-year-old girl with cerebellar atrophy and ataxia who came to be seen for an evaluation in the Genetics Clinic. She was born to a 20-year-old G1, P0->1 mother, vaginally at 41 weeks gestation. Birth weight was 8 pounds. Apgars were 5 and 9. There was perinatal complication resulting from nuchal cord, requiring oxygen. During pregnancy, amniocentesis was obtained because of increased risk for Down syndrome. Fetal movements were normal. The mother had a 55-pound weight gain during pregnancy.

The child began to fail to reach developmental milestones after the age of 3 months. At 9 months, she was unable to sit with support and had poor head control. She was unable to crawl and could not pull to stand. By 12 months, she was able to roll from back to front with delayed rolling to back and was babbling. She stood independently at 30 months and walked at 4 years. She said her first words at 30 months and is now making five-word sentences. She was toilet trained at 8 years and now is dry during the day. She attends high school and is in special classes

in the 10th grade. Additionally, she uses the swimming pool and has done horseback riding in the past.

At 3 months of age, she started having tremulousness and increased rigidity. The initial electroencephalography (EEG), which was done at 7 months shortly after the tremulousness started, appeared normal. During her ninth month of life, she had episodes where she would raise her arms and extend them back in a stiff position with her eyes rolling back and her mouth open. She was sent to follow up with neurology and subsequently admitted for further evaluation. The magnetic resonance imaging (MRI) obtained at 9 months showed diffuse abnormal signals from the cortex of the cerebellar hemisphere. The significance of the cerebellar findings was unclear. She was evaluated for metabolic and lysosomal disease. Testing included serum amino acids, urine organic acids, and lysosomal storage disorders to rule out Tay-Sachs and other conditions. Other potential testing in the future could be to rule out Niemann-Pick, Gauche disease, and perioxisomal disorders. An ophthalmology evaluation and video-EEG were obtained. Spinal tap protein and complete blood count (CBC) were normal. Urine organic acids showed glucuronide conjugates commonly seen in patients with valproate and chloral hydrate.

The MRI findings showed diffuse abnormality of cerebellar cortices bilaterally with normal white matter and normal supratentorial appearance. The differential considered was acute disseminated encephalomyelitis (ADEM) secondary to previous vascular infection. A cerebellar biopsy was obtained at 1 year of age. Histology was reviewed. This cerebellar tissue revealed a population of large, often swollen, Purkinje cells. The internal granular layer showed a decrease in the number of cells with many fine bluish granules, most consistent with nuclear pyknosis and karyorrhexis. Viral inclusions were not found. There was no evidence of a lysosomal storage disease, nor were mitochondrial morphological abnormalities seen. The histology was not typical

of Lhermitte-Duclos disease. The lesion on the cerebellum was not considered a hamartoma. The head bobbing and eye rolling reportedly improved at that time.

The Mayo Clinic Child and Neurology Service evaluated Sibling A when she was approximately 2 years old. Notably, she had significant problems with her balance and irregular eye movement and hand movement until the age of 12 months. Subsequently, she stabilized but remained developmentally delayed. It was also noted that all of her testing, including serum and urine amino acids, organic acid, methylmalonic acid, homovanillic acid, and mitochondrial DNA analyses and leukocyte lysosomal enzyme assay, between the ages of 1 and 2 years old were normal. A neurological examination at age 2 years indicated that she was able to point to her nose and eyes with her fingers with some ataxia. Optic disks were normal, and there was no retinal degeneration. There was occasional titubation of her head and trunk, and her stance was wide based. Other disorders in the differential included Angelman, Rett, ataxia, telangiectasia, and Wilson's disease. A repeat genetic consultation, repeat lumbar puncture, serum alpha-fetoprotein, and serum ceruloplasmin mutation analysis for Angelman and Rett syndrome were recommended. The parents report that after consultation locally, testing for these disorders was not yet pursued.

A repeat MRI at 11 months showed diffuse T2 prolongation throughout the cerebellar hemispheres with irregular enhancement within the cerebellum (Figure 3). This was not considered a typical appearance of a demyelinating disease. Exclusion of a metabolic abnormality was recommended. The findings were not thought to be suggestive of an inflammatory process. At this time, a geneticist who evaluated her sent blood for mitochondrial studies. No deletions or well-known point mutations were revealed. This geneticist also noted bobbing head movements and unusual eye movements. The MRI was sent for further evaluation,

and the opinion was that the changes were consistent with Joubert syndrome. Joubert syndrome is marked by abnormal ocular movement, hypotonia, respiratory difficulties, cognitive deficit, hypoplasia of the cerebella vermis, and malformation of the brain stem. Other features of Joubert syndrome include cystic kidneys, retinal dystrophy, hepatic fibrosis, and ataxia. The brain malformation leads to the molar-tooth sign on the cranial MRI, which is the hallmark of Joubert syndrome. This was definitely not noted on the brain MRI.

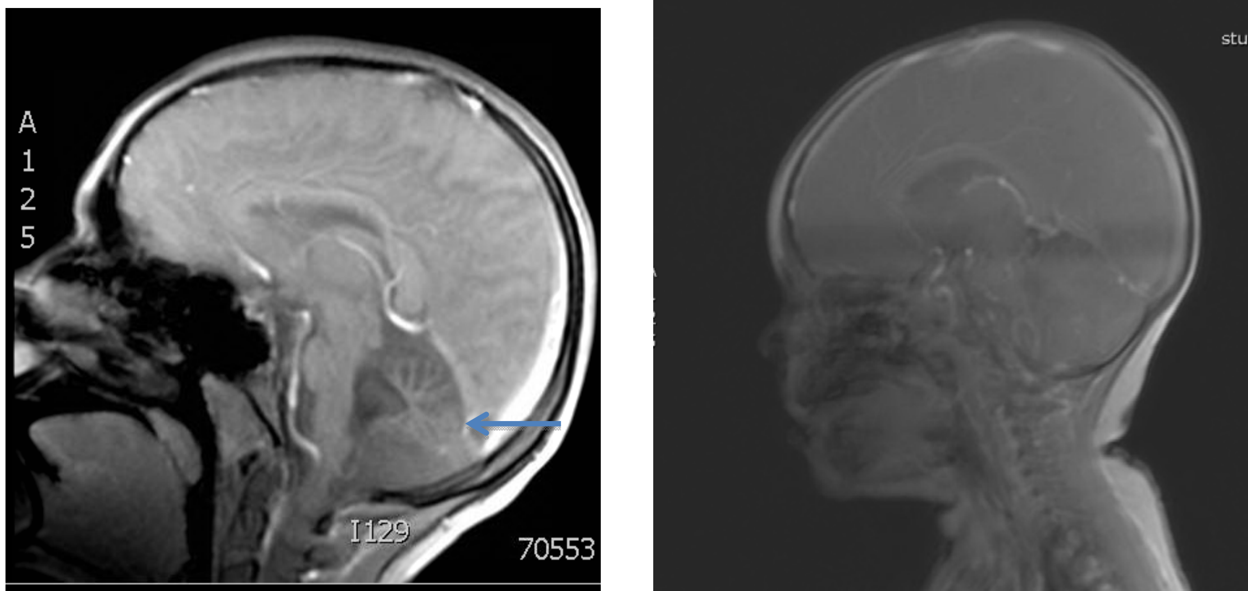

Figure 3. Sagittal Brain MRI Images of Siblings A and B. Features of note include severe atrophy and abnormal signaling pattern of the cerebellum in Sibling A (arrow). The pons is small. No signal abnormalities of the cerebral hemispheres and brainstem are noted. Sibling B has signal abnormalities of the cerebellum but does not have the severe atrophy noted in her sister. Her brainstem and cerebral cortex is spared.

Close to her third birthday, Sibling A was also noted to have developed microcephaly and presented a picture of cerebral palsy with hypotonia proximally in the upper and lower extremities with increased tone and tight heel cords in the lower extremities and upgoing toes. The recommendation was for ankle foot orthosis (AFO). Ophthalmologic examinations beginning at 5 years of age have been unremarkable.

For her clinical features of ataxia, a diagnosis of spinocerebellar ataxia was suggested. Another brain MRI was performed at 6 years of age, which showed severe global prominence of the sulci in the right and left cerebella line and T2 prolongation. The middle cerebellar, superior, and inferior peduncles were small. The pons was slightly small. The medulla and cervical cord and cranial vertebral junction were normal, as were the ventricles and the basal cisterns. A high-resolution T2-weighted image demonstrated no heterotopic gray matter, and the optic nerve was normal.

She was started on Carnitor and coenzyme Q10 for presumed mitochondrial dysfunction. Fours years later, the parents report no significant differences; however, it is felt that the nystagmus and ataxia of the hand was slightly improved.

By 11 years of age, the brain MRI findings showed diffuse hypoplasia/atrophy. The vermis was absent. The fourth ventricle was enlarged and had an unusual shape and communicated with a large cisterna magnum. The cerebellar peduncle was slightly elongated with a mild increase in depth of the interpeduncular fossa. There were normal spine MRI findings. At clinical evaluation, craniofacial examination did not reveal any dysmorphic features. She has ataxia of the hands, especially noted during the finger-nose testing. She has significant club feet with medial deviation of the medial malleolus. Her tone was normal in all extremities, and she was able to walk with her hands held. No nystagmus was noted. She has hyperreflexia, especially of the legs, with clonus and a positive Babinski test.

Beginning at the age of 11 years, a battery of new testing began. The first set of tests 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. These tests yielded normal results. As a result of parental

research, very low density lipoprotein receptor (VLDLR) associated cerebellar hypoplasia was suggested for molecular testing. This disorder is characterized by nonprogressive congenital ataxia that is predominantly truncal, resulting in delayed ambulation, cognitive deficits, dysarthria, strabismus, and seizures. Diagnosis is based on clinical, MRI findings, and molecular testing of the VLDLR gene. Again, no mutations were found. Other disorders tested included congenital disorders of glycosylation associated with developmental delay, hypotonia, subsequently hypoglycemia, and protein losing enteropathy. An entire ataxia sequencing panel was completed with no positive results for a mutation.

Medical History of Sibling B

Sibling B's history is similar to her sister's, but somewhat milder. She also presented with ataxia, developmental delay, cerebellar hypoplasia, and pontine hypoplasia.

She was born to a 31-year-old G3, P2 mother at term via vaginal delivery. Birth weight was 9.3 pounds, and length was 21.5 inches. There was no exposure to teratogens. Although the umbilical cord was around her neck and resuscitation was required, there was no neonatal intensive care unit (NICU) admission. The mother had a 40-pound weight gain during pregnancy. The triple screen was normal, and amniocentesis was not required.

Developmentally, she gazed at her parent's face at 2 months, smiled at 6 to 8 weeks, reached for objects at 3 months, rolled at 5 months, and sat at 15 months. She is walking and running and is very sociable. She has been saying "mom" and "dad" since 5 months. At 3 months of age, like her sister, Sibling B developed jerky eye movements and started shaking her head as if she was saying "No," especially when she was tired. She was evaluated at 10 months for cerebellar symptoms and was noted to have truncal hypotonia. Clinical examination revealed horizontal nystagmus, and no dysmorphic features were noted. She had

hyperreflexia in the arms and legs without clonus. She had mild tremor on reaching out for objects and exhibited side-to-side head shaking. She was suspected to have spinocerebellar ataxia. An MRI and EEG were arranged, and she was started empirically on carnitine and coenzyme Q10 for presumed mitochondrial dysfunction. Her EEG was normal.

MRI of the brain was obtained at 1 year of age and compared with that of her sister's taken at 11 years of age. The cerebellum showed diffuse increase in T2 signaling and mild diffuse hypoplasia. The cerebellar peduncles and pons were preserved. The lateral and third ventricles were normal. The fourth ventricle was mildly prominent. An MRI of the lumbar and thoracic and cervical spine was normal. A summary of testing is included in Appendix B.

Differences between the Siblings

It is interesting to note the significant differences between the sisters reported by their parents. We never underestimate parents' insights because they are usually with their children every day and have great insight into the subtleties of their children's actions and demeanors. A seemingly small detail can unlock the door to choosing an appropriate testing pathway. Differences occurred in all global areas of development. Cognitively, Sibling A has been "extremely delayed in academics." At the age of 15 years old, she does not express understanding of colors, shapes, numbers, and letters. She drags her feet when she walks so much that holes are worn in the toes. In addition, she has a hard time with opening and closing her hands, and her ataxia is quite apparent. Finally, she is easily irritated by touch and noise.

Sibling B, in contrast now 5 years old, "knows more" but is considered mildly delayed. Physically, she has a better gait and very little issue with ataxia in her daily life. She seeks hugs and affection and has moments of purposefully testing her upper limits of noise production.

Family History

The family history was reviewed (Figure 4). Both parents are of German ancestry and consanguinity was denied. The middle unaffected female sibling, age 8 years, has not had any genetic testing and has been deferred until she is of an age when she can give informed consent. Of note is the history of the paternal great grandmother having shaking of her hands. It is not known if she was formally diagnosed with Parkinson's disease or essential tremor. The father was recently evaluated for reported essential tremor; however, neurological examination was unremarkable. In addition, the siblings' parents had one miscarriage together, and the paternal grandmother had four.

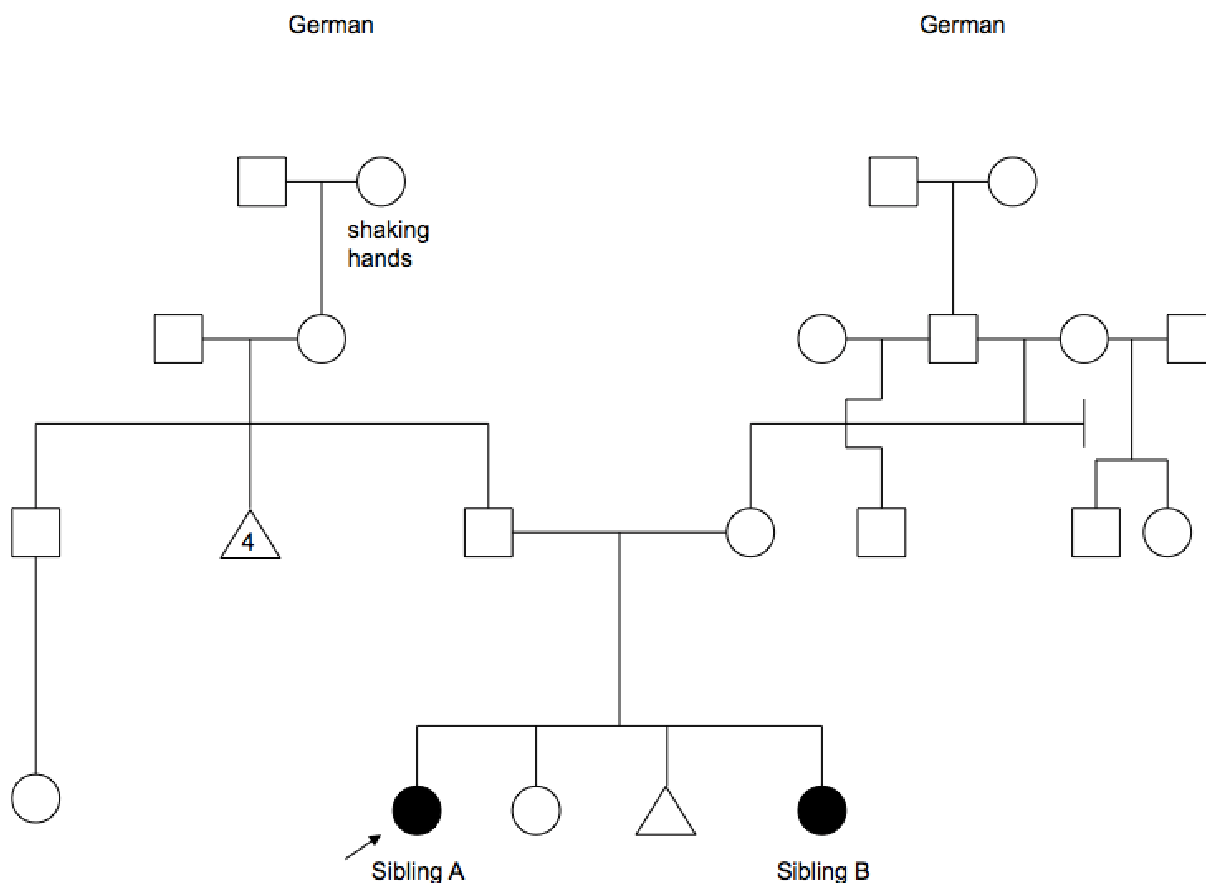

Figure 4. Pedigree of the Family.
Circles indicate females and squares males. The arrow indicates the proband Sibling A.

Whole Exome Sequencing

The exome testing was requested on October 25, 2012, when the two affected sisters, Sibling A and Sibling B, were 12 and 3 years old, respectively. Testing also included the parents. A first-tier with reflex to clinical diagnostic exome test was ordered through Ambry Genetics. This test first analyzes the approximately 4,000 genes listed in Human Gene Mutation Database (HGMD). If no positive variant cause is found, the reflex is to the much larger scale of approximately 20,000 genes. Filtering of variants was based on varying types of inheritance. Based on the family history, emphasis was placed on an autosomal recessive model of the proband, Sibling A. This filtering is suited for a clinical laboratory situation. Subsequently, a manual review of candidate variants was performed to exclude sequencing artifacts and polymorphisms. Finally, medical interpretation was completed to ensure the clinical indications paralleled the expected phenotype given the mutations found. The clinical indications for testing were provided in the form of medical records.

Once a “Notable Candidate Gene” list was compiled, co-segregation analysis was used to further investigate the feasibility of the disease-causing potential of the variant. This variant analysis was conducted using three large databases: HGMD, 1000 Genomes, and the Exome Variant Server from the Exome Sequencing Project (ESP) (University of Washington, 2014). HGMD is a comprehensive sequencing analysis tool, including a private database containing 152,000 mutation reports. 1000 Genomes is a repository for whole genome sequencing data, lightly sequenced, from a broad range of ethnicities. The goal of the project was to identify variants occurring at a 1 percent or more frequency in the human population. The project was planned and supported by an international consortium. The ESP includes an Exome Variant Server with a database in which whole exome sequencing data was compiled, the original scope

of which was to focus on finding novel genes associated with heart lung and blood disorders. Currently, ESP contains data from approximately 6,500 samples from 2 ethnic populations and is supported by the University of Washington and The Broad Institute. Both of these databases were used in the manual investigation of candidate genes during sample analysis at Ambry Genetics.

Mitochondrial Respiratory Chain Enzyme Analysis

Subsequent to finding the results from exome sequencing, mitochondrial respiratory chain enzyme analysis of muscle biopsy from quadriceps on February 27, 2013, in Sibling A was performed at Baylor College of Medicine Medical Genetics Laboratories. This test measures the efficacy of the electron transport chain (ETC) in the inner mitochondrial membrane by tracking the activity of the mitochondrial complexes I-IV in relation to the amount of citrate synthase to account for differences in cell culture environments (Kirby, 2007).

Pipelines used in the Analysis of the Data to Explain Phenotypic Differences in the Siblings

Two different pipeline analyses were conducted, one by Ambry Genetics and the other by Seven Bridges Genomics. Ambry Genetics makes the FASTQ files available to the ordering physician through an online portal. FASTQ format includes the sequence reads and the corresponding quality score of the read. The downloaded files were sent on a jump-drive to Seven Bridges Genomics and processed using a research-based bioinformatics pipeline designed to create lists of variant calls or candidate genes for the research setting. The manual investigation of the findings was performed based on a personally compiled gene list comprised of genes associated with various clinical findings within and suspected in the siblings (Table 1).

Table 1. Candidate Gene List for Focused Review by Seven Bridges Genomics Pipeline

	Chrom #	Position
Complex 1 Nuclear Encoded		
NDUFA1	X	(119871771..119876666)
NDUFA2	5	(140645363..140647785)
NDUFA10	2	(239957372..240025402)
NDUFA11	19	(5891276..5904013)
NDUFA12	12	(94971328..95003713)
NDUFA13	19	(19516210..19528204)
NDUFS1	2	(206123079..206159519)
NDUFS2	1	(161197419..161214395)
NDUFS3	11	(47579010..47584563)
NDUFS4	5	(53560635..53683341)
NDUFS5	1	(39026295..39034636)
NDUFS6	5	(1801382..1816051)
NDUFS7	19	(1383661..1395589)
NDUFS8	11	(68030617..68036647)
NDUFV1	11	(67606852..67612541)
NDUFV2	18	(9102630..9134345)
NDUFB7	19	(14566078..14572077)
NDUFB6	9	(32553526..32573184)
NDUFB8	10	(100523729..100529923)
NDUFA13	19	(19516210..19528204)
ETFDH	4	(158672101..158708713)
Complex 1 Nuclear Encoded Accessory Factors		
NDUFAF5	20	(13785026..13818421)
NDUFA3	19	(54102852..54106974)
NDUFA4	7	(10931953..10940186)
NDUFAF6	22	(42085526..42090884)
ACAD9	9	(128879490..128913114)
NUBPL	14	(31561385..31861224)
FOXRED1	11	(126269040..126278132)
Mitochondrial Complex Fe-S Cluster Biogenesis		
FXN	9	(69035563..69100178)
LYRM4	6	(5031753..5260950)
ABCB7	X	(75053172..75156340)
NFU1	2	(69396113..69438071)

Table 1. Candidate Gene List for Focused Review by Seven Bridges Genomics Pipeline

	Chrom #	Position
BOLA3	2	(74135401..74147912,
IBA57	1	(228165728..228197905)
CASK	X	(41514934..41923155)
CHMP1A	16	(89644431..89657785)
EXOSC3	9	(37779714..37785092)
OPHN1	X	(68042344..68433647)
RELN	7	(103471784..103989516)
SEPSECS	4	(25120005..25160582
TSEN54	17	(75515707..75524740)
TUBA1A	12	(49184795..49189324)
TUBA8	22	(18110687..18131731)
TUBB2B	6	(3224261..3227734)
TUBB3	16	(89922009..89936097)
VLDRL	8	(42109679..42141155)
VRK1	14	(96797347..96881614)
WDR81	17	(1716523..1738599)
CA8	8	(60188864..60281395)
ATP8A2	13	(25372071..26025851)
Ataxia, Developmental Delay Related		
EIF2B1	12	(124104953..124128323)
POLG	15	(89859534..89878092)
SACS	13	(23902965..24007841)
FRDA FXN	9	(71650175..71715094)
VLDLR	9	(2621793..2654485)
DNAJC19	3	(180701497..180707562)
SYNE1	6	(152442819..152958936)
PRICKLE1	12	(42852140..42984157)
EIF2B5	3	(183852810..184402546)
Interesting Hits from Ambry Genetics Data		
C5orf42	5	(37067870..37249428)
GNS	12	(64713442..64759446)
SHMT1	17	(18327860..18363563)
MTFMT	15	(65001512..65029639)
SYNE1	6	(152121684..152637399)

This list included known genes contributing to Complex I (CI), iron sulfur complex genes, ataxia-related genes, cerebellar hypoplasia/dysplasia-associated genes, and genes determined worth revisiting based on the data from Ambry Genetics. The gene list assembled from the Ambry Genetics data was chosen based on many factors, including appearance on Ambry Genetics' candidate known gene lists for autosomal dominant and autosomal recessive inheritance, including *de novo* mutations; an overlap of clinical phenotype and expected phenotype based on the candidate gene; and SIFT and PolyPhen scores (Table 1).

Based on results of the initial analysis with the Seven Bridges Genomics pipeline, the filters were changed. Because it was known that a variant was in an intronic region based on previous whole exome sequencing by Ambry Genetics, the filters used in the Seven Bridges Genomics pipeline were modified to search exons and to 30 base pairs into introns on either side of each exon. After the use of the Seven Bridges Genomics pipeline, Golden Helix annotated the VCF files separately for any extra information.

Analysis of Modifier Genes for Assessment of Increased Severity in Sibling A

Ambry Genetics has an in-house bioinformatics pipeline, Ambry Variant Analyzer (AVA), used by in-house lab directors and genetic counselors to assess the quality and significance of each variant call. Use of this pipeline allowed the ability to change some of the filtering parameters by which the FASTQ data files were initially analyzed. Two different inheritance models were run.

The first was autosomal dominant, including *de novo* variants. The allele frequency threshold was changed from default 0.1 percent to 1 percent to allow a larger number of variants to be considered as modifier genes, because these may lack a phenotype in and of themselves; therefore, they may be found more frequently in the general population. In these models,

Sibling A needed to have the variant, Sibling B could not have the variant, and the mother could or could not have the variant. Then each variant was assessed individually looking at the change that occurred, the PolyPhen (Adzhubei, 2010) and SIFT (Ng and Henikoff, 2006) scores, and the associated condition for that gene. The same exercise was continued for an autosomal recessive inheritance model. As before, the cutoff for allele frequency was changed, but the cutoff frequency was changed from 0.1 percent to 10 percent instead of to 1 percent (Table 2).

Table 2. Genetic Models for Modifier Genes

Name	Genetic Model	*De Novo*	Allele Cutoff %	Genotype		
				Sibling A	Sibling B	Mother
DNAD	Autosomal Dominant	Yes	1	Positive	Negative	Exclude
DNAR	Autosomal Recessive	Yes	10	Positive	Negative	Exclude

CHAPTER 2: RESULTS

Whole Exome Sequencing

Sequencing revealed compound heterozygous mutations in the *NUBPL* gene, located on chromosome 14q12, in both affected girls (Figure 5). One mutation found in the girls was c.815–27T>C, which has been documented in other patients with *NUBPL* mutations and similar MRI findings. This mutation resulted in an intronic nucleotide change from T to C 27 bases before exon 10. The mother does not have this mutation and, based on the familial co-segregation analysis, the father carries this alteration. Approximately 0.95 percent of the European American population has a C/T genotype at this location, which is considered relatively common. This is a branch point mutation causing the production of three distinct mRNA transcripts: a wild type, a truncated transcript, and an extended transcript. In the other documented cases of similar phenotypes caused by *NUBPL* mutations, the c.815-27T>C mutation is found in cis with another alteration noted to be a polymorphism (Calvo, 2010; Eis, 2014). This second alteration, p.G56R, which is found in all other documented cases of *NUBPL* disease, was not found in either sibling. The second mutation present in these girls, found in trans with the branch point mutation, is p.L104P, a novel mutation. The nucleotide change c.311T>C causes the amino acid leucine at position 104 to be replaced with proline, p.L104P. Both affected siblings carry this mutation, which they inherited from their mother (Table 3).

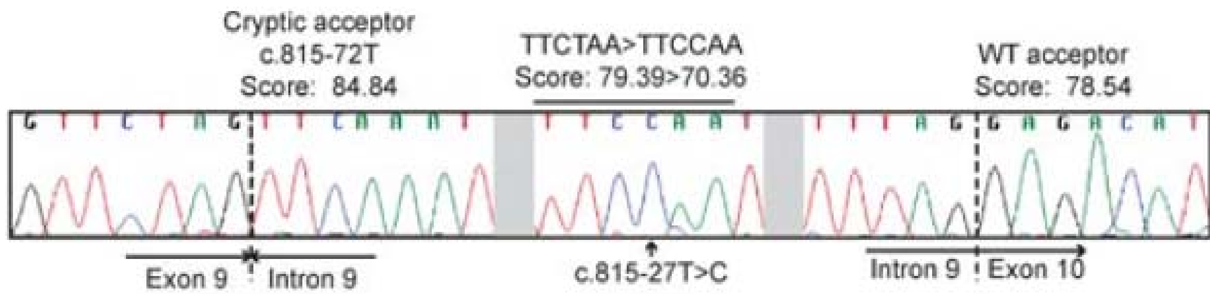

Figure 5. Sibling A and B: Chromatogram of Mutation Found in *NUBPL* Gene
(Kimonis, 2013)

Table 3. *NUBPL* Mutation Data in Sisters

Gene (RefSeq ID)	Protein	Alteration	Exon #	Genotype	Alteration Type	Sibling A	Sibling B	Mother	Father
NUBPL (NM_025152)	Nucleotide binding protein-like	c.311T>C (p.L104P)	Exon 4	Heterozygous	Missense	+/-	+/-	+/-	-/-
		c.815-27T>C	Intron 9	Heterozygous	Splice site	+/-	+/-	-/-	+/-

Mitochondrial Enzyme Testing

The interpretation of the results of the enzymatic testing reveals increased citrate synthase activity. This suggests mitochondrial proliferation and can be due to decreased mitochondrial function. In addition, no deficiencies of respiratory chain activities were detected before or after adjustment for the increased citrate synthase activity. These results would not be immediately predictive of a Complex I defect (Table 4).

Table 4. Mitochondrial Respiratory Chain Enzyme Analysis (ETC) – Skeletal Muscle

ETC Activities	ETC Complexes	Patient (% of mean)	Control ± SD*
NADH:Ferricyanide dehydrogenase	I	329 (117, 75)	280 ± 91
NADH:cylochrome c reductase	I+III		
Total		35.3 (125, 80)	28.2 ± 4.3
Rotenone sensitive		13.5 (149, 95)	9.1 ± 2.5
Succinate dehydrogenase	II	9.06 (112, 71)	8.11 ± 2.44
Succinate:cytochrome c reductase	II+III	7.25 (148, 94)	4.9 ± 1.1
Cytochrome c oxidase	IV	94.3 (323, 205)	29.2 ± 9.1
Citrate synthase		440 (157, 100)	280 ± 95

**nmol/min/mg protein*

Further Analysis to Determine Clinical Variations in Siblings

Five variants were returned by the AVA within genes that warranted further manual investigation based on associated phenotype, predicted PolyPhen scores, SIFT scores, and allele frequency (Table 5). Using an autosomal dominant inheritance pattern, including *de novo* mutations, the C5orf42 gene was altered in Sibling A and her mother, but not in Sibling B. The gene is located on chromosome 5p13.2, and the mutation is c.7829A>T, p.E2610V. It is associated with Joubert syndrome. This mutation is predicted to be benign by PolyPhen and damaging by SIFT, and it is rare in the population. In addition, GNS was mutated in both Sibling A and her mother, but not in Sibling B. GNS is located on chromosome 12q14. The mutation is c.632T>G, p.V211G. This gene is associated with mucopolysaccharidosis IIID (MPS IIID). The mutation is predicted to be benign by PolyPhen and tolerated by SIFT, and it is also considered rare in the population. Finally, SHMT1 was found to have a mutation in just Sibling A and not in Sibling B or her mother. SHMT1 is found on chromosome 17p11.2. The mutation is c.805A>G,

p.Y269C. Mutations in this gene are associated with neural tube defects. This mutation is predicted to be probably damaging by PolyPhen and tolerated by SIFT, and it is rare.

Table 5. Inheritance Model Interesting Mutations

	Gene				
	C5orf42	**GNS**	**SHMT1**	**MTFMT**	**SYNE1**
Genetic Model	Autosomal Dominant	Autosomal Dominant	Autosomal Dominant	Autosomal Recessive	Autosomal Recessive
Location	5p13.2	12q14	17p11.2	15q22.31	6q25
Mutation	c.7829A>T p.E2610V	c.632T>G p.V211G	c.805A>G p.Y269C	c.172T>A p.F58I	c.12150G>T p.K4050N
PolyPhen Prediction	Benign	Benign	Probably damaging	Probably damaging	Benign
SIFT Prediction	Damaging	Tolerated	Tolerated	Damaging	Tolerated
Sibling A's Genotype	+/-	+/-	+/-	+/+	+/+
Sibling B's Genotype	-/-	-/-	-/-	+/-	+/-
Mother's Genotype	+/-	+/-	-/-	-/-	+/-
Allele Frequency	Rare	Rare	Rare	Relatively common	Common
Condition(s) Associated with Gene Mutations	Joubert	MPS IIID	Neural tube defect	Leigh Syndrome and combined oxphos; deficiency CI deficiency	Spinocerebellar ataxia and pontocerebellar dysplasia

A second analysis was run using an autosomal recessive inheritance pattern, including *de novo* mutations. MTFMT, located on chromosome 15q22.31, has a homozygous mutation not carried by the mother but found in one gene copy of Sibling B. This mutation is c.172T>A and p.F581. It is associated with Leigh Syndrome, combined oxidative phosphorylation deficiency

and Complex I deficiency. PolyPhen predicted this mutation as probably damaging and SIFT predicted this same mutation as damaging, and the mutation is considered relatively common in the population. Lastly, a homozygous mutation in *SYNE1* on chromosome 6q25 was located. This mutation is c.12150G>T and p.K4050N. *SYNE1* mutations have been associated with spinocerebellar ataxia and pontocerebellar dysplasia (Laforce Jr., 2010). PolyPhen predicted this mutation to be benign and SIFT predicted it to be tolerated, and the variant is common in the population.

The final report generated by Ambry Genetics also included two mutations of unlikely clinical significance. The alterations were found in ALG1. c.1267C>G (p.L423V) was noted to have only some clinical overlap and was not included in further investigation. The second finding of unlikely clinical significance was found in the *GDAP1* gene.

Additionally, a bioinformatics pipeline by Seven Bridges Genomics was utilized for further analysis. Seven Bridges Genomics customizes data analysis workflows. A single variant in the *NUBPL* gene, p.L104P, was recognized; however, c.815-27T>C was not. Therefore, the Seven Bridges Genomics pipeline failed to make a diagnosis of Complex I disease due to mutations in both copies of the *NUBPL* gene. To expand the search parameters, the filters were adjusted to include 30 base pairs on either side of exons into intronic regions. Subsequently, the intron variant and the exon variant were found. The intron variant was at 14:32 319,298 T.C and was rs 118161498.

CHAPTER 3: DISCUSSION

Complex I (CI) and *NUBPL* Mutation Basics

The ETC in mitochondria is the machinery by which oxidative-phosphorlyation occurs and produces ATP. To generate ATP, the electrons generated by the citric acid cycle are pushed in and out of four large protein complexes dubbed Complexes I-IV that sit in the inner mitochondrial membrane. This creates an electrochemical gradient allowing electrons to flow down the respiratory chain and that ultimately drives ATP generation. (Vafai and Mootha, 2013 & Pagliarini, 2008)

Complex I is the first of these protein complexes, and its importance cannot be underscored enough. It is responsible for 40 percent of the electrochemical gradient, and 14 of its 45 subunits are conserved from prokaryotes (Papa and De Rasmo, 2012). The basic functions of Complex I, proton and electron movement, can be accomplished by just these ancient 14 subunits. This raises the interesting question of the true functions of the rest of the subunits. Perhaps these other subunits have an important function not completely understood (Vafai and Mootha, 2012). Complex I is coded by nuclear and mitochondrial genes and plays an important role in the assembly of other ETC complexes. Complex I genetic deficiencies are the most common of the primary neurological mitochondrial diseases (Papa and De Rasmo, 2012); however, genetic mutations have been identified as the cause for Complex I deficiencies in only approximately 50 percent of affected individuals (Wydro & Balk, 2013). Consequently, it has become evident that many other factors, including assembly factors, may contribute to Complex I deficiencies. A commonly recommended treatment for Complex I deficiencies is a mitochondrial cocktail therapy consisting of coenzyme Q10, carnitine, biotin, alpha lipoic acid, and vitamins E, C, and B complex. Both Siblings A and B are receiving this treatment, from the ages of 11 years and 1 year, respectively.

The NUBPL protein is a chaperone protein that assists in the assembly of Complex I. *NUBPL* mutations affect the assembly of the peripheral arm of Complex I, which contains eight Fe-S clusters and can be associated with Complex I deficiency. It has been shown that *NUBPL* mutations may decrease the number of Complex I subunits containing iron-sulfur clusters, which lead to Complex I deficiency (Tucker *et al.,* 2011). Such mutations are associated with the pathognomonic MRI presentation (Kevelam, 2013). Studies are underway to investigate the impact of the branch point mutation in the *NUBPL* gene using a yeast model (Wydro and Balk, 2013).

In previous cases of *NUBPL* disease phenotypes, molecular analysis found an intronic branch point mutation c.815-27T>C. *In vitro* analyses via mRNA, protein expression, and RT-PCR analyses demonstrated that the alteration reduces mRNA expression and protein levels and results in the production of three distinct transcripts. This branch site mutation produces a cryptic acceptor site that has a higher acceptor site score than the wild-type acceptor site score and is responsible for the longest of the three mRNA transcripts (Tucker *et al.,* 2011). It seems to follow that this could lead to a greater than usual proportion of non-wild-type transcript than wild-type within an individual, perhaps causing disease. Interestingly, approximately 0.95 percent of the European American population has a C/T genotype at this branch site location, which is a rather high allelic frequency for a disease-causing mutation. The c.311T>C has only been seen once in 11,780 alleles tested according to the ESP. It is in a highly conserved region and is predicted to be deleterious by SIFT in silico models and probably damaging by PolyPhen. Based on this data, Ambry Genetics has labeled this variant as likely to be deleterious.

Currently, seven other cases of Complex I deficiency due to *NUBPL* mutations have been documented worldwide. With Siblings A and B, the case count rises to nine (Table 6). All previous cases reported by Calvo have in cis both the branch site mutation c.815-27T>C and

another polymorphism c.166G>A. Among the previously reported cases, Patient 1 is homozygous for these alterations. Patients 2 through 7 also have varying other mutations found in trans to the branch site mutation. My reported Siblings A and B are particularly interesting because they do not carry the polymorphism in cis with the branch site mutation. If indeed this common alteration is not entirely benign in nature, it may be another factor to explain their relatively mild phenotype.

Table 6. Documented Cases of Complex I Deficiency Due to *NUBPL* Mutations

Patient	Country of Origin	c.DNA	Protein	Exon	State	Inheritance
1	Argentina	c.166G>A	p.Gly56Arg	2	Homozygous	Unknown
		c.815-27T>C		Intron 9	Homozygous	Unknown
2	Germany	c.166G>A	p.Gly56Arg	2	Heterozygous	Paternal
		c.667_668insCCTTGTGCTG	p.Glu223Alafs*4	8	Heterozygous	Maternal
		c.815-27T>C		Intron 9	Heterozygous	Paternal
3 & 4 (sibs)	Canada	c.166G>A	p.Gly56Arg	2	Heterozygous	Paternal
		c.313G>T	p.Asp105Tyr	4	Heterozygous	Maternal
		c.815-27T>C		Intron 9	Heterozygous	Paternal
5	United States	c.166G>A	p.Gly56Arg	2	Heterozygous	Paternal
		593+1G>A	p.?	Intron 8	Heterozygous	Unknown
		815-27T>C		Intron 9	Heterozygous	Paternal
6	Netherlands	c.166G>A	p.Gly56Arg	2	Heterozygous	Maternal
		c.579A>C	p.Leu193Phe	7	Heterozygous	Paternal
		c.815-27T>C		Intron 9	Heterozygous	Maternal
7	Australia	c.166G>A	p.Gly56Arg	2	Heterozygous	Paternal
		240-kb deletion (exons 1-4); 137-kb duplication (exon 7)		1-4 and 7	Heterozygous	Maternal
		c.815-27T>C		Intron 9	Heterozygous	Paternal
8[1]	France	c.205_206dlGT	Premature stop	Unknown	Heterozygous	Unknown
		c.815-27T>C		Intron 9	Heterozygous	Unknown
Siblings A & B	United States	c.311T>C	p.L104P	4	Heterozygous	Maternal
		c.815-27T>C		Intron 9	Heterozygous	Paternal

(Kevalam, 2013; Eis et al., 2014)

[1] It is not noted in the literature whether Patient 8 has the c.166G>A polymorphism.

Tools for *NUBPL* Mutation Identification

NUBPL mutations are assessed through a combination of four methods: phenotypic assessments, brain MRIs, enzyme assays for Complex I activity, and whole exome sequencing. Sibling A did not show a decrease in Complex I activity by enzyme assay of muscle tissue, and both showed abnormal brain MRIs typical of this disorder. Whole exome sequencing revealed that Siblings A and B have the c.815-27T>C mutation, neither has the c.166G>A mutation.

Six patients from an MRI database of more than 3,000 cases of unclassified leukoencephalopathy were recognized by MRI findings according to three criteria: (1) extensive cerebellar cortex signal abnormalities, (2) signal abnormalities in the corpus callosum, and (3) absence of signal abnormalities in the basal ganglia, thalami, and cerebral cortex (Kevelam, 2013). All patients from this study had both a c.815-27T>C in cis with a c.166G>A mutation (Kevelam, 2013).

Clinical presentations included gain or loss of motor skills and signs of cerebellar dysfunction at the end of the first or during the second year of life. All patients had motor problems due to ataxia, and plasma and cerebrospinal fluid (CSF) lactate was elevated in most patients.

NUBPL mutations can decrease the efficiency of Complex I. Respiratory chain enzyme assays in muscle and fibroblasts revealed a Complex I deficiency in all tested patients ranking 27 to 83 percent of the lowest reference value (Kevelam, 2013). Both of these mutations in a separate study were ascertained by Complex I deficiency patterns (Calvo, 2010). Although no Complex I deficiencies were detected in Siblings A and B, enzymatic testing revealed an increase in citrate synthase activity, which may suggest mitochondrial proliferation. When mitochondria are not working sufficiently to provide the energy needed, they may increase in

number to compensate for reduced activity. Given that Siblings A and B appear to have more benign clinical symptoms, perhaps the impact of their genetic abnormalities are minor enough to be offset by an increased quantity of mitochondria, thus leading to increased citrate synthase activity and undetectable Complex I deficiencies.

After exome sequencing was complete, brain MRI images were shared with Dr. Marjo S. Van der Knaap, a neuroradiologist familiar with *NUBPL*-related MRI changes. She indicated that the MRI changes in the siblings were similar to the previous reported patients with *NUBPL* mutations. Given the "selective involvement of the cerebellar cortex," according to this expert, and the normal ETC enzyme assay, it appears that these siblings are on the more "benign end of the spectrum" and would have been difficult if not impossible to diagnose via MRI and Complex I deficiency, the standard tools of diagnosis for these phenotypes (personal communications).

This case suggests that whole exome sequencing for rare disease phenotypes can provide distinct advantages to existing methodologies for shortening the diagnostic odyssey that some families affected by rare diseases face. Had the exome sequencing not been completed, it is highly unlikely that the *NUBPL* mutation would have been discovered as quickly given that their clinical presentation was not as severe as previous patients who were identified as having such mutations. In cases where distinct findings are difficult to ascertain with more standard methods, whole exome sequencing has the potential to be a strong complementary diagnostic tool.

Bioinformatic Pipelines

The goal of using filters in bioinformatic pipelines is to reduce the massive number of potentially significant variants to a manageable number of variants to be analyzed. All bioinformatic pipelines select their own unique set, order, and threshold of filters depending on

the goals of the particular researcher or laboratory. Seven Bridges Genomics provides customizable pipelines based on customer requests, and a pipeline for researchers in nonclinical settings was requested for this study. Ambry Genetics' pipeline focuses more on clinical applications; the lab is Clinical Laboratory Improvement Amendments (CLIA)-certified and intends for diagnostic results to be returned to patients. These pipelines were selected to compare how a research pipeline could potentially find modifier genes that would not have shown up in an initial clinically significant variant list.

Surprisingly, Seven Bridges Genomics' research pipeline did not find the c.815-27T>C mutation (i.e., the branch point intronic mutation). This may have been missed for many reasons. Seven Bridges Genomics may have used an intron filter because they are not coding regions, causing intronic variants to be dropped from candidate lists. Additionally, filters for read depth and/or variants for rare conditions may have also contributed to the missed mutation. To determine what exactly caused this difference in findings, the filters used for each pipeline would need to be analyzed and compared. Although the branch site mutation has an allelic frequency of 0.95 percent, the AVA pipeline did not filter it out in spite of a 0.1 percent filter, because the pipeline prioritizes retaining all variants annotated by the Human Gene Mutation Database (Stenson *et al.,* 2009).

This calls into question the finality of "negative" results and should encourage clinicians to remember that there is subjectivity within testing. Given the results of whole exome testing in this case, prior suggestions of using multiple pipelines to increase the chance of finding significant variants are supported (O'Rawe *et al.,* 2013).

Genetic Modifiers Can Change Phenotypes

Unlike previously reported cases of patients with *NUBPL* mutations, Siblings A and B do not have the c.166G>A mutation. Although this mutation is not expected to affect NUBPL import or processing based on *in vitro* import assays (Tucker, 2011), all other patients with this polymorphism had phenotypes with more severe clinical presentations than Siblings A and B (as mentioned previously, MRI testing and enzyme assays for Siblings A and B were considered on the "benign" end of the spectrum).

Given this distinction between the affected siblings and other patients with *NUBPL* mutations, it is possible that the c.166G>A mutation is not truly a benign polymorphism as believed. Instead, it may contribute to the severity of clinical presentations by increasing the difference between the acceptor site scores of the cryptic-type and wild-type, leading to greater transcription of the longer mutant mRNA. Another possibility is that this missense mutation leads to some instability of the NUBPL protein, making it unable to compensate for Complex I deficiency purely by mitochondrial proliferation. Further studies are needed to investigate the role that the c.166G>A mutation plays in the phenotype associated with *NUBPL* mutations and whether either of these explanations are valid. One should also note that these differences may also be explained by the variety of second mutations found on the opposite allele, which varied amongst families.

Assuming these sisters do in fact have the same condition, how can the differences in their clinical presentations be explained? It has been long observed that many conditions described to have simple Mendelian inheritance have variable expressivity or incomplete penetrance, and some have no strong phenotype-genotype correlation. To discuss this surprising situation, genetic modifiers have been postulated. In the original sense of the concept, a genetic

modifier is a locus that causes a change in the phenotypic output of a gene it interacts with – the target gene. Much work has been done in recent years to understand how genes impact each other. For instance, patients with cystic fibrosis present with a broad range of organ involvement and clinical severity, much of which is impacted by modifier genes (Cutting, 2010, from Hamilton and Yu, 2012). Results seem to demonstrate that genes do not necessarily act in a linear fashion but in more complicated networks. These conclusions make hypothesizing the different ways genes could affect one another broad and numerous. In fact, the endless routes of interaction allow for at least partial explanation of human variation. The ways that genes have been shown to interact include additive, epistatic, and protective effects. In the case of an additive effect of modifier genes, the risks to phenotype are inclusive of the risks of each of the alleles individually. Epistatic effects of a modifier gene effectively serve to hide the phenotype of the mutated target gene. Finally, the presence of certain alleles may in fact seemingly reduce the severity or penetrance of a phenotype associated with the condition revealed by a mutation in a target gene (Slavotinek and Bieseker, 2003).

The five genes flagged for further manual investigation (Table 5) were considered for the original diagnosis with *NUBPL.* Two of the variants found with the dominant inheritance pattern were C5orf42 and GNS. The mother carried both of these changes but remains asymptomatic. Neither of these nucleotides is well conserved across species, and each has at least one *in silico* prediction to be benign; however, the variants are very rare. In the case of the *SHMT1* gene, the mother does not carry the variant; therefore, the father must carry the change, or it is *de novo.* It is a rare variant, but it is predicted to be tolerated by SIFT and is only moderately conserved. Interestingly, C5orf42 is associated with Joubert syndrome, which was considered in the differential quite early in Sibling A's testing. GNS is associated with MPS IIID. Testing

performed on Sibling A included storage disease, which was negative. SHMT1 is associated with neural tube defects, which does not appear to be associated with the symptoms of Sibling A. I considered two variants found during sorting with an autosomal recessive inheritance pattern. MTFMT provided an exciting opportunity for consideration. This gene is associated with Leigh syndrome and Complex I deficiency; however, due to the read depth, only six reads were detected, including one of which was wild type; I discarded this possibility as artifact. Finally, the two mutations in the *SYNE1* gene I considered worthy of more thought. This gene is associated with spinocerebellar ataxia. The fact that the variant occurs in the human population at 8.5 percent and in the Chinese population at as high as 13.5 percent, according to 1000 Genomes, it is unlikely that this variant is the solitary cause of the girls' condition. This variant does not exist in ESP. Such conflicting approximations of allele frequency could be due to psuedogenes. More research is needed to clarify this conflict.

After considering the five genes manually chosen for closer inspection, none seemed to fit more exactly the primary diagnosis associated with two mutations found in *NUBPL*. Nevertheless, it is possible that these genes could be acting in conjunction with the *NUBPL* gene and may explain some of the differences seen between the two siblings. As a modifier gene, of particular interest is the *SYNE1* gene. Mutations in this gene lead to pontocerebellar dysplasia. The more-affected Sibling A acquired two mutations in both the *SYNE1* and *NUBPL* genes. Sibling B only carries one change in her *SYNE1* gene. In addition, it is active in mitochondria. More research into these genes and their variants is required to tease out any interaction that may occur between these genes.

Implications

Because rare diseases by definition affect extremely small numbers of people, it can be difficult to justify spending so much time and money to reach a diagnosis. Whole exome sequencing is still not regularly accepted as a gold standard by third-party payors, including private and public insurance providers. Consequently, it can be a long process to get such expensive, experimental testing authorized for patients. Further studies explaining the benefits of whole exome sequencing, particularly in the case of rare diseases, will benefit patients and the medical community by encouraging insurance companies to accept this testing as a necessary method of diagnosis in certain instances. Additionally, as this case shows, going to great lengths can result in novel methods of studying genes that open doors to many other lines of research in addition to a diagnosis that can empower the affected family.

As specialists have noted, complex diseases are typically caused by a combination of mutations associated with rare diseases. Patients with atypical or more widely varying symptoms are often shown to be affected by known mutations via whole exome sequencing. This presents an opportunity to expand the phenotypic spectrum for known disease-causing genes, thus benefitting the entire medical community. Moreover, studying networks of genes and their potentials as modifiers of target genes can help characterize more distinctly the ideas of variable expressivity, incomplete penetrance, and lack of phenotype-genotype correlation.

This case is emblematic of how rare disease research occurs in that it required the collaboration of many great minds across the world in various disciplines. The MRI specialist who reviewed the affected girls' MRI images hailed from the Netherlands and has amassed MRI findings from the few *NUBPL* patients found globally. In the United Kingdom, functional studies using yeast are being performed and watched closely by others internationally who study the

NUBPL gene. International researchers have found that Complex I can be a risk factor for developing hereditary forms of Parkinson's Disease (Papa and De Rasmo, 2013). American researchers recently presented notable results that heterozygous *NUBPL* mutations, including c.815-27T>C, have been implicated in increasing the risk of developing Parkinson's disease. In their study of Parkinson's patients (*n=478*), seven patients were confirmed to have a single *NUBPL* mutation by Sanger sequencing. Interestingly, three of the seven patients, or 43 percent, were found to have the branch point c.815-27T>C mutation (Eis *et al.,* 2014). These patients, like Siblings A and B, do not have the polymorphism. Researchers in Australia, led by David Thorburn, are also currently studying *NUBPL* mutations. This global, interdisciplinary effort continues to further our understanding of genetics, resulting in tangible benefits to the patients we all ultimately hope to serve.

REFERENCES

Adzhubei, I.A, *et al*. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* **7**(4):248-249.

Calvo, S. E. *et al*. 2010. High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nature Genetics* **42**, 851-858.

Centrum Medische Geneticz UZ Brussel. 2012. *Gene with introns and exons* [Diagram]. Retrieved from http://www.brusselsgenetics.be/genes.

Eis, P.S. *et al.* 2014. Genetic Variation in Mitochondrial Complex I Genes are Associated with Parkinson's Disease. Poster presented at the Parkinson's Disease: Genetics, Mechanisms and Therapeutics. March.

Gilissen, C. *et al.* 2011. Unlocking Mendelian disease using exome sequencing. *Genome Biology* **12(9)**, 228.

Hamilton, B.A. and B.D. Yu. 2012. Modifier genes and the plasticity of genetic networks in mice. *PloS Genetics* **8(4)**, e1002644.

Kevelam, S.H. *et al*. 2013. *NUBPL* mutations in patients with complex I deficiency and a distinct MRI pattern. *Neurology*, **80**, 1577-1583.

Kimonis, V. *et al*. 2013. Clinical Exome Sequencing Leads to the Diagnosis of Mitochondrial Complex I Deficiency in a family with global developmental delays, ataxia, and cerebellar and pons hypoplasia. Poster presented at ASHG annual meeting.

Kirby, D.M. *et al*. 2007. Biochemical assays of respiratory chain complex activity. *Methods in Cell Biology,* 80 pgs 93-119.

Laforce Jr., R. *et al*. 2010. Cognitive Impairment in ARCA-1, a Newly Discovered Pure Cerebellar Ataxia Syndrome. *Cerebellum*. **9**, 443-453.

Ng, P.C. and S. Henikoff. 2006. Predicting the effects of amino acid substitutions on protein function. *Hum Genet.* **7**:61-80.

O'Rawe, J. *et al*. 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine*. **5(28)**.

Papa, S. and D. De Rasmo. 2013. Complex I deficiencies in neurological disorders. *Trends in Molecular Medicine* **19(1)**, 61-69.

Pagliarini, D. J. *et al*. 2008. A mitochondrial protein compendium elucidates complex I disease biology. *Cell,* **134(1)**:112-123.

Slavotinek, A. and L.G. Bieseker. 2003. Genetic modifiers in human development and malformation syndromes, including chaperone proteins. *Human Molecular Genetics* **12(1)**, R45-R50.

Smeets, C.J.L.M. and D.S. Verbeek. 2014. Cerebellar ataxia and functional genomics: Identifying the routes to cerebellar neurodegeneration. *Biochimica et Biophysica Acta (BBA) – Molecular Basis of Disease* In Press. April 13.

Stenson, P.D., *et al*. 2009. The Human Gene Mutation Database: 2008 Update. *Genome Med* **1**(1)13.

Teisch, E.V. *et al*. 2012. Massive and exclusive pontocerebellar damage in mitochondrial disease and *NUBPL* mutations. *Neurology*. **79**, 391.

Tucker, E.J. *et al*. 2011. Next-generation sequencing in molecular diagnosis: *NUBPL* mutations highlight the challenges of variant detection and interpretation. *Human Mutation* **33(2)**, 411-418.

University of Washington. 2014. NHLBI Exome Sequencing project (ESP) Exome Variant Server. Retrieved from http://evs.gs.washington.edu/EVS/. June 1.

Vafai, S.B. and V.K. Mootha. 2013. A common pathway for a rare disease? *Science* **342**, 1453-1454.

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

Wellcome Trust. 2003. *Gene expression* [Diagram]. Retrieved from http://genome.wellcome.ac.uk/doc_WTD020757.html.

Wydro, M. and J. Balk. 2013. Insights into the pathogenic character of a common *NUBPL* branch-site mutation associated with mitochondrial disease and complex I deficiency using a yeast model. *Disease Models & Mechanisms.* **6**, 1279-1284.

1000 Genomes. 2014. 1000 Genomes A Deep Catalog of Human Genetic Variation. Retrieved from http://www.1000genomes.org/home. June 1.

APPENDIX A:

SIBLING A TESTING BY YEAR AND THE RESULTS OF THIS TESTING

Date	Age	Test	Results
10/27/98	0 yr, 0 mo	Cord Blood Type	O
10/27/98	0 yr, 0 mo	Cord Rh	Negative
10/28/98	0 yr, 0 mo	Phenylketonuria	Negative
10/28/98	0 yr, 0 mo	Galactosemia	Negative
10/28/98	0 yr, 0 mo	Congenital Primary. Hypoplasia	Negative
10/28/98	0 yr, 0 mo	Hemoglobinopathy	Normal
5/99	0 yr, 6 mo	EEG Dr. Fowler	normal
8/16/99	0 yr, 9 mo	Video EEG	normal
8/16/99	0 yr, 9 mo	24-hr electroencephalogram	No evidence of seizures
8/16/99	0 yr, 9 mo	MRI Brain	Abnormal – diffuse abnormal signal from cortex of cerebellar hemispheres, white matter normal, brainstem normal.
8/16/99	0 yr, 9 mo	CSF sample obtained	No abnormal protein or abnormal cells (r/o postinfection)
8/17/99	0 yr, 9 mo	Plasma AAs and UOAs	Normal
8/17/99	0 yr, 9 mo	Lysosomal storage screen. GM1, gangliosidosis, beta mannosidosis, Tay-Sachs, metachromatic leukodystrophy, Krabbe disease	Normal
In hosp before 9/17/99	<0 yr, 10 mo	MRI	Evidence of altered density of the cerebellum
9/17/99	0 yr, 10 mo	Blood for mitochondrial point mutations and Southern blot for deletions to CHLA	No evidence of deletions or specific mutations (i.e., those associated with ophthalmoplegia)
9/29/99	0 yr, 11 mo	Methylmalonic Acid. Homovanillic Acid	Normal
10/19/99	0 yr, 11 mo	Abdominal U/S. r/o Neurofibromatosis	Normal
10/20/99	0 yr, 11 mo	MRI Brain w/wo contrast	Abnormal – diffuse abnormal signal intensity within cerebellar hemispheres bilaterally shows irregular enhancement. Does not have appearance of demylenating disease. Does not appear to be inflammatory process.

Date	Age	Test	Results
11/23/99	1 yr, 0 mo	Brain biopsy path report	Arachnoid a little opaque and very pink – Increased vascularity on the surface of cerebellum, cerebellum tissue firm but not rubbery, population of large, often swollen Purkinje's cells, scattered tissue edema, large number of vessels some with thick-walled arteries perhaps suggestive of vascular malformation, some degenerative features, no evidence of – lysosomal storage disease, viral inclusions, parasitic organisms, mt morphologic abnormalities, not typical of Lhetmitte-Duclos-Cowden disease, hers is not hamartoma.
11/23/99	1 yr, 0 mo	Brain tissue sent	r/o storage disease, cerebellar dysplasia or dysgenesis
12/13/99	1 yr, 1 mo	Brain biopsy – Dr. Lucy Rorke looked at sample	Foliar architecture is clearly pathological, Purkinje cells large and irregularly distributed, molecular layer contains a jumbled population of Brigmann astrocytes, almost no cells in the internal granular layer but those seen undergoing karyorrhexis, excessive vessels "spectacular" "bizarre""remarkable indeed" consider Joubert syndrome
10/22/01	2 yr, 11 mo	MRI Brain w/wo contrast	Abnormal – diffuse abnormal signal of the cerebellar cortex both sides. Fourth ventricle is markedly large, normal white matter appearance
4/9/04	5 yr, 5 mo	Ophthalmology evaluation	Unremarkable
8/10/05	6 yr, 9 mo	MRI Brain w/wo contrast	Severe global prominence of the sulci in the r and l cerebellum, T2 prolongation in the r and l cerebellar hemispheres, middle, superior and inferior cerebellar peduncles are small, pons slightly small, no diffusion abnormalities, minimal T2 prolongation in the subcortical and periventricular white matter of the posterior left temporal lobe…impression – encephalomalacia involving the r and l cerebellar hemisphere and cerebellar peduncles, minimal nonspecific subcortical and periventricular white matter changes in the posterior left temporal lobe.
4/21/10	11 yr, 5 mo	MRI C-Spine	Normal with mild anterior wedging of c3,c4,c5
4/21/10	11 yr, 5 mo	MRI Head	Cerebellum abnormal showing diffuse hypoplasia/atrophy, vernis is apparently absent, 4th ventricle enlarged and unusual in shape, communicating with large cisterna magna, cerebellar peduncles are slightly elongated with very mild increase in depth of interpeduncular fossa impression – findings suggestive of congenital hypoplasia of the cerebellum rather than developmental neurodegenerative disease
4/21/10	11 yr, 5 mo	MRI T-spine	normal
4/21/10	11 yr, 5 mo	MRI L-spine	Mild disc bulge at L4-L5 and L5-S1

Date	Age	Test	Results
8/24/10	11 yr, 9 mo	Oligo Microarray	Normal
8/24/10	11 yr, 9 mo	MECP2 sequencing for Rett syndrome	No mutation detected
8/24/10	11 yr, 9 mo	7-Dehydrocholesterol (SLOS)	Normal
8/24/10	11 yr, 9 mo	Peroxisomal disorders (plasma Total Lipid Very Long Chain and Branched Fatty Acids phytanic acid, plasmalogens, pipecolic acid	Normal
8/26/10	11 yr, 9 mo	Carbohydrate deficient transferring for CDG (congenital disorder of glycosylation)	Normal
10/19/10	11 yr, 11 mo	Surgery	Exostosis right ring finger, pes planus R
11/25/10	12 yr, 0 mo	VLDLR sequencing	No mutation detected
6/15/11	12 yr, 7 mo	Athena Complete Ataxia Evaluation Panel	Genes tested – (sequencing) MSS, AVED, KCNC3, FXN, MIRAS-POLG, APTX, PRKCG, SETX, AFG3L2, SPTBN2, (repeat expansion) ATXN3, FXN, PPP2R2B, ATXN1, CACNA1A, ATNX8OS, ATXN10, TBP, ATN1, ATXN2, ATXN7. All negative
11/27/12	14 yr, 1 mo	Exome testing	*See Results section*
2/27/13	14 yr, 4 mo	Electron transport chain analysis skeletal muscle	Normal

APPENDIX B:

SIBLING B TESTING BY YEAR AND THE RESULTS OF THIS TESTING

Date	Age	Test	Results
05/05/10	1 yr	MRI Brain	Cerebellum abnormal in appearance, diffuse increase in T2 signal, mild diffuse hypoplasia, 4th ventricle mildly prominent in size (not as severe as sister at 11 yr, 5 mo)
05/05/10	1 yr	MRI L-Spine	urinary bladder severely distended
05/05/10	1 yr	MRI T-Spine	Apparently prominent (in size) distal esophagus w/ fluid in lumen – could be related to anesthesia
05/05/10	1 yr	MRI C-Spine	normal
09/09/13	4 yr, 4 mo	05/05/10 Images to Dr. Van der Knapp	Cerebellar cortex, diffusion-weighted images indicate diffusion restriction in the cerebellar cortical structures, almost exclusively affects the cerebellar cortex