2014-12-17

Identifying Novel Genes and Genetic/Phenotypic Spectra for Inherited Neurodegenerative Disorders Using Next-Generation Sequencing

Michael A. Gonzalez
University of Miami, MikeGonzalez2015@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/1351

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
IDENTIFYING NOVEL GENES AND GENETIC/PHENOTYPIC SPECTRA FOR INHERITED NEURODEGENERATIVE DISORDERS USING NEXT-GENERATION SEQUENCING

By

Michael Anthony Gonzalez

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the Doctor of Philosophy

Coral Gables, Florida December 2014
UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

IDENTIFYING NOVEL GENES AND GENETIC/PHENOTYPIC SPECTRA FOR INHERITED NEURODEGENERATIVE DISORDERS USING NEXT-GENERATION SEQUENCING

Michael Anthony Gonzalez

Approved:

Stephan Zuchner, M.D., Ph.D.
Professor of Human Genetics

Susan Blanton, Ph.D.
Professor of Human Genetics

Mustafa Tekin, M.D.
Professor of Human Genetics

John Gilbert, Ph.D.
Professor of Human Genetics

Vance Lemmon, Ph.D.
Professor of Neurological Surgery

M. Brian Blake, Ph.D.
Dean of the Graduate School
Length-dependent axonal degeneration, sometimes referred to as the "dying-back" of a nerve, results in axonal degeneration at the most distal extent of the axon. Degeneration of extended axons in the peripheral (PNS) and central nervous systems (CNS) is the pathological basis for a number of neurological disorders, including the axonal peripheral neuropathies known as Charcot-Marie-Tooth disease type 2 (CMT2) and the axonal degeneration of the corticospinal tract in the CNS known as hereditary spastic paraplegias (HSP). CMT2 and HSP are inherited neurodegenerative disorders characterized by progressive axonal degeneration, which can cause severe disabilities and have no clinically available treatment options. While length-dependent axonal degeneration in CMT2 occurs in the periphery and in HSP in the central nervous system, it is becoming more evident that there is significant biological and genetic overlap between these disorders. In addition, the known genes only explain ~60% of HSP cases and ~30% of CMT2 cases, which suggests many more disease genes are still left to be identified. New technologies, such as exome and genome sequencing, allow us to rapidly increase our knowledge on genes involved in these diseases. Therefore, the overall goal of this project was to use exome sequencing to build
upon and extend the existing knowledge of genetic factors and biological pathways involved in axonopathies. I analyzed >1,000 whole exome datasets from patients with CMT2 and HSP. This analysis has lead to the identification of the novel disease genes *MARS* and *DDHD2*, which cause CMT2 and HSP, respectively. However, while producing this large cohort of exome data, it became evident that novel strategies for analyzing genomic-scale data would be needed. To this end, I developed GEnomes Management Application (GEM.app) to address the computational challenges brought forth by genomic ‘big data’. This platform has lead to numerous gene identifications (*BICD2*, *GBA2*, *DDHD1*, *DDHD2*, *FBXO38*, *REEP2*, etc). Due to the scalability of the GEM.app platform, I was able to analyze large collections of exome datasets. These analyses lead to the findings that redefined the phenotypic spectrum of two disease genes *PNPLA6* and *VCP*. Given that much of our knowledge of the pathophysiology of these diseases has initially been identified via genetic studies, the value of the present work cannot be underestimated. Further characterization of molecular pathways involved in axonal degeneration will lead to better understanding of biological mechanisms and eventually lead to new therapeutic options for disorders characterized by axon degeneration.
DEDICATION

To my family, girlfriend, and life long friends.
ACKNOWLEDGEMENTS

The Zuchner lab was instrumental in my thesis projects: Fiorella Speziani, Gladys Montenegro, Dr. Adriana Rebelo, Alleeene Strickland, Justin Price, Eric Powell, Alex Abrams, Feifei Tao, Rafael Acosta and Cima Saghira.

Dr. Stephan Zuchner, Dr. Susan Blanton, Dr. Mustafa Tekin, Dr. John Gilbert, Dr. Vance Lemmon for their mentorship throughout my graduate training. Dr. Stephan Zuchner and Dr. Vance Lemmon were vital in building my scientific network.

Dr. Michael Gaines, for his mentorship through the Howard Hughes Medical Institute Bridges to the Future program at the University of Miami. Thank you all the guidance during my undergraduate education and for helping me discover my interests in biomedical research.

The faculty and staff at the Hussman Institute for Human Genomics for providing me with a great education and training on the latest technologies in human genomics research. Dr. Bill Scott and Dori McLean for all the reminders to meet important deadlines leading up to my dissertation defense.

Sandhya Ramsook for endless hours of editing this manuscript and all the encouragement throughout my graduate work. Thank you for always being
supportive even when I would work throughout the night. Your support, encouragement, quiet patience and unwavering love were undeniably the reason I survived the experience of graduate school.

My friends for always reminding me that I needed some time to let my mind wander from thinking about science.

Mom, Melissa, David and Grandma for your love and support during these crazy times. Your understanding that my time away from you was spent trying to accomplish my goals helped me stay focused and finish my graduate training.
TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................. x
LIST OF TABLES ...................................................................................................................... xii
LIST OF ABBREVIATIONS ..................................................................................................... xiii

CHAPTER
1. Introduction to Inherited Neuropathies: Clinical Features, Genetics, and Insights from Biomedical Research .......................................................... 1
   Overview of the Central Nervous System ................................................................. 1
   Inherited Upper Motor Neuron Disorders: Hereditary Spastic Paraplegia ........ 1
   Molecular Mechanisms in Hereditary Spastic Paraplegia ................................. 4
   Overview of the Peripheral Nervous System ....................................................... 7
   Inherited Peripheral Neuropathies: Charcot-Marie-Tooth Disease .............. 8
   Molecular Mechanisms in Axonal Charcot-Marie-Tooth Disease ............ 10
   Genetic and Biological Comparison in CMT and HSP .................................. 14
   Next-Generation Sequencing to Study Inherited Neuropathies ................ 19
   Current Challenges with Next-Generation Sequencing .............................. 21
   Summary .................................................................................................................. 21

2. Exome Sequencing Identifies a Significant Variant in Methionyl-tRNA Synthetase (MARS) in a Family with Late Onset CMT2 ....................... 23
   Perspectives ............................................................................................................. 23
   Materials and Methods ......................................................................................... 24
Genomic Studies .............................................................................73

Functional Evaluation .....................................................................75

Discussion .....................................................................................78

6. **PNPLA6 Mutations Cause Boucher-Neuhaüser and Gordon Holmes Syndromes as Part of a Broad Neurodegenerative Spectrum ..........82**

Perspectives ..................................................................................82

Whole Exome Sequencing of Index Patients with Boucher-Neuhaüser and Gordon Holmes Syndromes ...................................................83

Sanger Sequencing of the Candidate Gene in Additional Boucher-Neuhaüser Syndrome Families .........................................................85

**PNPLA6** Analysis in More than 500 Exomes of Index Patients with Ataxia, Hereditary Spastic Paraplegia and Charcot-Marie-Tooth Disease ..........85

Structural Models of Wild-Type and Mutant PNPLA6 Domains ........86

Clinical Assessment ........................................................................86

**PNPLA6** Causes Boucher-Neuhaüser and Gordon Holmes Syndrome .87

**PNPLA6** Causes a Wide Spectrum of Neurodegenerative Disease ......90

Structural in silico Protein Modeling ................................................93

Clinical Findings ..........................................................................98

Discussion .....................................................................................102

7. **Summary, Future Directions, and Discussion .........................106**

Summary of Dissertation Findings ................................................106

Expanding the Number of Axonopathy Genes: MARS and DDHD2 ....107

Managing the Challenges of Genome Analysis .................................108
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Central Nervous System</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Motor and Sensory Neurons</td>
<td>9</td>
</tr>
<tr>
<td>1.3</td>
<td>Peripheral Neuron</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Similarities and differences in the molecular pathology of CMT and HSP</td>
<td>17</td>
</tr>
<tr>
<td>2.1</td>
<td>Pedigree of family with MARS variant</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>Multiple sequence alignment of MARS</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Structural protein model of MARS</td>
<td>30</td>
</tr>
<tr>
<td>2.4</td>
<td>Study of MARS Variants in Yeast</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>Pedigrees and sequence traces of the DDHD2 families</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic of the DDHD2 gene</td>
<td>41</td>
</tr>
<tr>
<td>3.3</td>
<td>Cranial MRI/CT scans of DDHD2 patients</td>
<td>42</td>
</tr>
<tr>
<td>4.1</td>
<td>GEM.app graphical user interface (GUI)</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Example of output from a GEM.app analysis query</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>GEM.app annotation manager</td>
<td>58</td>
</tr>
<tr>
<td>4.4</td>
<td>Performance of GEM.app</td>
<td>60</td>
</tr>
<tr>
<td>4.5</td>
<td>Current usage of GEM.app</td>
<td>62</td>
</tr>
<tr>
<td>5.1</td>
<td>Pedigree of family with VCP variant</td>
<td>67</td>
</tr>
<tr>
<td>5.2</td>
<td>VCP gene-network</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>VCP gene schematic and protein model</td>
<td>77</td>
</tr>
<tr>
<td>5.4</td>
<td>Functional evaluation of the VCP mutation</td>
<td>79</td>
</tr>
</tbody>
</table>
6.1 Pedigrees and Segregation of the Mutations Identified in PNPLA6 ...91
6.2 Electropherograms of Mutations Identified with Sanger Sequencing .92
6.3 Evolutionary conservation of PNPLA6 mutations ……………………94
6.4 Schematic of PNPLA6 gene ………………………………………96
6.5 Structural model of EST domain (grey) of human PNPLA6 ………97
6.6 Structural models of wild type and mutant CNB1 and CNB2 domains in complex with cAMP ………………………………………………… 99
6.7 The continuous spectrum of PNPLA6-associated disease ………100
6.8 The phenotypic spectrum caused by PNPLA6 mutations ………104
7.1 Graph illustrating the nature of the reductions in DNA sequencing costs ………………………………………………………………………109
LIST OF TABLES

Table

2.1 Top variants remaining after applying filter criteria on exome sequencing data .........................................................27

2.2 Corresponding protein residues in human and yeast .........................32

3.1. Five segregating variants in family THI26003 after filtering of sequencing data ........................................................................................................38

3.2 Clinical features of genetically confirmed DDHD2 patients .............44

5.1 Clinical features of family with E187K VCP mutation .......................70

5.2 Motor Nerve Conduction Velocities ..............................................71

5.3 Sensory Nerve Conduction Velocities ............................................72

5.4 in silico scores ......................................................................................74

6.1 Clinical features of PNPLA6 patients ..............................................88

6.2 PNPLA6 mutations identified in this study ......................................89
LIST OF ABBREVIATIONS

AARS - alanyl-tRNA synthetase
ALS - amyotrophic lateral sclerosis
ARCAAs - Autosomal recessive cerebellar ataxias
ARS - aminoacyl-tRNA synthetases
BNS - Boucher-Neuhäuser syndrome
CMT - Charcot–Marie–Tooth disease
CMT1 - demyelinating CMT
CMT2 - axonal degenerative CMT
CMTNSv2 - CMT Neuropathy Score
cNMPs - cyclic nucleotide-monophosphate binding
ERGIC - cis-Golgi and the ER-Golgi intermediate compartment
FA2H - fatty acid-2 hydroxylase
GARS - glycyl-tRNA synthetase
GEDI - GEnome Data Import
GEM.app - GEnomes Management Application
GeNeMove - German Network of Hereditary Movement Disorders
GERP - Genomic Evolutionary Rate Profiling
GHS – Gordon Holmes syndrome
GJB1 – Cx32
GUI – graphical user interface
HSP - hereditary spastic paraplegias
IMBPFD - inclusion body myopathy, Paget's disease of the bone, frontotemporal dementia
KARS - lysyl-tRNA synthetase
LOD - log of odds
MARS - methionyl-tRNA synthetase
NGS - next-generation sequencing
PDC – phosphatidylcholine
PNPLA6 - phospholipase B/neuropathy target esterase
REEPs - receptor expression-enhancing proteins
SAM - sterile alpha motif
sATX – spastic ataxia
SPG12 – RTN2
SPG17 – seipin
SPG20 – spartin
SPG3 – Atlastin-1
SPG31 – REEP1
VCF – variant call format
WES - whole exome sequencing
YARS - tyrosyl-tRNA synthetase
UNIVERSITY OF MIAMI

IDENTIFYING NOVEL GENES AND GENETIC/PHENOTYPIC SPECTRA FOR INHERITED NEURODEGENERATIVE DISORDERS USING NEXT-GENERATION SEQUENCING

By

Michael Anthony Gonzalez

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the Doctor of Philosophy

Coral Gables, Florida
December 2014
UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

IDENTIFYING NOVEL GENES AND GENETIC/PHENOTYPIC SPECTRA FOR INHERITED NEURODEGENERATIVE DISORDERS USING NEXT-GENERATION SEQUENCING

Michael Anthony Gonzalez

Approved:

__________________________  ________________________
Stephan Zuchner, M.D., Ph.D.  Susan Blanton, Ph.D.
Professor of Human Genetics  Professor of Human Genetics

__________________________  ________________________
Mustafa Tekin, M.D.  John Gilbert, Ph.D.
Professor of Human Genetics  Professor of Human Genetics

__________________________  ________________________
Vance Lemmon, Ph.D.  M. Brian Blake, Ph.D.
Professor of Neurological Surgery  Dean of the Graduate School
Length-dependent axonal degeneration, sometimes referred to as the "dying-back" of a nerve, results in axonal degeneration at the most distal extent of the axon. Degeneration of extended axons in the peripheral (PNS) and central nervous systems (CNS) is the pathological basis for a number of neurological disorders, including the axonal peripheral neuropathies known as Charcot-Marie-Tooth disease type 2 (CMT2) and the axonal degeneration of the corticospinal tract in the CNS known as hereditary spastic paraplegias (HSP). CMT2 and HSP are inherited neurodegenerative disorders characterized by progressive axonal degeneration, which can cause severe disabilities and have no clinically available treatment options. While length-dependent axonal degeneration in CMT2 occurs in the periphery and in HSP in the central nervous system, it is becoming more evident that there is significant biological and genetic overlap between these disorders. In addition, the known genes only explain ~60% of HSP cases and ~30% of CMT2 cases, which suggests many more disease genes are still left to be identified. New technologies, such as exome and genome sequencing, allow us to rapidly increase our knowledge on genes involved in these diseases. Therefore, the overall goal of this project was to use exome sequencing to build
upon and extend the existing knowledge of genetic factors and biological pathways involved in axonopathies. I analyzed >1,000 whole exome datasets from patients with CMT2 and HSP. This analysis has lead to the identification of the novel disease genes MARS and DDHD2, which cause CMT2 and HSP, respectively. However, while producing this large cohort of exome data, it became evident that novel strategies for analyzing genomic-scale data would be needed. To this end, I developed GEnomes Management Application (GEM.app) to address the computational challenges brought forth by genomic 'big data'. This platform has lead to numerous gene identifications (BICD2, GBA2, DDHD1, DDHD2, FBXO38, REEP2, etc). Due to the scalability of the GEM.app platform, I was able to analyze large collections of exome datasets. These analyses lead to the findings that redefined the phenotypic spectrum of two disease genes PNPLA6 and VCP. Given that much of our knowledge of the pathophysiology of these diseases has initially been identified via genetic studies, the value of the present work cannot be underestimated. Further characterization of molecular pathways involved in axonal degeneration will lead to better understanding of biological mechanisms and eventually lead to new therapeutic options for disorders characterized by axon degeneration.
DEDICATION

To my family, girlfriend, and life long friends.
ACKNOWLEDGEMENTS

The Zuchner lab was instrumental in my thesis projects: Fiorella Speziani, Gladys Montenegro, Dr. Adriana Rebelo, Alleeene Strickland, Justin Price, Eric Powell, Alex Abrams, Feifei Tao, Rafael Acosta and Cima Saghira.

Dr. Stephan Zuchner, Dr. Susan Blanton, Dr. Mustafa Tekin, Dr. John Gilbert, Dr. Vance Lemmon for their mentorship throughout my graduate training. Dr. Stephan Zuchner and Dr. Vance Lemmon were vital in building my scientific network.

Dr. Michael Gaines, for his mentorship through the Howard Hughes Medical Institute Bridges to the Future program at the University of Miami. Thank you all the guidance during my undergraduate education and for helping me discover my interests in biomedical research.

The faculty and staff at the Hussman Institute for Human Genomics for providing me with a great education and training on the latest technologies in human genomics research. Dr. Bill Scott and Dori Mclean for all the reminders to meet important deadlines leading up to my dissertation defense.

Sandhya Ramsook for endless hours of editing this manuscript and all the encouragement throughout my graduate work. Thank you for always being
supportive even when I would work throughout the night. Your support, encouragement, quiet patience and unwavering love were undeniably the reason I survived the experience of graduate school.

My friends for always reminding me that I needed some time to let my mind wander from thinking about science.

Mom, Melissa, David and Grandma for your love and support during these crazy times. Your understanding that my time away from you was spent trying to accomplish my goals helped me stay focused and finish my graduate training.
# TABLE OF CONTENTS

| LIST OF FIGURES | ........................................................................................................... x |
| LIST OF TABLES | ......................................................................................................... xii |
| LIST OF ABBREVIATIONS | ........................................................................................................ xiii |

## CHAPTER

1. Introduction to Inherited Neuropathies: Clinical Features, Genetics, and Insights from Biomedical Research ................................................................. 1

   - Overview of the Central Nervous System ............................................ 1
   - Inherited Upper Motor Neuron Disorders: Hereditary Spastic Paraplegia.. 1
   - Molecular Mechanisms in Hereditary Spastic Paraplegia ..................... 4
   - Overview of the Peripheral Nervous System ....................................... 7
   - Inherited Peripheral Neuropathies: Charcot-Marie-Tooth Disease ........ 8
   - Molecular Mechanisms in Axonal Charcot-Marie-Tooth Disease .......... 10
   - Genetic and Biological Comparison in CMT and HSP .......................... 14
   - Next-Generation Sequencing to Study Inherited Neuropathies .......... 19
   - Current Challenges with Next-Generation Sequencing ....................... 21
   - Summary .............................................................................................. 21

2. Exome Sequencing Identifies a Significant Variant in Methionyl-tRNA Synthetase (MARS) in a Family with Late Onset CMT2 ........................................... 23

   - Perspectives .......................................................................................... 23
   - Materials and Methods ........................................................................ 24
Discussion .............................................................................................................. 28

3. Mutations in Phospholipase DDHD2 Cause Autosomal Recessive Hereditary Spastic Paraplegia ................................................................. 35
   Perspectives ...................................................................................................... 35
   Materials and Methods ...................................................................................... 36
   Gene Identification ............................................................................................ 37
   Clinical Description ........................................................................................... 40
   Discussion .......................................................................................................... 43

4. GEnomes Management Application (GEM.app): A New Software Tool for Large-Scale Collaborative Genome Analysis ............................ 49
   Perspectives ...................................................................................................... 49
   Bioinformatics ................................................................................................... 50
   GEDI Pipeline and GEM.app Graphical User Interface .................................... 51
   Basic Principles and Description of GEM.app ................................................... 51
   Data Security ...................................................................................................... 54
   GEDI Pipeline .................................................................................................... 54
   Graphical User Interface .................................................................................... 55
   Performance ....................................................................................................... 59
   Discussion .......................................................................................................... 61

5. A Novel Mutation in VCP Causes Charcot-Marie-Tooth Type 2 Disease .65
   Perspectives ...................................................................................................... 65
   Consent and Exome Sequencing ....................................................................... 66
   Clinical Findings ............................................................................................... 66
Genomic Studies ................................................................. 73
Functional Evaluation .......................................................... 75
Discussion ............................................................................. 78

6. PNPLA6 Mutations Cause Boucher-Neuhaüser and Gordon Holmes Syndromes as Part of a Broad Neurodegenerative Spectrum ........... 82
   Perspectives .......................................................................... 82
   Whole Exome Sequencing of Index Patients with Boucher-Neuhaüser and Gordon Holmes Syndromes .................................................. 83
   Sanger Sequencing of the Candidate Gene in Additional Boucher-Neuhaüser Syndrome Families ............................................................ 85
   PNPLA6 Analysis in More than 500 Exomes of Index Patients with Ataxia, Hereditary Spastic Paraplegia and Charcot-Marie-Tooth Disease ........ 85
   Structural Models of Wild-Type and Mutant PNPLA6 Domains ............. 86
   Clinical Assessment .................................................................. 86
   PNPLA6 Causes Boucher-Neuhaüser and Gordon Holmes Syndrome .. 87
   PNPLA6 Causes a Wide Spectrum of Neurodegenerative Disease ....... 90
   Structural in silico Protein Modeling ........................................... 93
   Clinical Findings ..................................................................... 98
   Discussion ............................................................................ 102

7. Summary, Future Directions, and Discussion .............................. 106
   Summary of Dissertation Findings ........................................... 106
   Expanding the Number of Axonopathy Genes: MARS and DDHD2 .... 107
   Managing the Challenges of Genome Analysis ............................. 108
# LIST OF FIGURES

## Figure

1.1 Central Nervous System ................................................................. 2
1.2 Motor and Sensory Neurons ......................................................... 9
1.3 Peripheral Neuron ................................................................. 12
1.4 Similarities and differences in the molecular pathology of CMT and HSP ................................................................. 17
2.1 Pedigree of family with MARS variant ........................................ 25
2.2 Multiple sequence alignment of MARS ........................................ 29
2.3 Structural protein model of MARS ............................................. 30
2.4 Study of MARS Variants in Yeast ............................................. 33
3.1 Pedigrees and sequence traces of the DDHD2 families ................. 39
3.2 Schematic of the DDHD2 gene .................................................. 41
3.3 Cranial MRI/CT scans of DDHD2 patients ................................... 42
4.1 GEM.app graphical user interface (GUI) ........................................ 53
4.2 Example of output from a GEM.app analysis query ....................... 56
4.3 GEM.app annotation manager ...................................................... 58
4.4 Performance of GEM.app ............................................................. 60
4.5 Current usage of GEM.app .......................................................... 62
5.1 Pedigree of family with VCP variant .......................................... 67
5.2 VCP gene-network ................................................................. 76
5.3 VCP gene schematic and protein model ...................................... 77
5.4 Functional evaluation of the VCP mutation .................................. 79
6.1 Pedigrees and Segregation of the Mutations Identified in PNPLA6 ...91
6.2 Electropherograms of Mutations Identified with Sanger Sequencing .92
6.3 Evolutionary conservation of PNPLA6 mutations ..........................94
6.4 Schematic of PNPLA6 gene ....................................................96
6.5 Structural model of EST domain (grey) of human PNPLA6 ..........97
6.6 Structural models of wild type and mutant CNB1 and CNB2 domains in complex with cAMP ................................................................. 99
6.7 The continuous spectrum of PNPLA6-associated disease ..........100
6.8 The phenotypic spectrum caused by PNPLA6 mutations ............104
7.1 Graph illustrating the nature of the reductions in DNA sequencing costs ..........................................................................................109
LIST OF TABLES

Table

2.1 Top variants remaining after applying filter criteria on exome sequencing data .................................................................27
2.2 Corresponding protein residues in human and yeast .................32
3.1. Five segregating variants in family THI26003 after filtering of sequencing data ........................................................................38
3.2 Clinical features of genetically confirmed DDHD2 patients ..........44
5.1 Clinical features of family with E187K VCP mutation .................70
5.2 Motor Nerve Conduction Velocities ........................................71
5.3 Sensory Nerve Conduction Velocities ....................................72
5.4 in silico scores ........................................................................74
6.1 Clinical features of PNPLA6 patients .......................................88
6.2 PNPLA6 mutations identified in this study ..............................89
LIST OF ABBREVIATIONS

AARS - alanyl-tRNA synthetase
ALS - amyotrophic lateral sclerosis
ARCA s - Autosomal recessive cerebellar ataxias
ARS - aminoacyl-tRNA synthetases
BNS - Boucher-Neuhäuser syndrome
CMT - Charcot–Marie–Tooth disease
CMT1 - demyelinating CMT
CMT2 - axonal degenerative CMT
CMTNSv2 - CMT Neuropathy Score
cNMPs - cyclic nucleotide-monophosphate binding
ERGIC - cis-Golgi and the ER-Golgi intermediate compartment
FA2H - fatty acid-2 hydroxylase
GARS - glycyl-tRNA synthetase
GEDI - GEnome Data Import
GEM.app - GEnome Management Application
GeNeMove - German Network of Hereditary Movement Disorders
GERP - Genomic Evolutionary Rate Profiling
GHS – Gordon Holmes syndrome
GJB1 – Cx32
GUI – graphical user interface
HSP - hereditary spastic paraplegias
IMBPFD - inclusion body myopathy, Paget's disease of the bone, frontotemporal dementia
KARS - lysyl-tRNA synthetase
LOD - log of odds
MARS - methionyl-tRNA synthetase
NGS - next-generation sequencing
PDC – phosphatidylcholine
PNPLA6 - phospholipase B/neuropathy target esterase
REEPs - receptor expression-enhancing proteins
SAM - sterile alpha motif
sATX – spastic ataxia
SPG12 – RTN2
SPG17 – seipin
SPG20 – spartin
SPG3 – Atlastin-1
SPG31 – REEP1
VCF – variant call format
WES - whole exome sequencing
YARS - tyrosyl-tRNA synthetase
Overview of the Central Nervous System

The central nervous system is responsible for analyzing sensory input and stimulating muscle contraction and glandular secretion. It is comprised of the spinal cord and the brain, which consists of the cerebrum, cerebellum, and brainstem. The spinal cord is the channel for conducting signals between the brain and the rest of the body, whereas the brain is responsible for the integration of most sensory information and the coordination of body function. The pyramidal motor system includes the corticospinal tract, which is a collection of upper motor neurons that are involved in voluntary movement. These neurons have cell bodies in the cerebral motor cortex and axons that project to the spinal cord. Many diseases with motor symptoms, such as Parkinson's disease, multiple sclerosis, and the hereditary spastic paraplegias, are characterized by lesions in the corticospinal tract that leads to dysfunction of the central nervous system.

Inherited Upper Motor Neuron Disorders: Hereditary Spastic Paraplegia

Hereditary spastic paraplegias (HSP) are a collection of inherited neurologic disorders that were first described at the end of the 19th century by Adolph Strumpell (1880) and Maurice Lorain (1898). HSP is characterized by a
Figure 1.1. **Central Nervous System.** The central nervous system is comprised of the brain and the spinal cord. The brain is made up of the cerebral cortex, subcortical structures, brain stem, and cerebellum. *The Neuroscience on the Web Series: CSU, Chico, Patrick McCaffrey, Ph.D.*
length-dependent distal axonopathy of the corticospinal tracts, causing progressive weakness (paraplegia) and increased muscle tone and stiffness (spasticity) in the lower limbs. HSP is a clinically heterogeneous disorder of the central nervous system that is estimated to affect 3 in 100,000 people worldwide. Common symptoms of HSP are muscle spasms, balance difficulties, spastic gait, unexplained falls, and bladder dysfunction. Clinically, HSP can be divided into pure HSP and complicated HSP. In complicated forms of HSP, patients will have spastic paraplegia associated with additional neurological features such as: impaired control of voluntary movements (ataxia), hearing loss, vision loss, and mental retardation. HSP shows a very broad range of severity; onset can be as early as childhood or as late as 70 years of age. The onset of HSP is usually subtle, beginning with the development of leg stiffness. The German Network of Hereditary Movement Disorders (GeNeMove) consortium has attempted to establish clinical characteristics that are shared by most genetic subtypes of HSP: 1) spastic paraplegia is a major and severe feature of the phenotype and 2) symptoms progress slowly and continuously, without shortening the lifespan of the patient. However, there are exceptions to these guidelines. In some complicated forms of HSP, life expectancy can be reduced.

The genetics of HSP is complex due to the high rate of locus and allele heterogeneity; all modes of Mendelian inheritance have been observed. Over 50 loci have been mapped and at least 25 genes that contribute to HSP have been identified. The most common subtype of HSP (SPG4) is caused by
mutations in the gene \textit{SPAST}. These mutations account for more than 50% of autosomal dominant HSP cases and more than 15% of isolated HSP cases \cite{12}. Therefore, it is important to screen patients for \textit{SPAST} mutations before considering whole exome sequencing. Mutations in \textit{ATL1} are the second most common form of HSP, and account for less than 10% of HSP. Many of the additional genes account for less than 1\% of HSP cases \cite{7}.

\textbf{Molecular Mechanisms in Hereditary Spastic Paraplegia}

Many groups around the world have shown that lipid transport, mitochondrial dynamics, and endosomal trafficking/structure play an important role in the pathogenesis of HSP. Years of research have elucidated additional genes that are involved in these mechanisms. Some of these findings suggest that a specialized neuronal cytoskeletal scaffold is important to maintain transport of molecules within axons of the corticospinal tract \cite{3}. These axons can be up to 1 meter long, requiring adaptable intracellular machinery for sorting and distributing proteins, lipids, mRNAs, organelles and other molecules over such a long distance \cite{13}. It has been established that defects in the pyramidal motor system lead to a number of neurological disorders, such as HSP.

Many genetic and functional studies have established that endosomal trafficking is a critical biological mechanism in HSP \cite{14}. \textit{SPAST}, the most commonly mutated gene in HSP, plays a critical role in shaping the morphology of the endoplasmic reticulum \cite{15}. Montenegro et al. demonstrated that \textit{SPAST},
along with other HSP genes (*RTN2, ATL1, and REEP1*) are vital for maintaining normal tubular endoplasmic reticulum morphology \(^{16}\). These findings suggest that mutations in genes that affect the formation and stabilization of the tubular ER network can lead to ER stress, which is hypothesized to cause HSP. Recently, Hashimoto et al. demonstrated that mutations in *ZFYVE27*, responsible for causing a subset of HSP known as SPG33, could also lead to ER stress \(^{17}\). They found that ZFYVE27 directly interacts with many proteins that are implicated in HSP or have very similar characteristics to HSP-associated proteins, in agreement with previous reports \(^{16, 17}\). Other studies have shown that additional endosomal trafficking pathway proteins *NIPA1, SPG20, and SPG21* can also cause HSP \(^{18-20}\). The discovery of mutations in these genes has solidified the role of endosomal trafficking in HSP.

The first gene to implicate lipid metabolism in HSP was reported by Goizet et al. where *CYP7B1*, which is involved in cholesterol metabolism, was first identified in nine families \(^{21}\). This came as a surprise to the HSP research community because lipid metabolism was not yet identified as an important biological process for HSP. Lipids have important biological functions that include signaling and acting as structural components of cell membranes \(^{22}\). Therefore, when one considers the large amounts of intracellular organelle membranes in motor neurons, it is not surprising that defects in genes involved in lipid metabolism can give rise to HSP. Together with various collaborators, I have contributed to the identification of several mutations in genes involved in lipid
metabolism: \textit{DDHD1, DDHD2, GBA2, PNPLA6, CYP2U1, B4GALNT1}. This gives further evidence for the significance of lipid metabolism in HSP \textsuperscript{23-27}. However, more work is needed to fully understand this biological mechanism and its role in axon degeneration.

Mitochondrial architecture and bioenergetics with increased oxidative stress have been shown to be important for HSP pathogenesis \textsuperscript{26}. Several HSP genes, including \textit{SPG7, C12orf65, and HSPD1}, have been reported to have mitochondrial functions. \textit{SPG7} knockout mice develop neuropathy that is directly correlated to the development of morphologically abnormal mitochondria \textsuperscript{28}. \textit{HSPD1} is a mitochondrial chaperon that is essential for the folding and assembly of newly imported proteins in the mitochondria. Mutations in \textit{HSPD1} found in HSP patients were replicated in \textit{E. coli}, and were observed to disrupt the normal growth of \textit{E. coli} cells, suggesting the pathogenicity of these variants are due to a loss-of-function of this specific mitochondrial chaperon \textsuperscript{29}. Shimazaki et al. report that mutations in \textit{C12orf65}, which encodes a mitochondrial matrix protein, can cause HSP, displaying yet another example of HSP-associated proteins with functions involved in mitochondrial processes \textsuperscript{30}. \textit{REEP1}, which has been established as one of the most commonly mutated genes in HSP, encodes a protein that localizes to the mitochondria, although more work is needed to further elucidate its function \textsuperscript{31}. 
Taken together, many of the genetic findings along with functional characterizations of HSP have helped to better characterize the underlying mechanisms of the disorder. Specifically, these findings have demonstrated that lipid metabolism, endosomal trafficking, and mitochondrial processes play important roles in HSP. Our knowledge of the mechanisms of how these defects can give rise to degeneration of the corticospinal axons is still limited, but the identification of novel genes associated with HSP is providing better insight into the mechanisms underlying neurodegenerative diseases.

**Overview of the Peripheral Nervous System**

The peripheral nervous system is composed of the nerves and neurons outside of the brain and spinal cord and the optic nerve. The peripheral nervous system includes the neurons that connect the central nervous system to effectors, glands and muscles throughout the body. It relays sensory information from the body to the central nervous system \(^1\). In order to accomplish this, there are two main portions of the peripheral nervous system: sensory neurons and motor neurons (Figure 1.2). Sensory neurons process external environmental stimuli and ultimately convey sensory information to the central nervous system \(^32\). Motor neurons are efferent neurons with cell bodies originating in the ventral root of the spinal cord. They relay messages from the central nervous system to the skeletal muscles, glands, and other effector tissues \(^32\). To transfer information to and from the cell body, sensory and motor neurons possess two forms of projections: axons and dendrites \(^33\). Dendrites are the short, branched projections
of a neuron that receive signals from other neurons (Figure 1.2). Axons are long, slender projections of neurons that propagate electrical impulses away from the cell body (motor neurons) or towards the cell body (sensory neurons) (Figure 1.2). Defects in the axonal structure are highly detrimental to the neurons in the peripheral nervous system. Failure to maintain healthy motor and sensory neurons has been implicated in a number of neurodegenerative diseases, such as Charcot-Marie-Tooth disease.

**Inherited Peripheral Neuropathies: Charcot-Marie-Tooth Disease**

Charcot–Marie–Tooth disease (CMT) represents a clinically heterogeneous neurologic disorder, which was first described by three physicians Jean Martin Charcot, Pierre Marie, and Howard Henry Tooth in the 19th century. CMT is one of the most common inherited neuromuscular disorders, with a prevalence of 1 in 2,500 worldwide. CMT is a neuromuscular disorder characterized by progressive and length-dependent degeneration of peripheral nerves that results in muscle weakness and wasting in distal limbs. CMT patients lose strength in their peripheral muscles and may exhibit classic hand and foot deformities, such as fingers being locked in a flexed position and pes cavus (high arches in the feet), respectively. CMT can be clinically divided into two groups: demyelinating CMT (CMT1) and axonal degenerative CMT (CMT2). CMT1 patients are observed to have slowed nerve conduction velocities and nerve demyelination, whereas CMT2 patients show normal nerve conduction and loss of myelinated axons without evidence of primary demyelination. Classically,
Figure 1.2. **Motor and Sensory Neurons.** The peripheral nervous system contains two main types of neurons, motor neurons and sensory neurons. These neurons contain dendrites, axons, cell bodies, and Schwann cells. Motor neurons are responsible for sending singles to targets. In contrast, sensory neurons transmit input stimulus towards the cell body. *Review of the Universe, Nervous System, Nerves and Neurons: universe-review.ca*
CMT1 has an earlier onset compared to CMT2, but there is a wide range of onset and disease severity \(^\text{38}\). In CMT2, there is a chronic length-dependent axonopathy of the peripheral nerve, which is a similar pathogenic theme to what is observed in HSP \(^\text{39}\).

CMT is a heterogeneous Mendelian disorder with more than 80 genes implicated in the disease \(^\text{40}\). CMT1 makes up approximately one-third of all CMT cases, in which 70-80\% of these cases are caused by \textit{PMP22} duplications \(^\text{38}\). CMT2 cases make up a larger fraction of all CMT cases. Genes that have been identified so far explain only \(~30\%\) of CMT2 cases. Mutations in \textit{MFN2} and \textit{GJB1} account for about 25\% of the cases of CMT2 with a known causative mutation \(^\text{38}\). Therefore, my work has focused on whole exome sequencing of CMT2 families that are negative for mutations in \textit{MFN2} and \textit{GJB1} in order to identify novel genes involved in axonal degeneration.

\textbf{Molecular Mechanisms in Axonal Charcot-Marie-Tooth Disease}

In the human body, peripheral nerves are responsible for communicating information from the central nervous system to distant targets. Axons of peripheral nerves are myelinated by Schwann cells (Figure 1.3). Schwann cells insulate axons in order to allow electrical signals to travel at high speeds. Similar to axons in the central nervous system, axons in the peripheral nervous system require specialized machinery to transport proteins and other molecules across
long distances in order to maintain neuronal homoeostasis $^{41}$. Without intact axons and myelin sheaths, peripheral nerve cells are unable to activate target muscles or relay sensory information from the limbs back to the brain $^{36}$.

Mitochondria are dynamic organelles that are constantly fusing and dividing to form tubular networks, which is an important part of mitochondrial dynamics. Recent discoveries have demonstrated that these dynamic properties of mitochondria (fusion, fission, movement, and mitophagy) are vital to the health of neurons $^{42}$. A number of neurodegenerative disorders (HSP, Alzheimer's disease, Parkinson's disease) have been shown to be caused in part by mitochondrial dysfunction. Several genes that encode proteins that function in the mitochondria, such as $MFN2$, $DNM2$, and $GDAP1$, are mutated in patients affected by CMT2 $^{43-45}$. The most common CMT2-associated gene, $MFN2$, has been shown to play a major role in mitochondrial dynamics, specifically in the fusion of mitochondria $^{46}$. It has been demonstrated that mutations in $GDAP1$ and $DNM2$ lead to neurodegeneration caused by altered mitochondrial dynamics in $Drosophila$ and mouse models $^{47-49}$. Neurofilaments have also been implicated in disruption of mitochondrial dynamics. $NEFL$ and $NEFH$ belong to a class of intermediate filament proteins that are exclusively expressed in mature neurons and mutations in them have been reported to cause CMT2 $^{50, 51}$. Recently, it has been demonstrated that, when mutated, these proteins can affect mitochondrial morphology by interfering with mitochondrial dynamics $^{52, 53}$. 
Figure 1.3. **Peripheral Neuron.** The peripheral nerves are a collection of neurons. Each of these neurons consist of axons that are insulated by Schwann cells in order to transmit signals at high speeds. *Review of the Universe, Nervous System, Nerves and Neurons: universe-review.ca*
Previous studies have demonstrated that abnormal neurofilament disrupts the assembly of neurofilament, axonal transport, and the localization of mitochondria in neurons. Previous research has demonstrated that axonal transport is vital to maintaining healthy axons. Therefore, if axonal transport is impeded, molecules vital to maintain healthy axons may not travel to the distal ends of these neurons. This is a plausible hypothesis that can explain how CMT2 patients present with a length-dependent neuropathy. A 2009 review by Gentil and Cooper reported that a number of CMT2 genes (MFN2, KIF1B, DYNC1H1, Rab7, TRPV4, NEFL, HSPB1, MPZ, KIF1B, DYNC1H1 and HSPB8) play important roles in axonal transport and further suggested that the main underlying molecular mechanism in CMT is axonal transport dysfunction.

Just as the identification of the involvement of lipid metabolism was a surprise to the HSP community, the identification of mutations in genes coding for aminoacyl-tRNA synthetases (ARS) came as a surprise to CMT researchers. Antonellis et al. were the first to suggest this pathway for CMT2 when they identified missense mutations in GARS that co-segregated with disease in CMT2 families. It has since been demonstrated that in mice, mutant GARS proteins lead to a gain-of-function that is toxic to peripheral neurons. Since the association of GARS with CMT2 was published, there have been reports that link mutations in other ARS genes, KARS, AARS, MARS, and HARS, to CMT2. Although CMT2 caused by GARS mutations appears to be a gain-of-function mechanism, mutations in other aminoacyl-tRNA synthetases appear to be a loss-
of-function mechanism \(^{61-64}\). There are a number of hypotheses suggesting how defects in a basic cellular process, such as aminoacylation, can cause such a specific phenotype as peripheral neuropathy \(^{65}\). However, further research is still needed to understand the underlying mechanism of how these mutant aminoacyl-tRNA synthetases lead to peripheral neuropathy. Although the role of aminoacyl-tRNA synthetases in CMT2 was elucidated recently, it appears to play a significant role in a portion of peripheral neuropathies.

**Genetic and Biological Comparison in CMT and HSP**

The identification of the remaining CMT2 and HSP genes is expected to yield important insights into the pathways and pathophysiology associated with axonal degeneration. Our understanding of biological mechanisms that underlie disease pathology has significantly improved due to the number of genes identified for axonopathies \(^3, 66-67\). In addition, it is becoming evident that the phenotypic and genotypic intersection of motor neuron loss, axon degeneration, and other symptoms is more extensive than previously thought \(^{66}\). While length-dependent axonal degeneration occurs in the peripheral nervous system in CMT2, and in the central nervous system in HSP, it is becoming evident that there is significant biological and genetic overlap between these disorders \(^4, 66\). Therefore, identification of genes involved in CMT2 and HSP can directly influence our knowledge of neurodegeneration.
Many of the genes identified in CMT2 are involved with similar biological functions observed in HSP, which is not surprising when one considers the shared pathogenic theme of progressive axonal degeneration (Figure 1.4). It is clear that many molecular pathological causes of CMT2 and HSP overlap, such as disruption of axonal transport and mitochondrial dysfunction. Previous reports have suggested that dysfunction of mitochondrial mobility, fusion, fission, and axonal transport are key elements that lead to neuropathies. For example, there are a number of gene products that have been identified for CMT2, such as *MFN2*, *DNM2*, and *GDAP1*, which are associated with mitochondrial and endosomal trafficking. Reports have demonstrated that mutant *MFN2* proteins are stable and normally expressed in cells, but they fragment mitochondria and disrupt their axonal transport. *GDAP1* mutations have also been shown to induce mitochondrial fragmentation, further evidence for the role of mitochondrial dynamics in CMT2. Many genes that are known to cause HSP when mutated are involved with similar processes. *REEP1*, the third most common gene affected in HSP patients, was shown to localize to mitochondria, suggesting that it plays a role in mitochondrial dynamics. From these findings, it appears that axonal degenerative disorders may be biologically linked by alteration of mitochondrial dynamics.

A majority of the genes identified for HSP and CMT2 are involved with membrane trafficking and mitochondrial regulation. In addition, other biological functions such as organelle shaping, the process responsible for the 3D structure
of organelles, seem to be involved \cite{3,39}. Various studies have demonstrated that $SPA\text{ST}, \, ATL1, \, RTN2, \, REEP1$ are ER morphogens and are interacting partners $^{72-75}$. Many of the most common HSP genes ($SPA\text{ST}, \, ATL1, \, REEP1$) account for $\sim 60\%$ of cases and are involved with formation of the tubular endoplasmic reticulum $^{76}$. It appears that dysfunction of endoplasmic reticulum shaping is the most common pathogenic mechanism of disease in HSP $^{13}$. In contrast, mitochondrial dynamics appear to be one of the most common pathogenic themes in CMT2 where nearly $20\%$ of CMT2 cases are explained by mutations in $MFN2^{77}$.

A plethora of $in \, vivo$ studies have demonstrated that disrupting transportation along the axon may be pathogenic. It has been shown that mice with truncating mutations in $SPA\text{ST}$ exhibit disrupted anterograde axonal transport and manifest gait disturbances similar to those seen in HSP patients $^{78}$. Mutations that cluster in the kinesin domain of KIF5A, found in HSP cases, lead to halting of anterograde axonal transport in $Drosophila \, melanogaster^{79}$. Similar phenomena have been observed in animal models for CMT2. For example, $NEFL$ knockout mice have subtle phenotypes, but after severing of the peripheral nerves, $NEFL$ knockout mice show delayed regeneration of myelinated axons and slowed axonal traffic $^{80}$. 
Figure 1.4. Similarities and differences in the molecular pathology of CMT an HSP. The HSP associated genes are indicated in red and the CMT neuropathy associated genes are in blue. Genes involved in both HSP and CMT diseases are black. V. Timmerman et al. / Experimental Neurology 246 (2012) 14-25.
As we sequence the genomes of additional affected individuals, I believe we will observe that many neuropathy-associated genes harbor different alleles that can cause either CMT2 or HSP. When I began my work on CMT and HSP, *BSCL2* and *ATL1* were the only examples of such genes, but I have found throughout my graduate work that other genes can lead to a spectra of disorders. \(^66\) I have contributed to the identification of CMT2 mutations in *VCP*; a gene previously associated with HSP, that can also cause CMT2 \(^81, 82\). These findings demonstrate that we have just begun to identify the genetic etiologies of axonopathies.

In 2014, Novarino et al. exome sequenced an HSP cohort, performed network analysis, and described the “HSPome” \(^83\). They reported that genes associated with amyotrophic lateral sclerosis (ALS), Alzheimer’s disease, and Parkinson’s disease significantly overlap with HSP genes, suggesting there are underlying genetic/biological links in neurodegenerative disorders. In contrast, they did not find an association with sets for representative neurodevelopmental disorders such as autism spectrum disorders and epilepsy nor with non-neurological disorders represented by heart and pulmonary disorders \(^83\). This demonstrates that the “HSPome” is specifically associated with other neurodegenerative disease gene-networks and suggestive of an underlying mechanistic link between neurodegenerative diseases. I believe that these
findings are further evidence for the hypothesis that neurodegenerative disorders have common biological pathways.

It is clear that many molecular causes of CMT2 and HSP overlap in areas such as disruption of axonal transport and mitochondrial dysfunction. Tremendous progress has been made in understanding the biological mechanisms of CMT2 and HSP due to successful identification of genes. However, our knowledge of how mutations in these genes cause HSP and CMT2 is still limited. The identification of additional genes involved in axonal degeneration can help improve biological model systems and further describe pathways that are important for neurodegenerative diseases. Mapping of the remaining genes involved in axonal degeneration will lead to significant insight into the pathophysiological mechanisms of CMT2 and HSP and therapies for neurodegenerative diseases.

**Next-Generation Sequencing to Study Inherited Neuropathies**

Next-generation sequencing has revolutionized human genetics and the pace at which causative genes are identified. Whole exome sequencing, capture and sequencing of protein-coding regions of the genome, is a rapid, high-throughput, and cost-effective approach for gene identification. Whole exome sequencing has been widely used to identify pathogenic variation, especially in Mendelian disorders. Until whole genome sequencing becomes more affordable, sequencing of the exome is justified because of the large
amount of protein-coding variation implicated in Mendelian disorders \cite{87, 88}. Additionally, the recent validation of an excess of rare exonic variation (<0.5% minor allele frequency) in the human species provides further support for the hypothesis that rare exonic variants have led to higher rates of allelic and locus heterogeneity and will explain a significant portion of the burden in human diseases \cite{89-91}.

Recently, many studies have exploited next-generation sequencing technologies and have identified several novel genes involved in inherited neuromuscular disorders \cite{74, 92-94}. This is vital because the known genes only explain ~60% of HSP cases and ~30% of CMT2 cases, which suggests many more genes remain unidentified \cite{95, 96}. The success of whole exome sequencing (WES) to identify pathogenic variation in various diseases has stimulated the sequencing of disease cohorts \cite{97-99}. It is likely that the remaining genes for axonopathies will have a relatively low frequency. In order to detect rare novel genes, large sample sizes will be necessary. For example, to identify a gene that is responsible for 2% of recessive HSP cases, 300 samples are required to achieve ~95% power \cite{100}. Many small and medium-sized clinical and research-based investigative teams around the world are generating data that, if combined and shared, will significantly increase power for the entire community to identify new genes.
Current Challenges with Next-Generation Sequencing

The cost for DNA sequencing has been reduced by a factor of more than 100,000 in the past 10 years, which is arguably unprecedented for any field. As a result, we are now seeing the tip of the iceberg of the data deluge that genomics will soon experience. Computation of “big data” is not a problem that is unique to the field of genomics; many resources and new solutions are being implemented across fields. Examples include massive public cloud solutions (Amazon EC2), new types of databases (Google BigTable), new programming models such as MapReduce (Apache Hadoop), and improved platform-independent standardization of versatile web protocols (HTML5). Of course, highly specialized experts drive all of these new approaches. In genomics research (and clinical applications), the majority of scientists are non-bioinformaticians and it is unlikely that they all will become programmers. It appears that there is an extraordinary need and opportunity to train scientists who are familiar in both worlds: bioinformatics and molecular genetics.

Summary

Although the genetic basis of HSP and CMT is partly understood, only a fraction of affected individuals receive a genetic diagnosis. A better understanding of the underlying molecular causes of CMT and HSP will ultimately lead to a better understanding of the mechanisms underlying neurodegenerative disease in general. In my dissertation, the focus was on applying new sequencing technologies to identify novel genetic defects that
caused inherited neurodegenerative diseases. In the next chapters, I will discuss studies where I identified novel genes, developed a platform for analyzing genomic data, and refined the phenotypic spectrum of two neuropathy-associated genes. I hope these findings will provide information to the research community that will eventually lead to new diagnostics and therapeutics for neurodegenerative disorders.
CHAPTER 2

Exome Sequencing Identifies a Significant Variant in Methionyl-tRNA Synthetase (MARS) in a Family with Late-Onset CMT2

The data presented in this chapter were previously published in *The Journal of Neurology, Neurosurgery and Psychiatry* (Volume 84, Issue 11) \(^6\). I contributed to the identification of MARS and wrote the manuscript. I performed all of the work presented with three exceptions. First, Mary Reilly performed all clinical examinations and patient sample collections. Second, the laboratory of Anthony Antonellis performed all of the yeast functional studies. Finally, Xiang-Lei Yang and Min Guo performed all protein modeling.

**Perspectives**

Charcot-Marie-Tooth (CMT) disease is a genetically heterogeneous disorder of the peripheral nerve, which is clinically divided into primarily demyelinating CMT1 and axonal CMT2. Despite astounding progress in gene identification in CMT, up to an estimated 70% of CMT2 patients do not have a mutation in any of the known genes \(^77, 102\). The identification of the remaining CMT2 genes is expected to yield important insights into the pathways and pathophysiology associated with axonal degeneration. This will ultimately lead to focused studies aimed at drug development. In addition, it is becoming evident that the phenotypic and genotypic intersection of CMT2 with diseases related to motor neuron and axon degeneration is more extended than previously thought.
The recent introduction of exome sequencing and ultimately whole genome sequencing will be essential to the mapping of exact genotype/phenotype relationships in the coming decade. Currently, exome sequencing offers an economic opportunity to increase the gene discovery pace and, importantly, investigators are able to take advantage of relatively small families or even single cases. However, the presented case study offers insight into the challenges that may come with describing new rare causes of CMT2.

**Materials and Methods**

We have studied a family with late-onset CMT2 and incomplete phenotypic penetrance. Examination of the index case at age 50 (Figure 2.1; III.1) revealed bilateral foot drop, distal wasting in the upper and lower limbs, mild distal weakness in the upper limbs to Medical Research Council scale grade 4, but equal proximal and distal weakness in the lower limbs with both hip flexion and ankle dorsiflexion and plantar flexion being grade 4. The 81-year-old uncle (Figure 2.1; II.3) of the index case has an axonal neuropathy presenting at age 67. Other causes of a peripheral neuropathy including diabetes mellitus were excluded. The mother (Figure 2.1; II.2) of the index case was clinically unaffected and showed normal neurophysiology at age 85. Unusual features in the family include equal proximal and distal motor involvement in the lower limbs and the presence of neuropathic pain. These clinical results and the pedigree suggested
Figure 2.1. Pedigree of Family with MARS Variant. Pedigree of dominant CMT2 family displaying segregation of Arg618Cys. Two affected male individuals were analysed by exome sequencing (circles).
a hereditary late-onset X linked or autosomal-dominant CMT2 with incomplete penetrance.

To identify the underlying genetic cause we applied next-generation sequencing of whole exomes (Figure 2.1). Analysis focused on nonsynonymous, splice-site and protein-coding indel variants that segregated under an autosomal-dominant and X-linked model. First, we analyzed the known X linked CMT genes GJB1, AIFM1, PRPS1 and PDK3, but did not identify any variants despite excellent coverage in the exomes from the two male patients. A total of 3,044 variants that met the initial filtering criteria were further filtered for conservation (GERP>3 OR PhastCons>0.7), predicted consequence on protein function (PolyPhen2>0.5 OR unknown) and a minor allele frequency of less than 0.005% in the Exome Variant Server (http://evs.gs.washington.edu/EVS/) and dbSNP137. We also compared the results with 1,236 exomes from different phenotypes available in our own database. In this family, only four missense variants passed these filters (Table 2.1). The variant with the highest conservation score (GERP=5.04, PhastCons=1) resided in the methionyl-tRNA synthetase (MARS or MetRS) gene. Sanger sequencing validated segregation of the c.1852C<T (p.Arg618Cys, chr12:57906632 (hg19)) variant in MARS in the two affected male family members as an autosomal dominant trait with reduced penetrance. As expected, the clinically unaffected mother of the index case, II.2, was an obligate heterozygote. MARS is an excellent CMT2 candidate gene because four aminoacyl-tRNA synthetase (ARS) genes have been shown to
Table 2.1. Top variants remaining after applying filter criteria on exome sequencing data.

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>REF</th>
<th>VAR</th>
<th>Gene</th>
<th>Function</th>
<th>PolyPhen</th>
<th>PhastCons</th>
<th>GERP</th>
<th>Grantham</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>57906632</td>
<td>C</td>
<td>T</td>
<td>MARS</td>
<td>Missense</td>
<td>1.000</td>
<td>1.000</td>
<td>5.04</td>
<td>180</td>
</tr>
<tr>
<td>1</td>
<td>155030876</td>
<td>C</td>
<td>A</td>
<td>ADAM15</td>
<td>Missense</td>
<td>0.970</td>
<td>1.000</td>
<td>5.02</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>203407059</td>
<td>G</td>
<td>A</td>
<td>BMPR2</td>
<td>Missense</td>
<td>0.988</td>
<td>1.000</td>
<td>4.82</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>3839684</td>
<td>C</td>
<td>T</td>
<td>ATP2A3</td>
<td>Missense</td>
<td>0.996</td>
<td>1.000</td>
<td>4.11</td>
<td>56</td>
</tr>
</tbody>
</table>
cause axonal forms of CMT: glycyl-tRNA synthetase (GARS), tyrosyl-tRNA synthetase (YARS), alanyl-tRNA synthetase (AARS) and lysyl-tRNA synthetase (KARS)\textsuperscript{59, 63, 103, 104}.

We then Sanger sequenced all protein-coding exons of MARS in 400 unrelated CMT2 patients, but did not identify any additional pathogenic changes. A search of the entire database of 1,236 exomes, including 466 families with CMT, hereditary spastic paraplegia, amyotrophic lateral sclerosis (ALS) and other related phenotypes, only revealed one additional non-synonymous variant in MARS (c.1448G->A, p.Arg483His). The affected family, however, was previously diagnosed with CMT1A due to PMP22 gene duplication.

**Discussion**

The variant identified in the described family is located in the catalytic domain of MARS (MetRS) and the residue Arg618 is unusually strictly conserved from bacteria to human (Figure 2.2). In absence of a crystal structure, we performed *in silico* modelling of the core domains of human MetRS (Figure 2.3). Arg618 is located at the interface of the catalytic domain and the anticodon-binding domain. Interactions with both domains suggest that Arg618 plays an important role in stabilizing the domain interface. The Arg618Cys substitution could potentially cause a neomorphic structural opening in the protein. The hypothetical structural opening between the catalytic domain and the anticodon-binding domain may also affect the tRNA aminoacylation function of MARS,
Figure 2.2. Multiple Sequence Alignment of MARS. Sequence alignment of the MetRS proteins from bacteria to human showing that Arg618 is a strictly conserved residue during evolution.
Figure 2.3. Structural Protein Model of MARS. Structural model of human MetRS showing that Arg618 is located at the interface of the catalytic domain (blue) and the anticodon-binding domain (yellow), with the guanidinium side chain forming a strong salt-bridge with the side chain of Asp292 from the catalytic domain, and extensive hydrogen-bonding interactions with the backbone carbonyl oxygens of Phe347 and Asn348 from the catalytic domain and of Pro758 and Tyr759 from the anticodon-binding domain.
which is essential for the translation process, and for viability of all organisms. The result may be a gain of function or a loss of function as observed in other CMT-associated ARS mutants.\textsuperscript{105}

To assess the functional consequences of Arg618Cys MARS \textit{in vivo}, we modeled the variant in the yeast ortholog $MES1$, and determined the ability of Arg618Cys $MES1$ to rescue deletion of endogenous $MES1$ compared with wild-type $MES1$, and a common, non-synonymous, non-pathogenic missense change (Arg727Gln MARS, dbSNP rs113808165; Table 2.2). An insert-free pRS315 construct was unable to rescue the mes1Δ allele, whereas both wild-type and Arg727Gln $MES1$ were able to fully complement mes1Δ (Figure 2.4). These data are consistent with $MES1$ being an essential gene, and with the wild-type and R727Q experimental $MES1$ vectors expressing functional proteins, respectively. In contrast, Arg618Cys $MES1$ was unable to rescue the mes1Δ allele (Figure 2.4). Combined, these data indicate that Arg618Cys MARS represents a loss-of-function allele in vivo.

Our results represent a typical conundrum in the new arena of exome sequencing. While a gene has been identified in a single small CMT2 family with strong functional and evolutionary support, sequencing of 1,236 exomes from a diverse set of related and unrelated phenotypes and Sanger sequencing of 400 unrelated CMT2 cases revealed only a single additional relevant change in MARS, albeit in a CMT1 family already carrying a $PMP22$ duplication (CMT1A).
Table 2.2. Corresponding protein residues in human and yeast.

<table>
<thead>
<tr>
<th>Human MARS&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Yeast MES&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.Arg727Gln</td>
<td>p.Thr661Gln</td>
</tr>
</tbody>
</table>

<sup>1</sup>Amino-acid coordinates correspond to GenBank accession number NP_004981.2

<sup>2</sup>Amino-acid coordinates correspond to GenBank accession number NP_011780
**Figure 2.4. Study of MARS Variants in Yeast.** Three representative cultures of each yeast strain (indicated along the top of each panel) were inoculated and grown on solid growth medium containing 5-FOA (see Methods for details). Each strain was previously transfected with a vector containing no insert (pRS315), wild-type MES1 (wt MES1) or the indicated variant form of MES1. Two independently generated mutant-bearing constructs were analysed (DNA Set 1 and DNA Set 2). Before inoculating on 5-FOA-containing medium, each strain was resuspended in 100 µl water, then diluted 1:10.
This is somewhat further complicated by the late-onset of CMT2 in this family, which we believe contributes to the incomplete penetrance. A wide range of age-of-onset is common in CMT2 and can complicate segregation studies. With over 50 genes known in CMT only about 30% of CMT2 cases are explained as of today. It is to be expected that future genes are exceedingly rare causes of the disease and it will thus become difficult to fully verify their relevance. However, we believe that it will be valuable to carefully report such gene identifications as to provide the research field the opportunity to further support or disprove such claims in the future. Although our finding required a number of assumptions on variant filtering typical for exome sequencing studies, we are suggesting MARS as a novel rare CMT2 gene for the following reasons: (1) near complete evaluation of functionally strong variants in the studied family; (2) unusually strong conservation of the mutation across species; (3) four pre-existing ARS genes causing CMT phenotypes; and (4) a demonstrated loss-of-function effect of Arg618Cys MARS similar to many other CMT-associated ARS alleles. In summary, we are presenting data for a late-onset CMT2 family suggesting MARS as a novel, very rare cause of the disease. These results warrant further evaluation of the MARS locus for pathogenic mutations in patients with axonal CMT disease, and further underscore the importance of relevant functional evidence toward implicating rare variants in disease-onset in pedigrees too small for significant linkage analysis.
CHAPTER 3

Mutations in Phospholipase DDHD2 Cause Autosomal Recessive Hereditary Spastic Paraplegia

The data presented in this chapter were previously published in *The European Journal of Human Genetics* (Volume 21, Issue 11)\(^{24}\). I contributed to the identification of *DDHD2* and Rebecca Schule and I wrote the manuscript. I performed all of the work presented with one exception. Sheela Nampoothiri and Rebecca Schule performed all clinical examinations and patient sample collections.

**Perspectives**

The genetically diverse group of hereditary spastic paraplegias (HSP) is clinically defined by a progressive spasticity and weakness of the lower limbs, caused by distal axonopathy of the long motor axons of the corticospinal tract. Identification of more than 39 HSP genes highlights intracellular membrane trafficking, mitochondrial metabolism and myelin formation as key functions involved in HSP pathogenesis\(^3, 108\). Only recently lipid metabolism has emerged as another main theme in HSP pathophysiology, backed by the discovery of mutations in a number of genes involved in lipid metabolism and signaling including *CYP2U1, CYP7B1, DDHD1, FA2H, GBA2* and *PNPLA6*\(^{25, 26, 109-111}\). In this study, we have identified deleterious mutations in phospholipase *DDHD2*, the sister enzyme of PA-PLA1/DDHD1 in two families with complicated HSP.
Materials and Methods

Exome sequencing was performed in 79 index patients with familial forms of autosomal recessive HSP. The SureSelect Human All Exon 50 Mb kit (Agilent, Santa Clara, CA, USA) was used for in-solution enrichment; exome sequencing was performed using the Hiseq2000 instrument (Illumina, San Diego, CA, USA). Paired-end reads of 100 bp length were produced. BWA and GATK software packages were used to align sequence reads to the reference and call variant positions \cite{112, 113}. The data were then imported into GEM.app, a web-based database and analysis toolkit for next generation sequencing data for further analysis \cite{114}. An average of 82,613,347 sequence reads was produced per sample, 98.7% of which could be aligned to the targeted sequence. Mean coverage was 75.9-fold; 85.5% of the targeted sequence was covered by at least 10 reads. Variants were filtered for impact on the coding sequence, presence of either one homozygous or two heterozygous variants in one gene, frequency in public databases (minor allele frequency in dbSNP135 and NHLBI ESP6500 < 0.5%), conservation (GERP score > 2 or PhastCons score > 0.6) and genotyping quality (GATK quality > 30 and genotype quality GQ > 30). Additionally, variants segregating in more than two families in GEM.app were removed. In addition to 79 families with complex HSP, GEM.app contained ~130 families with pure HSP and ~450 families with non-HSP phenotypes at the time of analysis. Informed consent was obtained from all individuals and the Institutional Review Boards at the participating medical centers approved the study.
**Gene Identification**

In family THI26003, seven homozygous SNVs and one homozygous indel were present; two of the SNVs as well as the indel didn’t segregate with the disease. Segregating homozygous missense variants were found in *C14orf166* (NM_016039.2: c.311C>T, p.Pro104Leu; rs149288575), *FUT10* (NM_032664.3: c.473T>C, p.Leu158Pro), *IMPA1* (NM_001144879.1: c.542C>T, p.Thr181Ile) and *OTOGL* (NM_173591.3: c.3461A>G, p.Asp1154Gly); a homozygous nonsense variant in *DDHD2* (NM_001164234: c.859C>T, p.Arg287*) was the only truncating change present. Details on all the five segregating variants in family THI26003 are given in Table 3.1. As it was not possible to further narrow down the list of candidate genes in family THI26003 alone based on the available *in silico* parameters, the resulting list of five candidate genes was used as a seed and intersected with the resulting candidate gene lists obtained by exome sequencing in the remaining 78 autosomal recessive HSP families. Another deleterious *DDHD2* variant was identified in family IHG25194 (NM_001164232.1: c.1982_1983delAT, p.Tyr661Cysfs*8) that segregates with the disease (Figure 3.1). Parents of the two affected siblings of this family were not available for genetic analysis, therefore a genomic deletion of one *DDHD2* allele cannot be ruled out with certainty. Independent analysis of the exome data of family IHG25194 did not yield any nonsense or truncating variants other than the
Table 3.1. Five segregating variants in family THI26003 after filtering of sequencing data.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcript</th>
<th>Variant</th>
<th>Effect on aa sequence</th>
<th>MAF NHLBI ESP6500 (all)</th>
<th>dbSNP137</th>
<th>PhastCons score</th>
<th>GERP score</th>
<th>Polyphen-2 score v2.2</th>
<th>Mutationtaster prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>C14orf166</td>
<td>NM_016039.2</td>
<td>c.311C&gt;T</td>
<td>p.Pro104Leu</td>
<td>0.0847%</td>
<td>rs149288575</td>
<td>0.857</td>
<td>5.14</td>
<td>0.734</td>
<td>disease causing</td>
</tr>
<tr>
<td>FUT10</td>
<td>NM_032664.5</td>
<td>c.473T&gt;C</td>
<td>p.Leu158Pro</td>
<td>0.0077%</td>
<td>-</td>
<td>1.0</td>
<td>5.17</td>
<td>1.0</td>
<td>disease causing</td>
</tr>
<tr>
<td>IMPA1</td>
<td>NM_001144879.1</td>
<td>c.542C&gt;T</td>
<td>p.Thr181Ile</td>
<td>-</td>
<td>-</td>
<td>1.0</td>
<td>4.23</td>
<td>n.a.</td>
<td>polymorphism</td>
</tr>
<tr>
<td>OTOGL</td>
<td>NM_173591.3</td>
<td>c.3461A&gt;G</td>
<td>p.Asp1154Gly</td>
<td>0.116%</td>
<td>rs202085918</td>
<td>0.896</td>
<td>5.83</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>DDHD2</td>
<td>NM_001164234</td>
<td>c.859C&gt;T</td>
<td>p.Arg287*</td>
<td>-</td>
<td>-</td>
<td>1.0</td>
<td>2.53</td>
<td>0.72</td>
<td>disease causing</td>
</tr>
</tbody>
</table>
Figure 3.1. Pedigrees and sequence traces of the DDHD2 families. The c.859C4T mutation segregates in family THI26003 and leads to the formation of a preterminal stop signal at codon 287. The c.1982_1983delAT mutation segregates in family IHG25194; it results in a frameshift at amino acid position 661.
above-described *DDHD2* mutation. In none of the other four candidate genes additional variants were found in the remaining autosomal recessive HSP families. Both *DDHD2* mutations will destroy the integrity of the DDHD domain and thereby affect catalytic function, membrane localization, phosphoinositide binding and homo-oligomerization (Figure 3.2). Both mutations therefore likely lead to a complete loss of phospholipase *DDHD2* function.

**Clinical Description**

The two siblings of family THI26003 originating from Azerbaijan from a consanguineous family background (first degree cousins) developed a progressive spastic gait disorder since early childhood. At the time of examination (disease duration 22–13 years) they were still able to walk unsupported. Both patients had intellectual disability and had attended a special school for mentally disabled children; they were working in a sheltered workshop. Spastic paraplegia was further complicated by short stature, high arched palate and dysgenesis of the corpus callosum especially in the dorsal parts (Figure 3.3 A and B).

The two brothers of family IHG25194 were of Indian origin and had spastic paraplegia from early childhood. No consanguinity was reported, but both parents originated from the same village and were part of the Muslim community, suggesting a possible founder effect. Additional clinical signs and symptoms in
**Figure 3.2. Schematic of the DDHD2 gene.** The DDHD2 gene contains three known protein domains. The WWE domain is predicted to mediate protein interactions in ubiquitin and ADP ribose conjugation systems. The tandem SAM (sterile alpha motif domain) – DDHD domain is required for phosphoinositide binding.14 Integrity of the family defining DDHD domain, present in DDHD2 as well as its homolog DDHD1, is necessary for the PLA1 catalytic activity and homo-oligomerization of DDHD2. Catalytic function as well as a positively charged cluster in the SAM domain (Arg434-Lys435-Lys436) also required for phosphoinositide binding are necessary to promote membrane localization115-118. Mutations previously described are indicated in black, novel mutations in red132.
Figure 3.3. Cranial MRI/CT scans of DDHD2 patients. MRI/CT scans of DDHD2 patients show dysgenesis of the corpus callosum as well as some paucity of the periventricular white matter (d). (A): coronal MRI (T1 inversion recovery) of THI26003-4. (B): axial CT-scan of THI26003-3. (C): sagittal MRI (T1) of IHG25194-3. (D): axial MRI (T1) of IHG25194-3.
both the siblings included mental retardation, mild facial dysmorphism, short stature and dysgenesis of the corpus callosum (Figure 3.3 C and D).

Considering the usually high phenotypic variability in HSP, the phenotype between the two unrelated *DDHD2* families is astonishingly similar. Key features of *DDHD2*-related HSP appear to be spastic paraplegia, mental retardation, short stature and dysgenesis of the corpus callosum (Table 3.2).

Discussion

Phospholipids are a key component of biological membranes. They are metabolized by the large family of phospholipases that can be classified according to their site of cleavage. The phospholipase A1 family, consisting of extracellular and intracellular enzymes, hydrolyzes the ester bond at the sn-1 position of phospholipids, producing 2-acyl-lysophospholipids and fatty acids. In contrast to most eukaryotic organisms, mammals have three different intracellular phospholipase A1s: phosphatidic acid preferring phospholipase A1 (PA-PLA1/iPLA1a; DDHD1), the SEC23-interacting protein p125 (iPLA1b; SEC23IP) and phospholipase DDHD2 (iPLA1c; DDHD2). The biological function of phospholipase DDHD2 is not fully understood. DDHD2 is ubiquitously expressed; a cytosolic and a membrane-associated pool, localizing to the cis-Golgi and the ER-Golgi intermediate compartment (ERGIC) are in a dynamic equilibrium\(^{115-117}\). Overexpression of DDHD2 leads to dispersion of the Golgi and enlargement of the perinuclear ERGIC\(^ {116,118}\). Conflicting data exists about the
Table 3.2. Clinical features of genetically confirmed *DDHD2* patients

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Gender</th>
<th>Age at onset</th>
<th>Age at examination</th>
<th>Spasticity, UL/LL</th>
<th>Reflexes, UL/LL</th>
<th>Extensor plantar response</th>
<th>Bladder disturbance</th>
<th>Sensory deficits</th>
<th>Other signs(s)</th>
<th>Imaging</th>
<th>Neurophysiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH126003-3</td>
<td>M</td>
<td>3</td>
<td>25</td>
<td>-/+</td>
<td>+/+</td>
<td>+</td>
<td>-</td>
<td>None</td>
<td>Mental retardation short stature</td>
<td>TCC</td>
<td>NCV normal; SEP/MEP normal</td>
</tr>
<tr>
<td>TH126003-4</td>
<td>F</td>
<td>6</td>
<td>19</td>
<td>-/+</td>
<td>+/+</td>
<td>+</td>
<td>-</td>
<td>Vibration sense</td>
<td>Mental retardation sacral reflex short stature</td>
<td>TCC</td>
<td>NCV normal; SEP/MEP normal</td>
</tr>
<tr>
<td>IHG25194-3</td>
<td>M</td>
<td>0</td>
<td>2</td>
<td>-/+</td>
<td>+/+</td>
<td>+</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Mental retardation short stature</td>
<td>TCC</td>
<td>NCV normal; SEP/MEP normal</td>
</tr>
<tr>
<td>IHG25194-4</td>
<td>M</td>
<td>0</td>
<td>9</td>
<td>-/+</td>
<td>+/+</td>
<td>+</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Mental retardation short stature</td>
<td>TCC</td>
<td>Not done; SEP/MEP normal</td>
</tr>
</tbody>
</table>

Abbreviations: F, female; LL, lower limits; M, male; SEP, motor evoked potential; NCV, nerve conduction velocity; SEP, sensory evoked potential; TCC, thin corpus callosum; UL, upper limits; WMC, white matter changes.
effect of DDHD2 depletion. Morikawa et al. report a specific defect of retrograde transport from the Golgi to the ER; this defect is not confirmed by Sato et al. who instead postulate a anterograde transport defect from the Golgi to the plasma membrane\textsuperscript{115, 117}. The authors attribute these differences to their use of different oligonucleotides for the knockdown of DDHD2 in HeLa cells and possible off-target effects. In spite of these contradictions, the role of DDHD2 in intracellular membrane trafficking is further supported by its homology to another member of the PLA1 family – p125 – that has been shown to be involved in vesicular transport from the ER to the Golgi by interacting with Sec23p, a component of the COP II complex\textsuperscript{119}.

Changes in morphology and dynamics of Golgi and ER functionally connect several subtypes of HSP: Atlastin-1 (SPG3), the central nervous system expressed member of the atlastin family of GTPases, localizes predominantly to the tubular ER and to a lesser extent to the ERGIC and the cis-Golgi. It is required for the formation of three-way junctions of the ER\textsuperscript{120}. Receptor expression-enhancing proteins (REEPs) and reticulons form large oligomeric complexes in the ER membrane that are involved in curving the ER membrane, thereby forming the characteristic tubular structure of the smooth ER. Mutations in \textit{REEP1} (SPG31) as well as \textit{RTN2} (SPG12) cause autosomal dominant forms of HSP\textsuperscript{121}. Point mutations in \textit{BSCL2}, causing SPG17, lead to upregulation of ER stress markers\textsuperscript{122}. The long isoform of \textit{SPAST} mutated in the most common autosomal dominant subtype of HSP (SPG4) forms a protein complex with \textit{ATL1}
and REEP1 in the tubular ER network and is thought to coordinate microtubule regulation and membrane modeling. The identification of mutations in the gene encoding DDHD2, another protein acting at the ER-to-Golgi interface, therefore comes as no surprise.

In addition to its involvement in membrane trafficking, DDHD2 has been shown to possess phospholipase A1 catalytic activity towards phosphatidic acid and other phospholipids, a function shared also by DDHD1. Phosphatidic acid has diverse biological functions. It is a precursor for the biosynthesis of triacylglycerols and phosphoglycerols. Along with its structural functions in biological membranes, it is also involved in lipid signaling. This phospholipase A1 catalytic activity puts DDHD2 into context with a number of recently discovered HSP genes involved in lipid metabolism. Both seipin (SPG17) and spartin (SPG20) are involved in formation and regulation of lipid droplets. These dynamic organelles consist of a core of neutral lipids that are covered by a monolayer of amphiphatic lipids containing cholesterol and phospholipids. Phosphatidic acid, preferred substrate for DDHD1 and DDHD2, is essential for lipid droplet assembly. The acetyl-CoA transporter SLC33A1 located in the ER membrane and required for the formation of O-acetylated gangliosides is mutated in SPG42. Defects in the metabolism of complex lipids cause at least four more HSP subtypes: in SPG39 the deacetylation of phosphatidylcholine, the major membrane phospholipid, is defective due to mutations in phospholipase B/neuropathy target esterase (PNPLA6), mutations
in fatty acid-2 hydroxylase (FA2H) affect synthesis of 2-hydroxysphingolipids in SPG35 and CYP2U1 mutations in SPG49 lead to disturbed and w-1 fatty acid hydroxylation\textsuperscript{26, 111, 129}. GBA2 truncations, mutated in autosomal recessive HSP SPG46, lead to the deficiency of the conversion of glucosylceramide to free glucose and ceramide\textsuperscript{25}. In SPG5 mutations in the 7a-hydroxylase gene CYP7B1 not only lead to accumulation of oxysterols in SPG5 patients but may also influence neurosteroid metabolism\textsuperscript{130, 131}. During preparation of this manuscript Schuurs-Hoeijmakers et al. reported mutations in \textit{DDHD2} in four families with autosomal recessive HSP\textsuperscript{132}. The phenotype in these families is strikingly similar to the clinical features reported in families THI26003 and IHG25194 with early onset spasticity, mental retardation and TCC being the overlap between all affected family members. The seven \textit{DDHD2} mutations described in Schuurs-Hoeijmakers et al. and our study comprise four frameshift, two nonsense and only one missense mutation so far (Figure 3.1)\textsuperscript{132}. Of note, most mutations cluster in the DDHD domain, located in the C-terminal half of the protein. Further studies will have to clarify whether a toxic gain of function of potentially expressed truncated or mutant DDHD2 protein contributes to the phenotype. The nonsense mutation Arg287* that we identified in the Iranian family THI26003 has been described by Schuurs-Hoeijmakers et al. in another Iranian family; a founder effect is therefore possible\textsuperscript{132}.

The identification of mutations in \textit{DDHD2}, which is involved in Golgi-/ER membrane trafficking and lipid metabolism further demonstrates the critical roles
of these essential cellular processes in motor neuron function and helps to understand the molecular mechanisms underlying the pathogenesis of HSPs.
CHAPTER 4

GEnomes Management Application (GEM.app): A New Software Tool for Large-Scale Collaborative Genome Analysis

The data presented in this chapter were previously published in *Human Mutation* (Volume 34, Issue 6)\(^\text{24}\). I contributed to the development of the GEM.app framework (bioinformatics, GEDI, and software dealing with annotations) and wrote the manuscript. I performed all of the work presented with one exception: Rafael Acosta and Rick Ulloa assisted in the development of the graphical user interface.

**Perspectives**

The development of next-generation sequencing (NGS) technologies has revolutionized human genetics research\(^\text{133}\). Whole exome sequencing (WES), an early genomic application, is a rapid, high-throughput, and cost-effective approach and has been widely used to identify pathogenic variation especially in Mendelian disorders\(^\text{134}, \text{135}\). Such considerations have further stimulated the large-scale production of genomic datasets leading to significant challenges for efficient data management and analysis. The data-intensive nature and lengthy computational pipelines for genomic data have also increasingly removed clinically and molecular trained investigators from direct access to data analysis. This leads to miss opportunities in identifying novel causative gene variants and creates unnecessary bottlenecks in the discovery process. Finally, large disease
-oriented research consortia and collaborative networks of investigators are seeking new ways for communal data analysis and the sharing of variant data.

To address these concerns, a number of tools have been developed to analyze and visualize genomic variant data. Although these available tools are very useful, they are also limited in specific ways. VAAST is a powerful tool to identify genes likely involved in disease, but is not intended for the browsing of variant and annotation data. To visualize variant data, VARSIFTER was developed, but is designed for a desktop computer and can presently only manage a modest amount of data. Lastly, the sequence variant analyzer includes many powerful analysis packages, but is rather complex and does not easily facilitate collaborative efforts. To address these limitations, we have developed Genomes Management Application (GEM.app), which is an analysis toolset accessible via modern Web browsers allowing for easy, quick, and collaborative analysis of genomic data. GEM.app is currently growing significantly in size and we are identifying disease genes at the pace of >1 per month.

Bioinformatics

The Illumina CASAVA v1.8 pipeline was used to produce 100 bp sequence reads. BWA software was used to align sequence reads to the human genome (hg19) and variants were called using the GATK v1.4 software package. Variants were submitted to SeattleSeq for annotation. Further
annotation was obtained using data from dbSNP137, variant frequency data from the NHLBI Exome Sequencing Project (Exome Variant Server, NHLBI Exome Sequencing Project [ESP], Seattle, WA Project [Exome Variant Server, 2012]), the HGMD Human Gene Mutation Database, and the OMIM Online Mendelian Inheritance in Man database (OMIM Online Mendelian Inheritance in Man, December, 2012).

**GEDI Pipeline and GEM.app Graphical User Interface**

GEnome Data Import (GEDI) is an automated pipeline that uses different available software tools (VCFtools and ENSEMBL VEP) and python scripts to automate processing, annotation, backfilling of variant call files (VCF) files and data upload into our GEM.app database (mySQL 5.5). GEDI is optimized to extract relevant information for each variant and to transform and structure the genomic information to guarantee fast query execution using the GEM.app graphical user interface (GUI). The GEM.app GUI was implemented in layers to allow enough flexibility to handle data efficiently in a fast growing environment. To execute queries, API’s were developed using PHP 5.3. GEM.app’s user-friendly interface was built with JQUERY, HTML5, JSON, and CCS3. Slickgrid was used to display query results.

**Basic Principles and Description of GEM.app**

Several design principles went into the development of GEM.app. We created a graphical user interface for this Web application that makes genomic
data accessible for physician scientists and molecular trained PhDs (Figure 4.1). The easy to use interface allows a user to design custom queries and results are typically returned in seconds. This is a key feature as it allows for an iterative working approach, instant refining of filters, and immediate testing of different Mendelian segregation patterns—even if hundreds of exomes are being queried. GEM.app is a web application developed using the latest Internet standards and languages—JQuery, JSON, HTML5, SlickGrid and is compatible with Safari 5.1.7, Chrome 22.0.1229.94, and Firefox 16.0.2 or later versions. We have developed a powerful access-control system that assigns an account to each user. This account system is context-aware and shows or hides phenotype-specific customized filter-options and variant annotation. Each user only sees their own detailed data; yet, visible to all users are anonymous variant counts derived from the entire database. Examples include global minor allele and genotype frequencies in GEM.app, number of familial segregation events of a specific variant under a selected Mendelian trait, or number of SNVs/indels per gene of interest. These latter features encourage collaboration for studies that are focused on rare Mendelian-type variants, as they give hints on the existence of a second family for a given new candidate gene. The access system further allows for sharing of access to specific exomes and a collaborative analysis.
Figure 4.1. GEM.app graphical user interface (GUI). GEM.app is accessible via any web browser. The user-friendly GUI allows users from all computational backgrounds to analyze NGS data.
Examples of existing successful collaborations in GEM.app include the international Inherited Neuropathy Consortium with >250 exomes in GEM.app and a network of 15 collaborating groups in as many countries working on hereditary spastic paraplegia (HSP) (>400 exomes).

Data Security

We only store deidentified data in GEM.app. This includes a numerical identifier for each sample and family, sex, and deidentified pedigrees. Access is password controlled and a VeriSign class 3-server certificate encrypts all data transfer between users and servers, which is comparable to online banking security. Further, GEM.app currently resides on servers of the University of Miami, which are behind a firewall and monitored 24/7 for cyber attacks. In the future, we envision moving GEM.app to a true cloud environment.

GEDI Pipeline

Part of GEM.app is a processing pipeline, GEDI module, which handles processing of VCF files, annotation, “backfilling” of variant data, cosegregation analysis within families, and calculation of counts across all samples, including minor allele frequencies. When data are processed and imported into GEM.app, information on each sample is required such as: affection status, pedigree individual ID, family ID, and possible Mendelian inheritance patterns. Using this information for each sample allows GEDI to automatically determine whether a
specific variant follows a given segregation pattern (i.e., autosomal dominant). If a user selects a particular inheritance pattern to analyze, GEM.app will return variants that fit this model in a table format (Figure 4.2). GEDI achieves annotation by utilizing the SeattleSeq annotation server (http://snp.gs.washington.edu/), which includes conservation and amino acid substitution scores (GERP, PhastCons, Grantham, PolyPhen2)\(^{141-143}\). In addition, GEDI obtains data from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), NHLBI EVS (http://evs.gs.washington.edu/EVS/), OMIM (http://www.ncbi.nlm.nih.gov/omim), String-db (http://string-db.org/), and ENSEMBL (http://www.ensembl.org/).

Backfilling is a process that retrieves sequence information from all previously added VCF files for each novel variant incorporated into GEM.app. Each addition of new exomes initiates a reanalysis of variant counts, recalculation of allele and genotype frequencies, etc. The GEDI process is fully automated and takes advantage of the 5,000-node computer cluster named Pegasus at the University of Miami.

**Graphical User Interface**

After the GEDI/GEM.app pipeline is completed, data are accessible to registered users through an online graphical user interface (https://genomics.med.miami.edu). Users interact with query modules, which are presented as tiles with descriptive names. Currently seven different modules are available. These include simple modules (“Quick finds”) that search for genes
Figure 4.2. Example of output from a GEM.app analysis query. Results are returned directly on the client side web browser. We provide many helpful linkouts to directly link variants to databases such as UCSC genome browser, NCBI, OMIM, ClinVar, and pubmed.
(“Gene look-up”) or genomic positions (“Position look-up”). Further, all accessible exomes are individually listed (“My samples”) complete with basic phenotypes, external and internal numbering, quality measures from alignment and variant calling, possible traits, processing details, and other information. Most queries happen in extended filtering modules. “Variants within families” provides detailed options for variant filtering within single families, but can process as many families at once as requested. Fifteen different filter option fields are present, including selection of genomic positions, variant function class (synonymous, nonsynonymous, etc.), conservations scores, quality scores, Mendelian traits, or lists of known genes for phenotypic groups (i.e., inherited peripheral neuropathy genes). The module “Genes across families” allows for identifying genes that have multiple hits in the same gene across multiple families and/or phenotypes. Advanced filter modules contain two-step or nested queries. This allows for filtering with a strict set of criteria resulting in few hits in a first step and then in a second step these gene are taken into account for a query with more relax criteria to find additional evidence for a gene in a larger set of exomes. All filter modules have the option of choosing from four preset filter criteria: user-defined, relaxed, moderate, or strict (Figure 4.1).

The output is presented in a table that contains 28 different columns with annotation. Each row presents a variant. There are an additional 44 columns of annotation that can be added via a “column manager” (Figure 4.3). Columns can also be sorted and rearranged via drag and drop. Several fields in the output
Figure 4.3. GEM.app Annotation Manager. The column manager allows for adding up to an additional 44 columns, deleting columns or resorting of columns.
table contain hyperlinks that directly link to pedigrees, to the UCSC genome browser (http://genome.ucsc.edu/), a gene-network viewer (http://string-db.org), NCBI (http://www.ncbi.nlm.nih.gov/), or OMIM (http://omim.org/) (Figure 4.2). GEM.app connects several publicly available databases directly to data being analyzed within the same Web page, thus making follow-up analysis less tedious (Figure 4.2).

**Performance**

Queries of individual families are typically finished in less than 1 sec. Benchmarking of more than 10,000 queries of over 60 different users demonstrate that all the different query modules achieve an average query time of less than 8 sec, with the exception of the two-step module (Figure 4.4). The most popular query module, “Variants within families,” produces output in about 4 sec. Factors that negatively influence query times include: (1) querying across multiple phenotypes, (2) using fewer filter criteria, and (3) using nested filter options. Generally, the speed and simplicity of GEM.app allows an investigator to iteratively test different filtering strategies (such as multiple possible traits) within a few minutes without the need of programming. Searching across a large number of samples/families is fast: a query for conserved and rare mutations in 124 known genes for related neurodegenerative diseases (CMT, HSP, distal hereditary motor neuron, etc.) across 481 exomes obtained results within 10 sec. Querying these same 124 genes across 1,200 samples finished within 15 sec, which suggests this is a scalable platform.
Figure 4.4. Performance of GEM.app. Average search times over 10,000 queries from over 60 different users. By far, the most popular module is “Variants within families,” which returns results in \(~4\) sec across 1,200 exomes. Individual families are typically instantly returned.
Currently, 103 users are registered to use GEM.app, which includes 40 principal investigators from 15 different countries studying over 50 different phenotypes (Figure 4.5 A). Since the release of GEM.app in April 2012, we have experienced a significant increase in the number of users and queries (Figure 4.5 B and C).

**Discussion**

The development of GEM.app was motivated by the need of NGS analysis tools for physician scientists and biomedical investigators with limited computational experience. Importantly, we needed a tool to manage and organize large sets of exome data as they become available from large-scale sequencing projects. Increasingly large sample sizes are required to identify genes for rare, highly heterogeneous Mendelian disorders and rare familial forms of phenotypes with a complex genetic architecture. The modular structure of GEM.app allows for future implementation of new computational strategies, such as multicore processing and distributed systems, to address the increasing size of data and scaling to whole genomes. Because investigators worldwide are producing small to large exome-genome datasets, the exchange of variant information and uniform analysis provides an often-untapped opportunity for increasing genetic power via collaborations. A variety of collaborative models from loose networks to tightly integrated consortia are feasible within the flexible access system in GEM.app. Existing projects range from the ability of directly
Figure 4.5. Current usage of GEM.app. A: Geographical overview of principle investigators with data in GEM.app. B: The number of registered users has grown to >103 in 2013. C: The usage of GEM.app has increased significantly since its release. ASHG—American Society for Human Genetics annual meeting.
shared access to full exomes at different sites to using the anonymous variant
counts across all datasets, which are available to every registered user. Further,
consortia can decide to limit data analysis to specific sites, or, potentially more
advantageous, data analysis can be spread over a larger number of institutions.
The intuitive interface allows for the participation of investigators with different
skill sets, including expert geneticists, physician scientists, or basic scientists.

The GEM.app framework has recently been utilized to identify clinically
relevant variants in a number of disorders, such as inherited deafness, CMT,
HSP, and dilated cardiomyopathies 16, 144-147. GEM.app has also been applied to
identify novel genes for a variety of disorders 16, 25, 26, 135, 148, 149. The majority of
these studies are led by physician scientists interested in the application of whole
exome sequencing for the elucidation of genetic variation involved in their
disorders of interest, thus demonstrating the power of GEM.app to connect
investigators to their NGS datasets. Some of the newer findings have only been
possible by connecting relatively small datasets from around the world in this
centralized resource and the ability of individual investigators to share their
results on a trusted platform. In summary, GEM.app will enable researchers of all
computational backgrounds to visualize and analyze genomic variant data. Using
an automated pipeline, GEM.app organizes thoroughly annotated variant data to
be directly connected to various genomics resources and allows investigators to
directly analyze and interpret mutations from large sets of samples. GEM.app
offers the ability for biomedical researchers to share data and perform joint
analysis simultaneously. These features promote collaborations leading to the identification of novel disease associated genes.
A Novel Mutation in VCP Causes Charcot-Marie-Tooth Type 2 Disease

The data presented in this chapter were previously published in *Brain* (Volume 137, Issue 1)\(^8^2\). I contributed to the identification of VCP and wrote the manuscript. I performed all of the work presented with four exceptions. First, Michael Shy performed all clinical examinations and patient sample collections. Second, Susan Blanton performed the linkage analysis. Third, the lab of Christopher Wiehl performed all VCP enzymatic assays. Finally, Matthew Danzi and Alleene Strickland assisted in the network view of VCP.

**Perspectives**

Charcot–Marie–Tooth disease type 2 (CMT2) is a clinically and genetically heterogeneous disorder that leads to axonal degeneration of the peripheral nerve. At the time of writing, it is estimated that up to 70% of patients with CMT2 do not obtain a genetic diagnosis\(^7^7,^{10^2}\). However, whole exome sequencing has led to rapid progress in gene identification for CMT2, which will ultimately improve genetic diagnostics\(^4^0\). In addition, the identification of >50 genes for CMT2 has improved our understanding of the phenotypic and genotypic intersection of peripheral neuropathies with other motor neuron and axonal degenerative disorders. The gene encoding valosin-containing protein (VCP) is an example of a gene associated with a spectrum of related disorders due to
reports that have linked missense mutations in VCP to cause inclusion body myopathy, Paget’s disease of the bone, frontotemporal dementia (IBMPFD), hereditary spastic paraplegia, and 1–2% of familial amyotrophic lateral sclerosis (ALS)\(^81,150-152\). VCP, in association with ubiquitin binding proteins, facilitates the degradation of proteins through both the proteasomal and autophagic pathways\(^153\). Recent evidence suggests that pathogenic VCP mutations preferentially disrupt autophagy leading to cellular degeneration\(^154\). Using whole exome sequencing in a family with autosomal dominant CMT2, we identified a novel missense mutation (p.Glu185Lys) in VCP that is located in the same L1 domain as other pathogenic mutations. Our results suggest that the phenotypic spectrum associated with VCP should be expanded to include CMT2.

**Consent and Exome Sequencing**

A CMT2 family with eight individuals was ascertained. Informed consent was obtained from all individuals and the Institutional Review Boards at the participating medical centers approved the study. Neurological exam and neurophysiological studies were performed.

**Clinical Findings**

We evaluated eight family members from a single generation of a family thought to have a dominantly inherited peripheral neuropathy (Figure 5.1).
Figure 5.1. Pedigree of Family with VCP Variant. Pedigree of the studied family with genotypes of the VCP mutation in all available individuals. Black symbols – affecteds.
Dominant inheritance is presumed based on family history describing that the identical twins Subjects 1001 and 1002 were reported to have always had long, thin legs, high arches and hammertoes and developed weakness and sensory loss in their fourth and fifth decade of life. Subject 1001 passed away at 87 years of age and Subject 1002 passed away at 84 years, both of natural causes. Five individuals had abnormalities on their neurological examinations and nerve conduction studies. The onset of symptoms began in early childhood for Subject 0101, early adulthood in Subject 0108, and after age 50 in Subjects 0102, 0103 and 0104. The earliest affected individual required shoes connected by a bar to aid in ambulation in early childhood. She was a slow runner and could not keep up with her peers as a girl. She underwent triple arthrodesis in 1998 and 2000. She continued to have difficulty with balance and noticed a progressive inability to feel touch below her knees. She began wearing bilateral ankle foot orthoses in 2009, and in 2010 she used these along with forearm crutches for mobility. She subsequently developed difficulty using her hands for fine movements such as manipulating buttons or fastening jewelry. In 2009, neuropsychiatric testing and evaluation resulted in the diagnosis of adjustment disorder with changes in mood and behavior requiring anxiolytic antidepressant or other psychotropic medication. She has long-standing dysarthria, dyspnea, and persistent cough. Subject 0108 first noted weakness in her early 20s when she needed to wear high heels at work. She obtained orthotics in her late 20s and has gradually developed problems with balance. The three remaining subjects all developed problems with balance, loss of sensation, and difficulties with fine movements of
their hands after the age of 50. The adult CMT Neuropathy Score (CMTNSv2) was used to measure impairment in all five subjects. Impairment was considered to be mild in two, moderate in two and severe in one individual. All affected subjects had absent Achilles deep tendon reflexes. Four of the five had clear pes cavus; the fifth had high arches but not hammer toes. The clinical features of the affected five individuals are summarized in Table 5.1. Four of the five affected patients examined gradually developed symmetrical, length-dependent weakness resulting in an inability to walk on their heels and frequent ankle pains. The same four patients subsequently developed weakness in intrinsic hand muscles (Table 5.1). Notably, weakness in the proximal leg or arm muscles was not observed. No scapular winging was present, creatine kinase values were normal, and no fasciculations were observed. Atrophy was detected in weak calf and intrinsic hand muscles. All five patients had length-dependent small (pinprick) and large (vibration) fiber sensory loss (Table 5.1). All five patients had difficulty with balance that was exacerbated by situations in which vision was impaired as by darkness or walking in a crowd. Nerve conduction studies were abnormal in all five affected individuals. The abnormalities were characterized as intermediate slowing of motor conduction velocities and axonal features of both motor and sensory nerves (Table 5.2 and 5.3). The EMG did not demonstrate any spontaneous activity as positive sharp waves or fibrillations. The proband (Subject 0101) had previously undergone commercial testing for CMT that did not reveal mutations in GJB1 (Cx32), MPZ,
Table 5.1. Clinical features of family with E187K VCP mutation

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Patient age (years)</th>
<th>CMTNSv2</th>
<th>Distal weakness LL</th>
<th>Proximal weakness LL</th>
<th>Distal weakness UL</th>
<th>Proximal weakness UL</th>
<th>Vibration LL</th>
<th>Vibration UL</th>
<th>Cutaneous LL</th>
<th>Cutaneous UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0101</td>
<td>59</td>
<td>22 (severe)</td>
<td>+ (4–4.5)</td>
<td>– (5.5)</td>
<td>+ (4–4.4–)</td>
<td>– (5,5,5)</td>
<td>Red Toes,</td>
<td>Red fingers,</td>
<td>Abs toes,</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wrists,</td>
<td>wrist</td>
<td>ankles, knees</td>
<td></td>
</tr>
<tr>
<td>0102</td>
<td>64</td>
<td>12 (moderate)</td>
<td>+ (4–4.5)</td>
<td>– (5.5)</td>
<td>+ (4,5,4)</td>
<td>– (5,5,5)</td>
<td>Normal</td>
<td>Normal</td>
<td>Abs toes,</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Red fingers,</td>
<td>Red fingers,</td>
<td>Red ankles</td>
<td>Normal</td>
</tr>
<tr>
<td>0103</td>
<td>66</td>
<td>16 (moderate)</td>
<td>– (5,5)</td>
<td>– (5.5)</td>
<td>– (5,5,5)</td>
<td>– (5,5,5)</td>
<td>Abs toes,</td>
<td>Abs toes,</td>
<td>Abs toes,</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Red ankles,</td>
<td>Red ankles,</td>
<td>Red ankles,</td>
<td>Normal</td>
</tr>
<tr>
<td>0108</td>
<td>48</td>
<td>5 (mild)</td>
<td>+ (4,5)</td>
<td>– (5.5)</td>
<td>+ (4,4,4,5)</td>
<td>– (5,5,5)</td>
<td>Abs toes,</td>
<td>Abs toes,</td>
<td>Abs toes,</td>
<td>Normal</td>
</tr>
<tr>
<td>0104</td>
<td>66</td>
<td>7 (mild)</td>
<td>+ (0,3)</td>
<td>+ (4,4–)</td>
<td>+ (4,5,4)</td>
<td>– (5,5,5)</td>
<td>Abs toes,</td>
<td>Abs toes,</td>
<td>Red toes</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Motor weakness based on MRC scale (0–5): ‘+’ = weakness present, ‘−’ = no weakness detected. Lower limb distal weakness assessed by anterior tibialis and gastrocnemius; lower limb proximal weakness assessed by iliopsoas and quadriceps; upper limb distal weakness assessed by first dorsal interosseous, abductor pollicis brevis, and adductor digiti minimi; upper limb distal weakness assessed by deltoids, biceps brachii, and triceps. Vibration based on Rydell tuning fork with ‘5’ on scale of ‘8’ being considered normal and cutaneous based on Pinprick sensation: Normal is no decrease compared to the examiner. Red = reduced; abs = absent up to level indicated. Both motor and sensory evaluations were based on worst score observed of the two limbs. CMTNSv2 scores are separable into 510 (mild), 11–20 (moderate) or 420 (severe) impairment. LL = lower limb; UL = upper limb.
Table 5.2. Motor Nerve Conduction Velocities

<table>
<thead>
<tr>
<th>Indiv.</th>
<th>Age</th>
<th>Side</th>
<th>Ulnar Nerve</th>
<th>Median Nerve</th>
<th>Peroneal Nerve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>DML (ms)</td>
<td>NCV1 (m/s)</td>
<td>NCV2 (m/s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;3.4</td>
<td>&gt;49</td>
<td>&gt;50</td>
</tr>
<tr>
<td>0101</td>
<td>59</td>
<td>L</td>
<td>2.8</td>
<td>51</td>
<td>58</td>
</tr>
<tr>
<td>0102</td>
<td>64</td>
<td>L</td>
<td>3.6</td>
<td>47</td>
<td>46</td>
</tr>
<tr>
<td>0103</td>
<td>57</td>
<td>L</td>
<td>4.3</td>
<td>48</td>
<td>44</td>
</tr>
<tr>
<td>0108</td>
<td>48</td>
<td>L</td>
<td>2.7</td>
<td>49</td>
<td>43</td>
</tr>
<tr>
<td>0104</td>
<td>66</td>
<td>L</td>
<td>3.2</td>
<td>43</td>
<td>41</td>
</tr>
</tbody>
</table>

Normal values for each nerve in top row under DML, NCV and CMAP. DML – distal motor latency, NCV – Nerve Conduction Velocity; NCV1 – NCV between wrist/elbow; NCV2 – NCV around elbow; NCV3 – NCV between ankle/knee; NCV4 – NCV around knee. CMAP – compound muscle action potential; **Bold letters signify abnormal values.**
Table 5.3. Sensory Nerve Conduction Velocities

<table>
<thead>
<tr>
<th>Individual</th>
<th>Side</th>
<th>Median Nerve</th>
<th>Ulnar Nerve</th>
<th>Radial Nerve</th>
<th>Sural Nerve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lat. (ms)</td>
<td>NCV (m/s)</td>
<td>SNAP (µV)</td>
<td>Lat. (ms)</td>
</tr>
<tr>
<td>0101</td>
<td>L</td>
<td>&lt;3.5</td>
<td>&gt;44</td>
<td>&gt;20</td>
<td>&lt;3.1</td>
</tr>
<tr>
<td>0102</td>
<td>L</td>
<td>3.1</td>
<td>45</td>
<td>14</td>
<td>2.8</td>
</tr>
<tr>
<td>0103</td>
<td>L</td>
<td>3.3</td>
<td>42</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>0108</td>
<td>L</td>
<td>3.4</td>
<td>41</td>
<td>15</td>
<td>2.6</td>
</tr>
<tr>
<td>0104</td>
<td>L</td>
<td>3.0</td>
<td>47</td>
<td>13</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Normal values listed for each nerve in top row under Lat., NCV, and SNAP., Lat. - latency, NCV – Nerve Conduction Velocity; SNAP-Sensory Nerve Amplitude Potential. **Bold letters signify abnormal values.**
NEFL, GDAP1, MFN2, LMNA, RAB7A, GARS and HSPB1. Alkaline phosphatase and creatine kinase levels were normal.

Genomic Studies

We used the GEM.app software to focus our genomic analysis on rare and conserved variants that followed an autosomal dominant mode of inheritance in whole exome sequencing data \(^{24}\). Only variants were considered that meet the following criteria: (i) non-synonymous variant; (ii) not present in NHLBI EVS 6500; (iii) not present in 2,700 families within the GEM.app database; (iv) Genomic Evolutionary Rate Profiling (GERP) > 3 OR PhastCons > 0.6; and (v) Genotype Quality > 75 (as previously described in \(^{146}\)). This resulted in the identification of three candidate variants of which only a variant in VCP co-segregated with disease. This missense change, c.553C>T (p.Glu185Lys) in VCP (NM_007126.3) was not observed in 9,200 exomes, had a very high GERP conservation score of 6.07, PhastCons score of 1, and was predicted to be ‘damaging’ by four of five protein prediction programs (PolyPhen-2, MutationAssessor, LRT, MutationTaster, SIFT) (Table 5.4). VCP had previously been shown to cause a spectrum of phenotypes from ALS, frontotemporal dementia, inclusion body myopathy, Paget’s disease of bone, and hereditary spastic paraplegia, but not CMT2.

Parametric linkage analysis was performed with the following assumptions: (i) 100% penetrance; (ii) autosomal dominant mode of inheritance;
Table 5.4. *in silico* scores

<table>
<thead>
<tr>
<th>Severity prediction</th>
<th>Value</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>GERP score</td>
<td>6.07</td>
<td>–12.3 to 6.17</td>
</tr>
<tr>
<td>PhastCons</td>
<td>1</td>
<td>0 to 1</td>
</tr>
<tr>
<td>PolyPhen-2</td>
<td>Probably damaging</td>
<td>Benign, possibly damaging, probably damaging</td>
</tr>
<tr>
<td>MutationTaster</td>
<td>Disease causing</td>
<td>Polymorphism to disease causing</td>
</tr>
<tr>
<td>MutationAssessor</td>
<td>High</td>
<td>Neutral, low, medium, high</td>
</tr>
<tr>
<td>LRT</td>
<td>Deleterious</td>
<td>Unknown, neutral, deleterious</td>
</tr>
<tr>
<td>SIFT</td>
<td>Tolerated</td>
<td>Tolerated to damaging</td>
</tr>
</tbody>
</table>

Nearly all *in silico* scores predicted a significantly damaging mutation.

GERP = Genomic Evolutionary Rate Profiling.
(iii) disease frequency of 0.000001; and (iv) minor allele frequency of 0.000001. This analysis resulted in a log of odds (LOD) score of 2.107 for the c.553C>T VCP mutation in this family.

To understand the role of VCP in a gene network composed of known Mendelian genes for ALS, hereditary spastic paraplegia and CMT, we analyzed unbiased, experimentally derived protein–protein interactions based on data from BioGRID. VCP directly interacts with 14 known CMT, hereditary spastic paraplegia, and ALS proteins, and further interacts with 19 'intermediate' proteins that directly connect to proteins in the known disease gene clusters (Figure 5.2). Thus, VCP appears to represent an important node negotiating between these different disease gene clusters, which may explain its involvement in a number of phenotypes.

**Functional Evaluation**

The E185K variant resides within the linker domain between the N-domain and the D1 ATPase domain. Several other known pathogenic mutations are within this amino acid stretch and include R191Q, R191G, and L198W (Figure 5.3 A). More importantly, the E185 mutant residue is adjacent to other pathogenic mutations when superimposed onto the crystal structure of VCP (Figure 5.3 B and C).
Figure 5.2. VCP Gene Network. Visualization of VCP yeast 2-hybrid interactions with known CMT, hereditary spastic paraplegia and ALS proteins, allowing for one intermediate interacting node. Direct VCP interactions are highlighted. The figure was generated using interaction data from BioGRID and the GEM.app application.
Figure 5.3. VCP gene schematic and protein model. (A) Linear diagram of VCP with the location of 29 identified missense mutations at 18 different residues (in red). The location of the E185K mutation is shown in blue. (B–C) Renderings of the crystallographic structure of three subunits from a VCP hexamer (each individual monomer is a unique colour). (B) Hexameric VCP with the positions of the D1 domain (top barrel) and D2 domain (bottom barrel) with the N and C domains protruding to the sides. The dotted oval denotes where all 18 mutant residues (blue) reside. (C) Close-up view of the N-D1 interface where all known mutations reside (in blue) and the position of E185 (white).
To further assess the potential pathogenicity of the E185K mutation, we compared the intrinsic ATPase activity of VCP-WT, the previously reported IBMPFD mutations VCP-R155H and VCP-A232E with VCP-E185K. As demonstrated previously, the intrinsic ATPase activity of recombinant VCP-R155H and VCP-A232E was elevated by 3–4-fold (Figure 5.4 A)\(^{157}\). In contrast, VCP-E185K’s ATPase activity was unchanged as compared to VCP-WT (Figure 5.4 A). To further assess the pathogenicity, we transiently transfected VCP-WT, VCP-A232E and VCP-E185K fused to a C-terminal green fluorescent protein tag into U20S cells and evaluated the levels of VCP, the autophagosome protein LC3, SQSTM1 and a loading control actin via immunoblot (Figure 5.4 B). As previously described, VCP-A232E expression results in an increase in LC3II protein levels (Figure 5.4 B)\(^{154}\). Similarly, VCP-E185K expression caused an increase in LC3II levels consistent with the previously defined disruption in autophagosome maturation by VCP disease mutations (Figure 5.4 B and C).

**Discussion**

By applying a comprehensive genomic approach and linkage analysis we have identified a CMT2 family with a significant mutation in VCP. The combination of length-dependent sensory loss clinically and by electrophysiology, intermediate slowing of motor nerve conduction suggestive of myelin abnormalities and length-dependent weakness represents a novel phenotype for patients with VCP mutations. The phenotype is distinct from ALS or inclusion body myopathy. Because of the intermediate slowing on motor nerve
Figure 5.4. Functional evaluation of the VCP mutation. (A) Recombinant human VCP protein was purified from E.coli and its basal rate of ATPase was assayed. The ATPase activity of VCP-WT was arbitrarily designated as 1. (B) Representative immunoblot for VCP, LC3, SQSTM1 and actin of U20S cells transiently expressing VCP-WT-GFP, VCP-A232E-GFP or VCP-E185K-GFP for 2 days. (C) Quantitation of LC3II/actin levels from three independent experiments. *P<0.05 and **P-value<0.005 by student t-test.
conductions, this family could also be considered to have dominant-intermediate CMT rather than CMT2. The expansion of the phenotypic spectrum of VCP to include CMT is part of a broader trend, especially in neurological diseases, where genotypic and phenotypic overlap is increasingly recognized. This is mainly owed to the wider application of next-generation sequencing and other comprehensive forms of genome studies.

The E185K mutation in VCP is unique in that it causes distinctly one phenotype, CMT2, within a family. Most other previously reported VCP mutations have caused a spectrum of disease phenotypes ranging from inclusion body myopathy, Paget's disease, and FTD to ALS. The exception being patients identified with sporadic ALS and unique VCP variants. A CMT2 phenotype may be considered a milder form of IBMPFD/ALS as some patients with IBMPFD/ALS have been reported to have sensory/motor neuropathies in addition to the other phenotypes. This milder phenotypic expression may explain the normal ATPase activity of the E185K mutation. An elevated basal ATPase activity is characteristic for most but not all IBMPFD/ALS.

Studies suggest that IBMPFD mutations in VCP lead to an impairment in protein degradation and specifically autophagy. Expression of IBMPFD mutations in tissue culture or animal models leads to the accumulation of non-degradative autophagosomes. More recently it has been demonstrated that VCP is instrumental in the autophagic degradation of mitochondria or mitophagy.
VCP facilitates the extraction of ubiquitinated MFN1 and MFN2 from the outer mitochondrial membrane thus initiating mitophagy \(^{162}\). Interestingly, mutations in \textit{MFN2} lead to CMT2 suggesting a functional connection in the pathogenesis of these two hereditary neuropathies \(^{44}\). Our protein network analysis further underlines the role of \textit{VCP} as a hub with surprisingly many connections into CMT, ALS and hereditary spastic paraplegia gene/protein clusters. Future studies may determine that mutations at E185K specifically disrupt mitophagy as opposed to global macroautophagy.
CHAPTER 6

PNPLA6 Mutations Cause Boucher-Neuhäuser and Gordon Holmes Syndromes as Part of a Broad Neurodegenerative Spectrum

The data presented in this chapter were previously published in *Brain* (Volume 137, Issue 1)\(^{23}\). I contributed to the identification of PNPLA6 and wrote the manuscript along with Matthis Synofzik. I performed all of the work presented with three exceptions. First, Rebecca Schule, Charles Lourenco, Marcos Lima-Martinez, Dider Hannequin, Matthis Synofzik, and Wilson Marques Wilson Jr. performed all clinical examinations and patient sample collections. Second, the laboratory of Amjad Farooq performed all of the protein modeling. Finally, Adriana Rebelo and Marie Coutelier assisted in Sanger sequencing experiments.

**Perspectives**

Autosomal recessive cerebellar ataxias (ARCAs) are a clinically and genetically heterogeneous group of spinocerebellar diseases, often associated with additional non-cerebellar features\(^ {163,164}\). Two clinically defined syndromes combine early-onset ARCA with hypogonadotropic hypogonadism: (i) Boucher-Neuhäuser syndrome (MIM 215470), which is additionally associated with chorioretinal dystrophy\(^ {165-167}\); and (ii) Gordon Holmes syndrome (MIM 212840), which is additionally associated with brisk reflexes\(^ {168}\). Despite various descriptions of familial occurrence, the genetic basis of these syndromes has remained elusive. Moreover, it has remained unclear whether these syndromes
present clinically and genetically distinct entities or, alternatively, phenotypic clusters on a phenotypic continuum of neurodegenerative diseases caused by mutations in the same gene.

Here we demonstrate that Boucher-Neuhäuser and Gordon Holmes syndromes are indeed allelic diseases and reveal the major disease gene for these clinical disease entities. By using the significant genetic variants identified in Boucher-Neuhäuser and Gordon Holmes syndrome families as a seed we also analysed 4,500 exomes from patients with hereditary ataxia and/or spasticity syndromes and establish that variants in the new Boucher-Neuhäuser syndrome/ Gordon Holmes syndrome disease gene are not specific to these particular hypogonadism syndromes, but rather cause these presentations as part of a wider spectrum of neurodegenerative disease.

Whole Exome Sequencing of Index Patients with Boucher-Neuhäuser Syndrome and Gordon Holmes Syndrome

Whole exome sequencing was performed on two index patients with familial Boucher-Neuhäuser syndrome and one index patient with sporadic Gordon Holmes syndrome. Informed consent was obtained from all individuals and the Institutional Review Boards of the participating medical centers approved the study. The SureSelect Human All Exon 50 Mb kit (Agilent) was used for in-solution enrichment and exome sequencing was performed using the HiSeq2000 instrument (Illumina). Paired-end reads of 100-bp length were produced. BWA
and GATK software packages were used to align sequence reads to the reference and call variant positions, respectively \(^{112, 139}\). All data were then annotated and imported into GEnomes Management Application (GEM.app), a web-based tool for next generation sequencing data analysis \(^{24}\). An average of 73,609,687 sequence reads was produced per sample, 98.8% of which could be aligned to the targeted sequence. Mean coverage was 69-fold; 71% of the targeted sequence was covered by at least 20 reads.

Assuming a common cause for Boucher-Neuhäuser and Gordon Holmes syndromes, we used the GEM.app analysis module ‘Genes Across Families’ to filter for non-synonymous homozygous or compound heterozygous variants, with low frequency in public databases (minor allele frequency in dbSNP137 and NHLBI ESP6500 < 0.5%), moderate conservation (GERP score > 2 OR PhastCons score > 0.6) and moderate genotype quality (GATK quality index > 30 and genotype quality GQ > 30) in genes shared across the three families. Only one gene, \(PNPLA6\) (NCBI reference NM_001166111.1), remained as a candidate gene between the Boucher-Neuhäuser and Gordon Holmes syndrome index patients with segregating variants that were conserved, rare, and predicted to be damaging by at least three different \textit{in silico} algorithms (MutationTaster; MutationAssessor; Likelihood Ratio Test; and PolyPhen2).
Sanger Sequencing of the Candidate Gene in Additional Boucher-Neuhaüser Syndrome Families

To confirm the significance of PNPLA6 mutations in the pathogenesis of ataxia-hypogonadism syndromes, we screened the PNPLA6 gene in index patients from four additional Boucher-Neuhäuser syndrome families by conventional Sanger sequencing. Oligonucleotide sequences are available upon request.

PNPLA6 Analysis in More Than 500 Exomes of Index Patients with Ataxia, Hereditary Spastic Paraplegia and Charcot–Marie–Tooth Disease

To explore the possibility that Boucher-Neuhäuser and Gordon Holmes syndromes are clusters on a wide spectrum of neurodegenerative disease, PNPLA6 was used as a seed in the analysis of exome data from 538 unrelated patients with the following non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome neurodegenerative diseases: non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome early-onset ataxia (age of onset >30 years) (n = 67), non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome complex hereditary spastic paraplegia (n = 144), pure hereditary spastic paraplegia (n = 192) and Charcot–Marie–Tooth disease (CMT) type 2 compatible with recessive disease (CMT2; n = 135). In addition, we screened for PNPLA6 mutations in 1,637 additional whole exomes from a wide range of (non-ataxia, non-hereditary spastic paraplegia, non-CMT2) neurological phenotypes. The latter were used as a control cohort to scrutinize the possibility that recessive mutations observed in
our target cohort are just a result of genetic variability of the PNPLA6 gene with common occurrence of unique or rare variants not related to the disease.

**Structural Models of Wild-Type and Mutant PNPLA6 Domains**

Structural models of various wild-type and mutant domains of human PNPLA6 (UniProt# Q8IY17-4) were built using the MODELLER software based on homology modeling \(^{169}\). Briefly, the structural model of wild-type phospholipid esterase domain (EST), which represents the site of location of Ser1045Leu, Thr1058Ile, Phe1066Ser, Val1100Gly, Val1110Met and Pro1122Leu mutations, was constructed using the crystal structure of the homologous EST catalytic domain of the plant patatin PAT17 (PDB# 1OXW) \(^{170}\). The structural models of various wild-type and mutant CNB1/2 domains in complex with cyclic adenine monophosphate were obtained using the crystal structure of the homologous CNB2 domain of the regulatory alpha-subunit of cyclic adenine monophosphate-dependent protein kinase A bound to cyclic adenine monophosphate (PDB# 1RGS) \(^{171}\). In each case, a total of 100 atomic models were calculated and the structures with the lowest energies, as judged by the MODELLER Objective Function, were selected for further analysis. The atomic models were rendered using RIBBONS \(^{172}\).

**Clinical Assessment**

All index patients carrying two pathogenic PNPLA6 variants as well as their affected siblings received a systematic clinical assessment for disturbances
in multiple neurological systems (Table 6.1). In all subjects, routine MRI was performed.

**PNPLA6 Causes Boucher-Neuhäuser Syndrome and Gordon Holmes Syndrome**

By intersecting the identified variants in whole exomes from two index patients with Boucher-Neuhäuser syndrome and one index patient with Gordon Holmes syndrome under recessive inheritance models, only *PNPLA6* remained as a candidate gene. All variants identified in these cases were conserved, had low frequency in the general population, and were predicted to be damaging by at least three different *in silico* algorithms (Table 6.2). We identified a homozygous missense mutation (c.[3173C>T];[3173C>T], p.[Thr1058Ile];[Thr1058Ile]) in Family IHG25190, compound heterozygous splice/missense mutations in Family ARCA_05 (c.[2212-1G>C];[3328G>A], p.[Val738Glnfs*98];[Val1110Met]), and compound heterozygous frameshift/missense mutations (c.[3084_3085insGCCA](;)[4084C>G], p.[Arg1031Glufs*38](;)[Arg1362Gly]) in Family IHG25330. The splice mutation in Family ARCA_05 destroys a known splice acceptor site in intron 19-20, most likely resulting in skipping of exon 20 (76 bp) and a consecutive shift of the reading frame; *in silico* analysis did not indicate activation of cryptic splice sites (Table 6.2; for pedigrees and electropherograms see Figure 6.1 and Figure 6.2) \(^{173}\).
Table 6.1. Clinical features of PNPLA6 patients

<table>
<thead>
<tr>
<th>Family identifier</th>
<th>Subject</th>
<th>Origin</th>
<th>Gender</th>
<th>Phenotypic syndrome</th>
<th>Age of onset of first symptom (years)</th>
<th>Age at last examination (years)</th>
<th>Callus</th>
<th>Hypomagnesemia</th>
<th>Ocular dystrophy</th>
<th>Spasticity</th>
<th>MRI/CT</th>
<th>Tendon reflexes</th>
<th>Cognitive impairment</th>
<th>Imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIG25910</td>
<td>1</td>
<td>B</td>
<td>F</td>
<td>BNS</td>
<td>2, visual impairment</td>
<td>56</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>B</td>
<td>M</td>
<td>BNS</td>
<td>3, visual impairment</td>
<td>55</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>B</td>
<td>F</td>
<td>BNS</td>
<td>1, visual impairment and ataxia</td>
<td>53</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>B</td>
<td>F</td>
<td>BNS</td>
<td>2-3, visual impairment</td>
<td>48</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td>AICA-05</td>
<td>1</td>
<td>I</td>
<td>F</td>
<td>BNS</td>
<td>12, visual impairment and ataxia</td>
<td>44</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>I</td>
<td>F</td>
<td>BNS</td>
<td>6, gait ataxia</td>
<td>42</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>Mild</td>
</tr>
<tr>
<td>HIG25953</td>
<td>1</td>
<td>B</td>
<td>M</td>
<td>BNS</td>
<td>6, gait ataxia</td>
<td>61</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>B</td>
<td>M</td>
<td>BNS</td>
<td>6, gait ataxia</td>
<td>57</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
<tr>
<td>HIG25957</td>
<td>1</td>
<td>V</td>
<td>M</td>
<td>BNS</td>
<td>14, delayed puberty; visual impairment</td>
<td>26</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
<tr>
<td>HIG25933</td>
<td>1</td>
<td>F</td>
<td>M</td>
<td>GHS</td>
<td>14, delayed puberty</td>
<td>63</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
<tr>
<td>HIG26041</td>
<td>1</td>
<td>G</td>
<td>F</td>
<td>HSP</td>
<td>20, spasticity</td>
<td>54</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
<tr>
<td>HIG26117</td>
<td>1</td>
<td>G</td>
<td>M</td>
<td>spastic ataxia</td>
<td>4, gait ataxia</td>
<td>48</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>/-</td>
<td>+/---</td>
<td>1/1</td>
<td>N</td>
</tr>
</tbody>
</table>

M = male; F = female; N = normal; B = Brazilian; I = Italian; F = French; G = German; V = Venezuelan; BNS = Bocher-Neuhäuser syndrome; GHS = Corliss-Holmes syndrome; HSP = hereditary spastic paraplegia; O = absent; LL = lower limbs; PTR = plantar tendon reflex; ATR = achilles tendon reflex. Cognitive impairment was rated by clinical impression.
Table 6.2. *PNPLA6* mutations identified in this study

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Phenotype</th>
<th>cDNA change</th>
<th>Protein change</th>
<th>CVS function</th>
<th>GERP</th>
<th>PhastCons</th>
<th>phylop</th>
<th>dbSNP137 MAF</th>
<th>NHLBI EVS MAF</th>
<th>GEM, appMAF</th>
<th>MT</th>
<th>MA</th>
<th>LRT</th>
<th>PP2</th>
<th>SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCA_05</td>
<td>BNS</td>
<td>c.2212-1G&gt;C</td>
<td>p.Val738Gln*98</td>
<td>Splice-site</td>
<td>4.8</td>
<td>0.83</td>
<td>np</td>
<td>np</td>
<td>np</td>
<td>D L N D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCA_05</td>
<td>BNS</td>
<td>c.3326G&gt;A</td>
<td>p.Val1109Val</td>
<td>Missense</td>
<td>3.91</td>
<td>0.994</td>
<td>np</td>
<td>np</td>
<td>np</td>
<td>D L N D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25117</td>
<td>ATX</td>
<td>c.3086_3087insCGA</td>
<td>p.Arg1031Glu*38</td>
<td>Frameshift</td>
<td>4.94</td>
<td>1</td>
<td>np</td>
<td>0.07</td>
<td>0.00032</td>
<td>0.00033</td>
<td>D L N D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25117</td>
<td>ATX</td>
<td>c.3297T&gt;G</td>
<td>p.Val1100Gly</td>
<td>Missense</td>
<td>4.97</td>
<td>1</td>
<td>1.88</td>
<td>np</td>
<td>np</td>
<td>0.00033 D H D D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25641</td>
<td>HSP</td>
<td>c.767G&gt;A</td>
<td>p.Val256Ser</td>
<td>Missense</td>
<td>5.2</td>
<td>0.998</td>
<td>2.44</td>
<td>np</td>
<td>np</td>
<td>0.00033 D L D D 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25641</td>
<td>HSP</td>
<td>c.2514G&gt;A</td>
<td>p.Gly838Glu</td>
<td>Missense</td>
<td>4.66</td>
<td>1</td>
<td>2.59</td>
<td>np</td>
<td>np</td>
<td>0.00033 D M D D 0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25337</td>
<td>BNS</td>
<td>c.3134C&gt;T</td>
<td>p.Ser1045Leu</td>
<td>Missense</td>
<td>4.98</td>
<td>0.661</td>
<td>2.32</td>
<td>np</td>
<td>np</td>
<td>D M D D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25337</td>
<td>BNS</td>
<td>c.3305C&gt;T</td>
<td>p.Pro1102Leu</td>
<td>Missense</td>
<td>4.09</td>
<td>0.996</td>
<td>2.12</td>
<td>np</td>
<td>np</td>
<td>D M D D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25393</td>
<td>BNS</td>
<td>c.1732G&gt;T</td>
<td>p.Gly578Trp</td>
<td>Missense</td>
<td>5.17</td>
<td>0.999</td>
<td>2.41</td>
<td>np</td>
<td>np</td>
<td>D H D D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25393</td>
<td>BNS</td>
<td>c.3197T&gt;C</td>
<td>p.Phe1066Ser</td>
<td>Missense</td>
<td>5.1</td>
<td>1</td>
<td>1.93</td>
<td>np</td>
<td>np</td>
<td>D M D D 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25330</td>
<td>GHS</td>
<td>c.3084_3087insCGA</td>
<td>p.Arg1031Glu*38</td>
<td>Frameshift</td>
<td>4.94</td>
<td>1</td>
<td>-0.07</td>
<td>np</td>
<td>0.00032</td>
<td>0.00033</td>
<td>N N N B 0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25330</td>
<td>GHS</td>
<td>c.4084C&gt;G</td>
<td>p.Arg1362Gly</td>
<td>Missense</td>
<td>2.22</td>
<td>0.039</td>
<td>1.11</td>
<td>np</td>
<td>np</td>
<td>0.00033 N N N B 0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHGC25510</td>
<td>BNS</td>
<td>c.3173C&gt;T</td>
<td>p.Thr1058Leu</td>
<td>Missense</td>
<td>4.98</td>
<td>1</td>
<td>2.32</td>
<td>np</td>
<td>np</td>
<td>0.00066 D M D D 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BNS = Boucher-Neuhouser syndrome; ATX = ataxia telangiectasia; HSP = hereditary spastic paraplegia; GHS = Gordon-Hailes syndrome; MAF = minor allele frequency; MT = MutationTaster; MA = MutationAssessor; LRT = Likelihood Ratio Test; PP2 = PolyPhen2; SIFT = Sorting Tolerant From Intolerant; D = damaging; I = low probability of damaging effect; H = high probability of damaging effect; B = benign; M = medium probability of damaging effect.

N = neutral; P = probably damaging; U = unknown; np = not present.

GRAP = Genome: Evolutionary Rate Profiling; GVS = Genome Variant Server.

SIFT scores < 0.05 represent damaging effect.
To confirm the significance of *PNPLA6* mutations in the pathogenesis of ataxia-hypogonadism syndromes, index patients from four additional independent Boucher-Neuhäuser syndrome families were screened for *PNPLA6* mutations by conventional Sanger sequencing. We identified two additional Boucher-Neuhäuser syndrome families with *PNPLA6* mutations: (Family IHG25357: c.[3134C>T]; [3365C>T], p.[Ser1045Leu]; [Pro1122Leu]; and Family IHG25353: c.[1732G>T]; [3197T>C], p.[Gly578Trp]; [Phe1066Ser]) (Table 6.2; for pedigrees and electropherograms see Figure 6.1 and Figure 6.2).

**PNPLA6 Causes a Wide Spectrum of Neurodegenerative Disease**

To explore the possibility that Boucher-Neuhaüser syndrome and Gordon Holmes syndrome are only clusters on a wider spectrum of neurodegenerative disease, *PNPLA6* was used as a ‘seed’ in the analysis of exome data from 538 unrelated patients with non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome neurodegenerative diseases (non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome early-onset ataxia; non-Boucher-Neuhäuser syndrome/Gordon Holmes syndrome hereditary spastic paraplegia; CMT type 2) where family history was compatible with recessive disease. This analysis identified a spastic ataxia patient (Subject IHG26117), who carried significant
Figure 6.1. Pedigrees and Segregation of the Mutations Identified in **PNPLA6**. Roman numerals indicate the generation and numbers represent individuals within a generation. Circles represent females, squares represent males, filled symbols represent affected individuals, and arrows represent the proband. Abbreviations: mut – mutation and + - wild type.
Figure 6.2. Electropherograms of Mutations Identified with Sanger Sequencing. Corresponding mutations and individuals are indicated to the left of each electropherogram.
compound heterozygous changes (c.[3084_3085insGCCA]; [3299T>G], p.[Arg1031Glufs*38]; [Val1100Gly]) and a case with hereditary spastic paraplegia and mild motor neuropathy (Subject IHG26041) with compound heterozygous variants (c.[787G>A]; [2519G>A], p.[Val263Ile]; [Gly840Glu]) in PNPLA6 (Table 6.2).

In total, we report seven families from five different countries with significant novel mutations in PNPLA6. These variants co-segregated with the disease in all families where family members were available (Figure 6.1). All affected base pairs are highly conserved across species (Figure 6.3) and were absent or had low minor allele frequency in GEM.app (2,175 exomes), dbSNP137, and NHLBI ESP (6,500 exomes) (Table 6.2). Remarkably, in 1,637 additional whole exomes from a wide range of neurological phenotypes no additional homozygous or compound heterozygous changes were present in PNPLA6 under relaxed filter conditions (minor allele frequency in GEM.app, dbSNP137 and NHLBI ESP 6500 exomes <3%, low quality scores, low conservation scores) (Fisher’s exact test, two-tailed, P = 0.0001).

**Structural in silico Protein Modeling**

The majority of identified PNPLA6 mutations clustered within a short stretch of <100 residues in the EST domain. This domain has been shown to de-esterify phosphatidylcholine, a major component of biological membranes, into its
Figure 6.3. Evolutionary conservation of *PNPLA6* mutations. The inter-species conservation of the amino acids affected by missense mutations identified in this study.
constituent fatty acids and glycerophosphocholine (Figure 6.4) \(^{174-177}\). Glycerophosphocholine serves as a precursor for the biosynthesis of acetylcholine, a key neurotransmitter involved in mediating cellular signalling in the nervous system. On the CNB1-CNB2-CNB3EST modular architecture of PNPLA6 (Figure 6.4), we observe that mutations from residues Gly840–Ser1031 associated with spasticity (spastic paraplegia and spastic ataxia) largely cluster at the N-terminal side of the EST domain. Towards the C-terminal end of the EST domain, a second mutational cluster between residues Ser1031–Pro1122 causes cerebellar ataxia and/or hypogonadism (Boucher-Neuhäuser syndrome, Gordon Holmes syndrome, spastic ataxia). Next, we built structural models of CNB1/2 and EST domains using the crystal structures of the homologous CNB2 domain of the regulatory alpha-subunit of cyclic adenine monophosphate-dependent protein kinase A bound to cyclic adenine monophosphate (Protein Data Bank #1RGS) and the homologous EST catalytic domain of the plant patatin PAT17 (Protein Data Bank #1OXW) (Figure 6.5) \(^{170,171}\). Our structural models reveal that the catalytic center of the EST domain adopts a funnel-like shape that constitutes the entry route for the phosphatidylcholine substrate, where the active site Ser1014/Asp1144 catalytic dyad is located deep at the base of the funnel. The EST missense mutations may result in misalignment of active site residues Ser1014/Asp1144 so as to compromise its enzymatic function. It is also highly probable that mutations such as Thr1058Ile and Pro1122Leu may block the entry of phosphatidylcholine substrate to the active site residues located deep at the
Figure 6.4. Schematic of PNPLA6 gene. Schematic of the exon-intron arrangement of PNPLA6 (NCBI reference NM_001166111.1), with positions of mutations identified in five families. Exons are indicated as black boxes. CNB1/2 and the phospholipid esterase functional domains are indicated by purple and orange boxes, respectively. The mutations are indicated and color-coded by phenotype observed.
Figure 6.5. Structural model of EST domain (grey) of human PNPLA6. The side chain moieties of active site residues (Ser1014/Asp1144) and those identified here to be prone to mutations (Ser1045/Thr1058/Phe1066/Val1100/Val110/Pro1122) are colored red and blue, respectively. The funnel-shaped catalytic centre of the EST domain—where the active site residues Ser1014/Asp1144 are located deep at the base of the funnel whereas the rim of the funnel represents the entry route for the phosphatidylcholine (PDC)_substrate (green arrow)—is represented by yellow dashed lines. Note that the structural model of EST domain was built using the crystal structure of the homologous EST catalytic domain of the plant patatin PAT17 (PDB# 1OXW) in MODELLER\textsuperscript{169, 170}. A total of 100 atomic models were calculated and the structure with the lowest energy, as judged by the MODELLER Objective Function, was selected for further analysis. The structural model was rendered using RIBBONS\textsuperscript{172}. 
base of the catalytic funnel. Finally, although the highly conserved Val1100Gly mutation is located at a distal site away from the catalytic center, it may also indirectly affect protein dynamics that could ultimately shut down the enzymatic activity of PNPLA6. Additionally, two changes were located in the CNB1 and 2 domains (p.Val263Ile, p.Gly578Trp) (Figure 6.4). Our analysis suggests that these two mutations will likely compromise the ability of one or both regulatory CNB1/2 domains to bind cyclic nucleotide-monophosphate binding domains (cNMPs) and their failure to undergo a conformational change in response to changes in intracellular concentration of cNMPs would thus keep the EST domain in an auto-inhibited state (Figure 6.6).

Clinical Findings

We aggregated clinical data from 12 affected subjects belonging to seven families. All four index subjects with Boucher-Neuhauser syndrome presented with the classical triad: visual impairment, ataxia and delayed puberty (Figure 6.7 A–C). Symptoms in the individuals with Boucher-Neuhäuser syndrome started before the age of 8 years and were slowly progressive in all affected individuals with variable progression rates both between and within families, leading to wheelchair-dependency in the most severely affected subjects. In-line with the original publications of Boucher-Neuhäuser and Gordon Holmes syndromes, two of four subjects with Boucher-Neuhäuser syndrome as well as the subject with Gordon Holmes syndrome showed clinical signs of upper motor
Figure 6.6. Structural models of wild type (left) and mutant (right) CNB1 (a) and CNB2 (b) domains in complex with cAMP. In each case, the CNB domain is colored yellow and cAMP red. The sidechain moieties of Val263Ile (CNB1) and Gly578Try (CNB2) mutations identified here are depicted in blue.
Figure 6.7. The continuous spectrum of PNPLA6-associated disease (A) Photograph of the full body of a male patient with Boucher-Neuhäuser syndrome (Subject IHG 25357) at age 26 years illustrating incomplete secondary sex characteristics with lack of body hair and gynecomastia. (B) Exemplary fundus photography of the Boucher-Neuhäuser syndrome Patient ARCA_05 showing chorioretinal degeneration, characterized by diffuse atrophy of choroidal vessels and the retinal pigment epithelium with pigment clumps. The optic nerve shows no signs of atrophy. (C) Sagittal T2-weighted MRI of this subject with Boucher-Neuhäuser syndrome shows marked cerebellar atrophy.
neuron disease (spasticity, positive extensor plantar reflex and/or brisk reflexes), including a spastic Boucher-Neuhäuser syndrome phenotype (Family ARCA_05).

Achilles tendon reflexes were reduced or absent in all subjects with Boucher-Neuhäuser syndrome, providing clinical evidence that peripheral neuropathy is commonly associated with Boucher-Neuhäuser syndrome in our subjects. Neuropathy was of sensorimotor axonal type in those subjects with Boucher-Neuhäuser syndrome where nerve conduction studies were available. Affected siblings within Boucher-Neuhäuser syndrome families showed broadly similar phenotypes and age of onset (Table 6.1). However, the initial symptom of disease differed between and within Boucher-Neuhäuser syndrome families: whereas disease started with visual impairments in some subjects with Boucher-Neuhäuser syndrome, it presented with gait ataxia or delayed puberty in others (Table 6.1). Moreover, in Boucher-Neuhäuser syndrome Family ARCA_05 one of the affected siblings (Subject 2) did not show evidence for chorioretinal dystrophy indicating this as a non-obligate feature. This is further supported by the findings that the cases with hereditary spastic paraplegia and spastic ataxia showed neither retinal dystrophy nor hypogonadism (Table 6.1), thus demonstrating that hypogonadism is not an obligate feature of disease. Disease in the cases with hereditary spastic paraplegia and spastic ataxia also started before the age of 20 years, corroborating the notion that early-onset neurodegeneration is a common denominator across all affected subjects.
Discussion

To date, the genetic basis of Boucher-Neuhäuser and Gordon Holmes syndromes has not been identified. Here we use exome sequencing to uncover the genetic cause of these two clinical entities and demonstrate that they are allelic diseases, both caused by recessive *PNPLA6* mutations. *PNPLA6* seems to be a major cause of Boucher-Neuhäuser syndrome with four of six Boucher-Neuhäuser syndrome families carrying *PNPLA6* mutations. As not all subjects with Boucher-Neuhäuser syndrome carried *PNPLA6* mutations, the genetic basis of Boucher-Neuhäuser syndrome might either be heterogeneous or caused by *PNPLA6* mutations not detected by routine sequencing procedures (e.g. deletions or intronic mutations). Further heterogeneity of ataxia-hypogonadism syndromes is also supported by the recent observation that mutations in *RNF216* and the combination of mutations in *RNF216* and *OTUD4* cause ataxia-hypogonadism syndromes complicated by dementia \(^{178}\). In contrast to RNF216/OTUD4-associated ataxia-hypogonadism, Boucher-Neuhäuser syndrome and Gordon Holmes syndrome are not typically accompanied by dementia.

Complicated hereditary spastic paraplegia has been considered the prominent phenotype of *PNPLA6* thus far \(^{179}\). However, here we show that complicated hereditary spastic paraplegia represents only a relatively special case along a multi-dimensional *PNPLA6*-associated spectrum of neurodegenerative disorders. This spectrum includes at least four clinical key
features: ataxia, motor neuron disease (upper motor neuron disease with or without additional lower motor neuropathy), hypogonadism and chorioretinal dystrophy (Figure 6.8). Although these clinical features appear to be frequent in PNPLA6 disease none of them is an obligate feature of the disease (Table 6.1 and Figure 6.8). We postulate that new phenotypic combinations on this PNPLA6 associated disease spectrum will be identified as more patients representing broader phenotypes will be screened for mutations (Figure 6.8). PNPLA6 therefore appears to be another representative of a growing number of disease-related genes that have been shown to cause a spectrum or even a multitude of seemingly unrelated phenotypic expressions. Driven by the more comprehensive screening capabilities of next-generation sequencing based approaches, we will likely uncover a more complex structure of phenotype/genotype correlations in the coming years. On a functional level, proteins like PNPLA6, on which different phenotypes converge, may represent functional platforms or hubs that connect a diversity of pathways with multiple biological functions. These ‘hub proteins’, intersection nodes in phenotypic as well as cellular disease networks, might offer promising opportunities for therapeutic intervention for a number of diseases.

PNPLA6 belongs to a protein family of nine patatin-like phospholipase domain-containing proteins. The phospholipid esterase domain (EST) is altered by intoxication of organophosphorous compounds. These compounds are used in industry, agriculture, suicide attempts, chemical warfare, terrorist
Figure 6.8. The phenotypic spectrum caused by PNPLA6 mutations. The clinical spectrum of PNPLA6 mutations unfolds along four different neurological key features: ataxia, chorioretinal dystrophy, hypogonadotropic hypogonadism and motor neuron disease (upper motor neuron disease with or without additional lower motor neuropathy). Accordingly, Boucher-Neuhäuser syndrome (BNS) and Gordon Holmes syndrome (GHS) are not distinct entities, but clusters on a continuous spectrum of PNPLA6-associated disease, extending from Boucher-Neuhäuser syndrome via Gordon Holmes syndrome to spastic ataxia (sATX) and pure hereditary spastic paraplegia (HSP). The phenotype of complicated hereditary spastic paraplegia, which has been considered the most prominent phenotype of PNPLA6 so far, is only the ‘tip of the iceberg’ of this broad disease spectrum.
incidents (1995 Tokyo subway incident) and adulterated alcoholic beverages (Jamaica ‘Ginger Jake’ during the Prohibition era) \(^1\)\(^{111,182}\). Structural analysis of the various PNPLA6 mutations indicated that all mutations identified here harbor the potential to seriously affect the enzymatic activity of the EST domain. Based on current knowledge there are two main pathways of action for PNPLA6. Firstly, PNPLA6 de-esterifies phosphatidylcholine, a major component of biological membranes, into its constituent fatty acids and glycerophosphocholine \(^1\)\(^{174-177}\). Glycerophosphocholine serves as a precursor for the biosynthesis of acetylcholine, a key neurotransmitter involved in mediating cellular signaling in the nervous system. This may lead to disturbance of development and maintenance of synaptic connections in a variety of neuronal networks. Secondly, PNPLA6—which possesses lysophospholipase activity \(^1\)\(^{176}\)— has recently been suggested to catalyse 2-arachidonoyl lysophosphatidylinositol, which brings it into close functional relationship with other hereditary spastic paraplegia genes \(DDHD1/SPG28, CYP2U1/SPG49\) \(^2\)\(^{26}\). Thus, PNPLA6 disease might need to be added to this rapidly increasing list of genes involved in lipid metabolism associated with neurodegenerative disease \(^2\)\(^{25-27,132}\). Our PNPLA6 findings demonstrate that the phenotypic spectrum of these neurodegenerative phospholipid diseases is much broader than previously thought, extending from Boucher-Neuhäuser syndrome/Gordon Holmes syndrome to spastic ataxia and hereditary spastic paraplegia with or without motor neuropathy.
Summary of Dissertation Findings

In this final chapter, I will summarize the important findings of my dissertation and explain how the results demonstrate the vital role that mapping of disease-associated mutations plays in understanding the underlying mechanisms of axonal degenerative disorders. Prior to the work in this thesis, mutations in 20 genes and mutations in 50 genes were implicated in HSP and CMT, respectively. However, these genes only explained a fraction of cases suffering from these diseases. In my thesis, I present the genomic and functional evaluation of mutations discovered in several novel disease-associated genes in patients suffering from a range of inherited neuropathies. The findings in my dissertation took advantage of NGS, which is a powerful technique for exploring the underlying genetics of human diseases, but has introduced new computational challenges. I have started to address the challenges and difficulties with the analysis of NGS datasets by developing GEM.app. Together, this body of work further highlights the importance of efficient NGS analysis strategies to explore the genetics of human disease and how it can lead to the discovery of disease-associated mutations ultimately resulting in better resolution of the underlying biological pathways and mechanisms of axon degeneration.
The development of NGS has revolutionized human genetics research and how we approach identifying the underlying genetic variant in many Mendelian disorders. To take advantage of this new sequencing technology, I developed a novel analysis platform to efficiently analyze NGS data for inherited axonopathies, which has led to the identification of numerous genes $^{23-27, 61, 82, 135, 147, 184-192}$. The GEM.app framework created here will serve as a platform for assisting researchers to continue to identify novel gene defects that cause inherited neuropathies and I hope it will provide additional insights into the pathways and biological mechanisms in axonal degenerative disorders.

**Expanding the Number of Axonopathy Genes: MARS and DDHD2**

My work has yielded several significant advances towards further characterizing the pathways involved in inherited neuropathies. Many of the genes identified throughout this thesis have provided further evidence for emerging biological pathways contributing to CMT2 and HSP. First, the identification of a significant mutation in *MARS* in a late-onset CMT2 family further underscores the importance of protein translation and the aminoacyl tRNA synthetase pathway in peripheral neuropathies. This finding is supported by strong evolutionary conservation and strong functional evidence that the p.Arg618Cys mutation in *MARS* represents a loss-of-function allele *in vivo*. This is consistent with reports by other groups that have demonstrated that variants in ARS genes result in loss-of-function $^{65}$. In addition, I have identified mutations in *DDHD2* in two families with complicated HSP emphasizing the role of fatty-acid
and phospholipid metabolism in axonal degenerative disorders. Together with collaborators, we have observed that mitochondrial-membrane lipid composition is critical for maintaining proper bioenergetic functions. The alteration of lipid metabolism has already been shown to impact mitochondrial functions and trigger secondary cellular dysfunction. Thus, the action of DDHD2 might regulate the content of PA, a phospholipid on the surface of mitochondria that regulates fusion, which suggests DDHD2 involvement of mitochondrial fusion.

**Managing the Challenges of Genome Analysis**

Over the last decade, the throughput of sequencing has increased exponentially as the decrease in cost of sequencing has rivaled Moore’s law (Figure 7.1). This has improved the accessibility of NGS to a larger audience of physicians and biomedical researchers. However, this has created a new ‘bottleneck’, which is the analysis and management of the enormous amount of data produced by these technologies. Specifically, there are three main challenges posed by NGS data: (1) effectively managing larger exome and/or genome datasets, especially for smaller labs; (2) direct hands-on analysis and contextual interpretation of variant data in large genomic datasets; and (3) sharing of data produced by many small and medium-sized clinical and research-
Figure 7.1. **Graph illustrating the nature of the reductions in DNA sequencing costs.** The green line represents the cost of a whole genome sequence and the white line represents the hypothetical data reflecting Moore's Law (http://genome.gov/sequencingcosts).
based investigative teams around the world. If we can improve the efficiency of analyzing this data and facilitate secure and easy ways to share sequencing data, it will significantly increase the opportunities for the entire community to identify new genes in axonopathies. To address these challenges, I have significantly contributed to the development of GEM.app, a tool that has facilitated large-scale data-sharing and analysis.

**Defining the Phenotypic Spectrum of Disease Genes: PNPLA6 and VCP**

Large collections of WES from neurodegenerative familial diseases offers the opportunity to identify the genetic basis of clinical syndromes and to uncover the shared genetic etiology of clinical entities not previously considered biologically related. I conducted analysis of more than 500 exomes from families with varying clinical expressions of ataxia (with or without complicating features such as hypogonadism; n=67), complex hereditary spastic paraplegia (n=144), and pure hereditary spastic paraplegia (n=192) and Charcot-Marie Tooth disease type 2 (n=135). I discovered that PNPLA6 mutations can cause a broad spectrum of phenotypic features including ataxia, spasticity with or without motor neuropathy, hypogonadism and chorioretinal dystrophy. The previously clinically defined disease entities are not distinct entities, but rather represent phenotypic clusters on this spectrum. Another example of the expansion of clinical spectra associated with shared underlying genetic variation was observed with the identification of a novel mutation in VCP in a CMT2 family. Previous research has implicated mutations in VCP as resulting in ALS and IBMPFD.
Here, I argue that a CMT2 phenotype may be considered a milder form of IBMPFD/ALS since some patients with IBMPFD/ALS have been reported to have sensory/motor neuropathies in addition to the other clinical features. These findings provide convincing examples of how genetics can uncover that some clinically defined disorders may be biologically related. In the future, this will result in redefining of their clinical definitions. As genomic sequencing becomes more accessible and more patients undergo genetic testing, we will undoubtedly discover that many clinical syndromes actually share underlying genetic networks.

**Translational Implications and Discussion**

Despite astounding progress in Mendelian disease gene identification in axonopathies, specifically Charcot-Marie-Tooth disease and hereditary spastic paraplegia, an estimated 70% of axonal CMT2 patients and 40% of HSP patients do not have a mutation in any of the known genes. To address the limitations in current genetic diagnostics, I have been a part of many collaborative studies that have identified novel mutations in genes causing HSP or HSP-related disorders (\textit{B4GALNT1}, \textit{DDHD1}, \textit{CYP2U1}, \textit{DDHD2}, \textit{GBA2}, \textit{REEP2}, \textit{KIF1C}, \textit{STUB1}, \textit{PNPLA6}, \textit{C19orf12}) and novel peripheral neuropathy associated genes (\textit{MARS}, \textit{FBXO38}, \textit{VCP}, \textit{SYT2}, \textit{HARS}, \textit{RFVT2}, \textit{GARS}, \textit{BICD2}). These findings are already assisting patients in receiving genetic diagnoses that have been elusive for many years.
I played a major role in the identification of mutations in \textit{VCP} and \textit{PNPLA6} that refined the phenotypic spectrum associated with these specific genetic defects. I believe understanding the phenotypic spectrum associated with genetic variants is important for proper clinical and genetic diagnosis. Some suggest that clinical diagnosis should move toward describing patient symptoms, rather than classifying specific diseases, especially in heterogeneous disorders. In heterogeneous disorders such as CMT and HSP, patients can present with a number of different symptoms, which may be caused by a varying set of genes. Therefore, understanding the genetic/phenotypic associations can assist physicians in correctly deciding which genetic diagnostic test is most appropriate for each patient.

The identification of the remaining genes involved in axonal degeneration will undoubtedly lead to more patients receiving a genetic diagnosis. A better understanding of which alleles cause disease will not only lead to further insights into the biological mechanisms, but will eventually reveal useful therapeutic targets to treat these disorders. In summary, my dissertation research has advanced our understanding of the underlying genetic causes of axonal degenerative disorders and has resulted in the development of a novel analysis platform for further investigating the genetics of inherited neurodegenerative disorders.
References


