IDENTIFICATION OF DISEASE CAUSING MUTATIONS IN EARLY-ONSET NEUROPATHIES BY WHOLE EXOME SEQUENCING

Rosanna Pöyhönen
Pro gradu
Faculty of Medicine
University of Helsinki
2013
TABLE OF CONTENTS

1. Abbreviations .................................................................................................................. 6
2. Introduction ..................................................................................................................... 8
3. Review of the literature .................................................................................................. 9
    3.1 Clinical features of Charcot-Marie-Tooth neuropathy ........................................... 9
        3.1.1 Axonal Charcot-Marie-Tooth neuropathies (CMT2) ....................................... 11
            3.1.1.1 The subtypes of CMT2 and their clinical characteristics ....................... 12
            3.1.1.2 Clinical diagnostics of CMT2 ................................................................. 15
            3.1.1.3 The treatment of CMT2 patients ............................................................. 16
        3.1.2 The disease genes associated with CMT2 ....................................................... 17
            3.2.1 Disease genes for autosomal dominant CMT2 ......................................... 18
            3.2.2 Disease genes for autosomal recessive CMT2 ....................................... 20
    3.2 Next-generation sequencing ...................................................................................... 22
        3.3.1 Whole exome sequencing ............................................................................... 23
        3.3.2 Applying whole exome sequencing to Mendelian disorders ......................... 24
        3.3.3 The workflow of whole exome sequencing .................................................... 24
            3.3.3.1 Technical steps ....................................................................................... 24
            3.3.3.2 Filtering ................................................................................................. 25
        3.3.4 The challenges of whole exome sequencing .................................................. 26
        3.3.5 Whole exome sequencing as a diagnostic tool ............................................. 26
            3.3.5.1 Identifying recessive and dominant variants .......................................... 27
            3.3.5.2 Identifying new disease genes by whole exome sequencing .................. 28
            3.3.5.3 Identifying mutations in known disease genes by whole exome sequencing 28
    4. Aims of the study .......................................................................................................... 30
    5. Materials and methods ............................................................................................... 31
        5.1 Patients ..................................................................................................................... 31
            5.1.1 Patient 1 ......................................................................................................... 31
            5.1.2 Patient 2 ......................................................................................................... 32
            5.1.3 Patient 3 ......................................................................................................... 32
            5.1.4 Patient 4 ......................................................................................................... 32
            5.1.5 The collection of patient samples .................................................................... 33
        5.2 Whole exome sequencing ........................................................................................ 33
            5.2.1 Whole exome sequencing .............................................................................. 33
            5.2.2 Analyzing tools of WES data ....................................................................... 34
6.1.5 The TRIM2 variants are not present in the general population

6.2 TRIM2, A possible novel disease gene for early-onset neuropathies

6.3 mTrim2 site-directed mutagenesis

6.4 Western blotting of mTrim2 constructs

7. Discussion

7.1 Improving the process of analyzing variants from WES data

7.1.1 Analyzing WES data

7.1.1.1 The analyzing steps of WES

7.1.1.2 The success of WES analysis

7.1.1.3 The classification of the interesting variants

7.2 Identifying the molecular cause of disease for four patients suffering from axonal neuropathies using WES

7.2.1 The re-sequencing of the variants

7.2.2 The missing cause for disease for patients 1, 3 and 4

7.2.3 Identifying candidate disease gene for patient 2

7.2.3.1 TRIM2 belongs to E3 ligase protein family

7.2.3.2 Previous investigations about TRIM2

7.2.4 The possible effects of the missense mutation and deletion on the function of TRIM2 protein

7.2.5 TRIM2 carrier frequency

7.2.5 The limitations of verifying the causative nature of TRIM2 mutations

7.3 Performing functional analysis of identified potential disease mutations

7.3.1 mTrim2 mutagenesis and western blotting

7.4 Subsequent research on verifying the causative nature of the TRIM2 missense mutation and deletion

7.5 Future prospects and ethical considerations of WES

7.5.1 What the future holds for WES

7.5.2 The ethical considerations of WES analysis

8. Acknowledgements

9. References
Charcot-Marie-Tooth (CMT) neuropathy is one of the most common forms of inherited peripheral neuropathies with the prevalence of one in 2500 individuals. CMT is phenotypically and genetically a very heterogeneous disease. It can be inherited as an autosomal recessive, dominant or X-linked trait. CMT is characterized by distal muscle weakness, atrophy and deformity of the feet as well as clumsiness of gait. The onset of CMT varies and also the symptoms of the disease can vary even among the members of a single family. So far more than 40 genes have been identified for CMT and the list is estimated to grow by 30-50 genes.

Whole exome sequencing (WES) is a new next generation sequencing technique, which targets the protein-coding area of the genome. Through WES analysis it is possible to search for disease causing mutations with all kinds of inheritance patterns. Patients suffering from CMT are good candidates for WES analysis because of the genetic heterogeneity of their disease. WES can be used for diagnosing Mendelian disorders with atypical symptoms as well as diseases, which are difficult to confirm using clinical criteria alone and which require costly evaluation, e.g. CMT.

In this master study new disease causing mutations for early-onset neuropathies are identified by whole exome sequencing (WES). The aims of this study include using WES for the molecular diagnosis of four patients suffering from early-onset axonal neuropathies, the functional analysis of possible causative variants and improving and developing the process of analyzing variants from whole exome sequencing data, especially the analyzing steps of insertion and deletion variants. Finding causative variants among the insertion and deletion variants has previously been often left out from the WES analysis because of the lack of systematic analysis technique.

As a result of the WES data analysis a new candidate disease gene, tripartite motif containing 2 (TRIM2) was identified. A missense mutation c.761T>A (p.E254V) and a deletion c.1779delA (p.K594Rfs7X) were found in patient 2, who suffers from severe CMT type 2. The carrier frequency was analysed to see whether the variants are present in the general population or not. The functional analysis of TRIM2 was started by preparing constructs carrying the missense mutation and the deletion and by setting up conditions for western blotting.

Keywords
Charcot-Marie-Tooth neuropathy, whole exome sequencing, disease gene, molecular diagnosis, functional analysis

Where deposited

Additional information
## 1. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AARS</strong></td>
<td>Alanyl-tRNA synthetase</td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td>Autosomal dominant</td>
</tr>
<tr>
<td><strong>AR</strong></td>
<td>Autosomal recessive</td>
</tr>
<tr>
<td><strong>Bim</strong></td>
<td>Bcl-2-interacting mediator of cell death</td>
</tr>
<tr>
<td><strong>BSN</strong></td>
<td>Bassoon presynaptic cytomatrix protein</td>
</tr>
<tr>
<td><strong>CAMP</strong></td>
<td>Compound action motor potential</td>
</tr>
<tr>
<td><strong>CMT</strong></td>
<td>Charcot-Marie-Tooth neuropathy</td>
</tr>
<tr>
<td><strong>CNV</strong></td>
<td>Copy number variation</td>
</tr>
<tr>
<td><strong>DYNCR1</strong></td>
<td>Dynamin, cytoplasmic 1, heavy chain 1</td>
</tr>
<tr>
<td><strong>FIMM</strong></td>
<td>Institute for Molecular Medicine Finland</td>
</tr>
<tr>
<td><strong>FXN</strong></td>
<td>Frataxin</td>
</tr>
<tr>
<td><strong>GARS</strong></td>
<td>Glycyl-tRNA synthetase</td>
</tr>
<tr>
<td><strong>GDAP1</strong></td>
<td>Ganglioside induced differentiation associated protein 1</td>
</tr>
<tr>
<td><strong>GTF2IRD2</strong></td>
<td>GTF2I repeat domain containing 2</td>
</tr>
<tr>
<td><strong>HMSN</strong></td>
<td>Hereditary motor and sensory neuropathy</td>
</tr>
<tr>
<td><strong>HSPB1</strong></td>
<td>Heat shock 27 kDa protein 1</td>
</tr>
<tr>
<td><strong>HSPB8</strong></td>
<td>Heat shock 22 kDa protein 8</td>
</tr>
<tr>
<td><strong>Hsp</strong></td>
<td>Heat shock protein</td>
</tr>
<tr>
<td><strong>IGV</strong></td>
<td>Integrated Genome Viewer</td>
</tr>
<tr>
<td><strong>IM</strong></td>
<td>Inner mitochondrial membrane</td>
</tr>
<tr>
<td><strong>IMS</strong></td>
<td>Intermembrane space</td>
</tr>
<tr>
<td><strong>Indel</strong></td>
<td>Insertion and deletion variants</td>
</tr>
<tr>
<td><strong>KIFB1</strong></td>
<td>Kinesin family member 1B</td>
</tr>
<tr>
<td><strong>LMNA</strong></td>
<td>Lamin A/C</td>
</tr>
<tr>
<td><strong>LRSAM1</strong></td>
<td>Leucine rich repeat and sterile alpha motif containing 1</td>
</tr>
<tr>
<td><strong>MFN</strong></td>
<td>Mitofusin 2</td>
</tr>
<tr>
<td><strong>mNCV</strong></td>
<td>Motor nerve conduction velocities</td>
</tr>
<tr>
<td><strong>MPZ</strong></td>
<td>Myelin protein zero</td>
</tr>
<tr>
<td><strong>NCV</strong></td>
<td>Nerve-conduction velocities</td>
</tr>
<tr>
<td><strong>NEFL</strong></td>
<td>Neurofilament, light polypeptide</td>
</tr>
<tr>
<td><strong>NGS</strong></td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td><strong>OM</strong></td>
<td>Outer mitochondrial membrane</td>
</tr>
<tr>
<td><strong>OXPHOS</strong></td>
<td>Oxidative phosphorylation complex</td>
</tr>
<tr>
<td><strong>PCR</strong></td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td><strong>PLEC</strong></td>
<td>Plectin</td>
</tr>
<tr>
<td><strong>qPCR</strong></td>
<td>Real-time PCR</td>
</tr>
<tr>
<td><strong>RAB7A</strong></td>
<td>Member RAS oncogene family</td>
</tr>
<tr>
<td><strong>SNP</strong></td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td><strong>SNV</strong></td>
<td>Single nucleotide variation</td>
</tr>
<tr>
<td><strong>SYNM</strong></td>
<td>Synemin, intermediate filament protein</td>
</tr>
<tr>
<td><strong>TRIM2</strong></td>
<td>Tripartite motif containing 2</td>
</tr>
<tr>
<td><strong>TRPV4</strong></td>
<td>Transient receptor potential cation channel, subfamily V, member 4</td>
</tr>
<tr>
<td><strong>UCE</strong></td>
<td>Ultra-conserved element</td>
</tr>
<tr>
<td><strong>VEP</strong></td>
<td>Visual evoked potential</td>
</tr>
<tr>
<td><strong>WES</strong></td>
<td>Whole exome sequencing</td>
</tr>
</tbody>
</table>
2. INTRODUCTION

Charcot-Marie-Tooth (CMT) neuropathy is phenotypically and genetically a very heterogeneous disease. It can be inherited as an autosomal recessive, dominant or X-linked trait. CMT is the most common peripheral neuropathy with an incidence of 1 in 2500 individuals. It is characterized by distal muscle weakness, atrophy and deformity of the feet as well as clumsiness of gait. The onset of CMT varies and also the symptoms of the disease can vary even among the members of a single family. So far more than 40 causative genes have been identified and the list of disease causing genes is estimated to grow by 30-50 genes.

In this study causative gene mutations for patients suffering from early-onset axonal neuropathies were searched via whole exome sequencing (WES). Patients suffering from CMT are good candidates for WES analysis because of the genetic heterogeneity of their disease. WES is broadly used for diagnosing Mendelian disorders with atypical symptoms as well as diseases, which are difficult to confirm using clinical criteria alone and which require costly evaluation, e.g. CMT. It can be described as a fast and cost-effective next-generation sequencing technology, which targets the protein-coding area of the genome. This part accounts for only 1 % of the whole human genome but however 85 % of Mendelian diseases are caused by mutations in this genomic space.

At the moment the identification of possible causative mutations by WES can be described as laborious work, partly due to the large amount of sequencing errors and false-positive variants produced by WES. Nevertheless, through WES analysis it is possible to search for disease causing mutations with all kinds of inheritance patterns. Finding causative variants among the insertion and deletion (indel) variants has previously been sometimes left out of WES analysis, because of the lack of a systematic analysis technique.

In this study also the indel variants are viewed. Homozygous, compound heterozygous and heterozygous variants combined with heterozygous indel variants are searched for disease causing mutations. Also all the analyzing steps of identifying these variants are presented and further developed. As a result of WES analysis a potential disease gene, tripartite motif containing 2 (TRIM2) was identified in one of the studied patients suffering from severe CMT type 2. The functional analysis of this gene mutant was started and the role of TRIM2 as a new disease causing gene further investigated.
3. REVIEW OF THE LITERATURE

3.1 CLINICAL FEATURES OF CHARCOT-MARIE-TOOTH NEUROPATHY

Charcot-Marie-Tooth neuropathy (CMT), also known as hereditary motor and sensory neuropathy (HMSN) includes a group of motor and sensory peripheral neuropathies. These are the most common form of inherited peripheral neuropathies with the prevalence of one in 2500 individuals. As a disease CMT is genetically and phenotypically heterogeneous. It is also characterized by extensive locus heterogeneity. CMT affects the peripheral nerves, which are the nerves throughout the body that carry information to and from the spinal cord. CMT decreases the ability of these nerves to carry motor commands to the muscles, especially those furthest from the spinal cord located in the feet and hands. As a result, the muscles connected to these nerves eventually weaken. CMT also affects the sensory nerves that carry information from the limbs to the brain. Therefore, the patients with CMT also may have sensory loss.

CMT can appear in infancy or in adulthood, but usually the symptoms start before the age of 20 years. Some patients do not experience symptoms until their early thirties or forties. The severity of the symptoms and the progress of the disease can also vary, even among the members of the same family. CMT manifests clumsiness of gait, predominantly distal muscular atrophy of the limbs and deformity of the feet in the form of foot drop. Foot drop is the initial symptom, which is often seen early in the course of the disease. This can also cause claw toe, where the toes are fully curled. Also spasmodic muscular contractions can appear and be disabling when the disease activates. Other symptoms include scoliosis, malformed hip sockets, tremor and effects on the breathing, vision, hearing or speaking can follow. Pregnancy is known to exacerbate CMT, as well as extreme emotional stress and periods of prolonged immobility. Some CMT patients experience neuropathic pain, which varies in severity from case to case. CMT does not decrease the lifespan of the patients.

Different types of CMTs can be classified according to the inheritance pattern (autosomal dominant, autosomal recessive and X-linked), according to electrophysiological findings (demyelinating and axonal) or according to the causative disease gene (Figure 1). In 90% of the CMT cases the disease shows an autosomal dominant inheritance pattern. Defects in more than 40 genes have been described to cause different types of CMT making an accurate diagnosis increasingly possible. Autosomal dominant and X-linked CMTs are the most prevalent forms of CMT in the United Kingdom, Northern
European and American populations, whereas autosomal recessive CMT is more common in countries with high rates of consanguinity.

CMT patients can be assorted into three different categories based on the electrophysiological findings: demyelinating CMT1 patients, axonal CMT2 patients and intermediate CMT patients. Demyelinating forms of CMT affect approximately 80% of CMT patients. Demyelination means the loss of the myelin sheath insulating the nerves. When myelin degrades, conduction of signals along the nerve can be impaired or lost and the nerve eventually withers. One of the main functions of myelin is to increase nerve-conduction velocities (NCV) so it is logical that the most typical electrophysiological findings in demyelinating CMT are low NCVs. CMT1 patients demonstrate severely reduced motor nerve conduction velocities (NCVs < 38 m/s). Axonal forms of CMT affect approximately 20% of CMT patients and are caused by specific effects on the axon. Usually histopathologic studies show signs of axonal degeneration. NCVs of CMT2 patients are nearly normal (>38 m/s), while the amplitude of the motor and sensory action potentials are greatly reduced due to axonal loss. Intermediate forms of CMT (NCVs between 25 and 45 m/s) show signs of both demyelination and axonal loss.

This study concentrates on CMT type 2 axonal form, which is further introduced in the next section.
3.1.1 **Axonal Charcot-Marie-Tooth Neuropathies (CMT2)**

The CMT2 phenotype is highly heterogeneous with variable penetrance. Mutations associated with both early- and late-onset forms of the disease have been identified. CMT2 patients show normal or near-normal NCVs. Also peripheral nerves are not enlarged or hypertrophic. The subtypes of CMT2 are similar clinically and distinguished only by molecular genetics findings. CMT2 can be distinguished from CMT1 by electrophysiological study or DNA testing.

The majority of cases with CMT2 show an autosomal dominant inheritance pattern. However, some of the CMT2 cases are caused by recessively inherited mutations. Autosomal recessive forms of CMT2 are very rare compared to autosomal dominant ones. It has been known for several decades that the recessive cases of CMT present a more severe clinical phenotype and appear very early in childhood compared to the dominant cases. In some countries with a high prevalence of consanguineous marriages, autosomal recessive inheritance is likely to account for the great majority of all forms of CMT disease.
There are altogether 15 genes, which have been found to associate with CMT2. Many subtypes of CMT2 are associated with adult onset of symptoms 9.

Common symptoms of CMT2 patients comprise of muscle weakness and atrophy, sensory loss, reduced reflexes and abnormal gait. Possible additional symptoms of CMT2 patients include optical atrophy and tremors, atrophy of the limbs, migraine and effects on the central nervous system. Restless legs and sleep apnea have also been associated with CMT2 9.

3.1.1.1 The subtypes of CMT2 and their clinical characteristics

The most common form of CMT2 is CMT2A. There are also less frequent subtypes of CMT2: CMT2B-L, CMT2N, CMT2O, CMT2P, AR-CMT2A and AR-CMT2B. CMT2A and B can also be divided into A1, A2, B1 and B2 subtypes. In some cases the subtypes are named after single families or gene mutations and usually each subtype has its own typical symptoms. Still it is not always easy to distinguish the subtypes among all of the possible ones. From all of the CMT2 subtypes CMT2B1, CMT2B2, CMT2H and CMT2K are inherited in an autosomal recessive manner and all other subtypes are inherited in an autosomal dominant manner. CMT2P has been reported to be inherited in an autosomal recessive as well as in an autosomal dominant manner. The subtypes and the underlining gene mutations are presented in Table 1 in the following section 3.2.

CMT2A has a typical CMT phenotype with onset in the second or third decade of life. It is the most common form of CMT2, since 20 % of CMT2 patients suffer from CMT2A. The average symptoms of these patients comprise of distal muscle weakness and atrophy, less severe sensory loss and depressed tendon reflexes. NCVs usually fall within the normal or near-normal range, which is compatible with an axonal neuropathy. Some patients suffering from this subtype have also developed optic atrophy and pyramidal signs. Clinical heterogeneity exists in CMT2A such that patients are presented with both mild cases with gradual onset and severe cases with children requiring wheelchairs to be able to cope in their everyday life 10.

CMT2B is a rare disorder that presents with sensory impairment often including ulceromutilations, also called ulcerative-mutilating phenomena 4. The exact classification of CMT2B type is still somewhat controversy. CMT2B1 is characterized with a variable age of onset, although usually the disease begins in the second decade. Patients suffering from CMT2B have experienced functional disability ranging from mild to severe. CMT2B1 has been described in Northwestern African population as CMT2B2 has
occurred in a Spanish family with ancestry in Costa Rica. CMT2B2 patients present with an adult onset phenotype and the symptoms include for example symmetrical weakness and atrophy in the ankles. The NCVs of this family were normal or slightly reduced.

CMT2C exhibits paralysis of the vocal chords and phrenic nerves. It also affects the diaphragm, a muscle located horizontally between the chest cavity and stomach cavity. Some patients are registered with mild sensory loss. The phenotypes of these patients have been heterogeneous, ranging from the symptoms described above to scapulo-peroneal spinal muscular atrophy.

CMT2D is characterized by predominately distal motor weakness with wasting of the hand muscles. Patients usually present first with weakness of the small muscles of the hand and much later with the involvement of the distal lower limb muscles. This presentation is quite different than in most cases of CMT, where weakness and symptoms usually present first in the distal muscles of the lower leg.

CMT2E/1F has been reported in several families with a progressive sensory and motor neuropathy. CMT2E/1F comprises 1% of all CMT cases. Patients have often presented with moderately to severely reduced nerve conduction velocities. Clinically these patients are heterogeneous, as some patients have presented with classical slowly progressive phenotypes but many others have been severely affected. This heterogeneity is at least partly due to the variation in the different mutated parts of the causative gene’s protein product. The disease has been reported to begin with gait problems usually before the age of 13 years. Some patients have had a very early onset and presented with delayed motor milestones. Also all of the patients identified with CMT2E/1F have suffered from foot drop.

CMT2F and CMT2L have been reported in only two families. Inheritance in both autosomal dominant and autosomal recessive patterns has been described. Clinical phenotypes have been variable with most patients presenting with distal weakness of lower and upper limbs, sensory loss and decreased reflexes. Scoliosis, vocal cord paresis, glaucoma and cranial nerve involvement have also been reported. CMT2F patients usually show the classic axonal CMT phenotypes. Wasting of upper limbs muscles eventually leading to the clawing of the hands have been described as well. The disease onset of CMT2F patients has been between the age of 15 and 25 years with a slowly progressive course.

CMT2G was reported in a single large Spanish family. The patients presented with classical CMT2 symptoms such as foot deformity and difficulties with walking. The progression of their disease was very slow. Patients were also mildly disabled and the upper limbs were involved in only two patients.

Both CMT2H and CMT2K subtypes are both caused by mutations in a same gene. CMT2H has been associated with pyramidal features and patients have also suffered from vocal cord paralysis. The NCVs
of these patients have been between intermediate and slow numbers. The autosomal recessive forms of these subtypes have presented with a more severe phenotype than the autosomal dominant forms and also the symptoms have started earlier in the autosomal recessive forms. Symptoms of these subtypes also include absent tendon reflexes and optic atrophy.\(^1\)

CMT2I and 2J are also both caused by mutations in a same gene and in addition they exhibit conserved NCVs. The mutations causing both subtype have been predominantly associated with CMT1 but later also with late onset axonal CMT.\(^1\) Moreover, CMT2J is associated with hearing impairment, alterations of the pupils and severe sensory disruption.\(^4\) These subtypes are characterized with late-onset but the symptoms vary greatly. The molecular reasons for these varying phenotypes are not yet fully understood.\(^1\)

CMT2L has been reported in a single Chinese family, with onset between ages 15 and 33 years.\(^9\) Eighteen individuals from this large family were diagnosed with CMT2L that followed an autosomal dominant inheritance pattern. The NCVs of these patients were normal and the symptoms comprised of scoliosis, weakness of the lower limbs and mild sensory loss.\(^1\)

CMT2N has been reported in two French families and in one Australian family. All these patients have been found to present with a same loss-of-function mutation in one specific gene.\(^9\)

CMT2O has been reported in a large family with childhood onset of delayed motor milestones associated with progressive distal lower limb weakness and variable sensory loss. The patients had normal NCVs. Occasional proximal weakness and waddling gait were also noted.\(^9\)

CMT2P has been reported in a large Canadian family and in a three generation Dutch family. In the Canadian family the age of onset was young adulthood with progressive distal muscle weakness and atrophy. Nerve electrophysiology was consistent with an axonal neuropathy. The Dutch family was associated with onset in the second or third decade of slowly progressive distal weakness and atrophy with mild sensory loss similarly as in the Canadian family. Both of these CMT2P cases were caused by mutations in a same gene with either autosomal recessive or autosomal dominant inheritance. Also a multi-generational Sardinian family has been described with a splice site mutation resulting in a frame shift and stop codon.\(^9\)

AR-CMT2A has been described with proximal involvement and rapid progression. The patients have been shown to manifest also cardiomyopathy, lipodystrophy and muscular dystrophy. Hence, AR-CMT2A presents with differences when compared with other forms of CMT2. The wide variability
among patients for the age of onset, the disease course and the involvement of proximal limbs is quite unusual within the same subtype 13.

AR-CMT2B has been reported in a large consanguineous Costa Rican family. The clinical presentation of these patients was typical of a classical form of CMT. The symptoms of these patients usually started at adulthood and they were considerably milder than in the axonal-recessive forms of CMT caused by mutations in GDAP1 or LMNA genes. Patients with AR-CMT2 owing to GDAP1 mutations present a severe early onset phenotype variably associated with additional clinical features. Other symptoms have included distal symmetrical weakness, atrophy and hyporeflexia. The NCVs are usually normal or slightly reduced 13.

3.1.1.2 Clinical diagnostics of CMT2

The diagnosis of CMT2 patients is based on clinical, family and neurophysiological studies. The making of the clinical diagnosis for CMT2 patients starts by determining if the patient has a hereditary neuropathy or is the disease sporadic. Neuropathies can be caused by other than genetic factors, which need to be ruled out before determining the inheritance pattern of the disease.

When determining the inheritance pattern of the CMT2, it is most convenient to first rule out dominantly inherited causative mutations. If either one of the parents or some of their children present with symptoms, a dominant inheritance pattern can be suspected to underlie the disease. The occurrence of disease among siblings and parental consanguinity can suggest an autosomal recessive inheritance. Family surveys can be done in order to see if there exists a lineage of affected ancestors. However, sometimes the family survey turns out to be negative. In this case the clinicians can look at a number of factors, which point towards genetic neuropathies; childhood onset and prolonged as well as slowly progressive clinical course of the disease and the presence of foot drop 14. The absence of these symptoms, however, doesn’t rule out the possibility of a genetic neuropathy.

The next step is a neurophysiological examination, which includes the measurement of the nerve conduction velocities. This should be done in at least 3 nerves. NCVs are usually within the normal range (>40-45 m/s). Sometimes the NCVs can be in a mildly abnormal range (30-40 m/s). A median NCV of 38 m/s is often used as a threshold for differentiating CMT1 from CMT2 14. EMG testing can be done and results usually show evidence of an axonal neuropathy with such findings as polyphasic potentials, positive waves, fibrillations and reduced amplitudes of evoked motor and sensory responses.
Also the compound action motor potentials (CAMP) are greatly reduced\(^9\). Nerve biopsy, if done, shows loss of myelinated fibers with signs of regeneration, axonal sprouting and athropic axons with neurofilaments.

Usually the genetic background study of CMT2 begins with checking for point mutations in gene \(MFN2\). \(MFN2\) mutations are known to be the cause of CMT2 in 33\% of the cases in families where the inheritance pattern is known to be dominant. It is also known that mutations in the \(GJB1\) gene may explain up to 12\% of CMT cases (CMT2 and CMT1 together)\(^4\).

Some more challenge to the molecular diagnosis of CMT2 brings the fact that the CMT2 phenotype can result from mutations in genes primarily associated with CMT1. Some of the CMT2 subtypes also resemble other disease, which makes the diagnosis of these subtypes even more difficult.

### 3.1.1.3 The treatment of CMT2 patients

CMT is currently an untreatable disorder and at the moment the treatment of CMT is only supportive, as there are no drugs available that would halt the disease symptoms\(^15\). The therapy consists mainly of orthopedic surgeries, rehabilitation, symptomatic treatment of pain and depression and surgical corrections of foot and hand deformities\(^3\). Exercise is encouraged with the individual’s capability and in fact many patients remain physically active. Fewer than 5\% of CMT patients need wheelchairs. Forearm crutches or canes can be used for gait stability. Treatment of sleep apnea and restless legs may also be required. CMT patients are advised to avoid obesity, which can make walking and moving around difficult\(^9\).

It is particularly important that patients with CMT avoid other known risk factors for the development of neuropathies, such as the use of neurotoxins (e.g., drugs and alcohol)\(^9\). Even though CMT is at the moment incurable, there are some promising results from patient studies using ascorbic acid or progesterone in the treatment of CMT. Ascorbic acid has been found to help some CMT patients, especially patients suffering from CMT1A\(^\text{16}\).
3.2 The disease genes associated with CMT2

CMT has been discovered to be caused by mutations in more than 40 different genes. As previously mentioned, CMT can be inherited as an autosomal recessive (AR), autosomal dominant (AD) or X-linked trait.

The 15 genes known to be associated with the CMT2 subtypes are \textit{KIF1B} (CMT2A1), \textit{MFN2} (CMT2A2), \textit{RAB7A} (formerly \textit{RAB7}) (CMT2B), \textit{LMNA} (CMT2B1), \textit{MED25} (CMT2B2), \textit{TRPV4} (CMT2C), \textit{GARS} (CMT2D), \textit{NEFL} (CMT2E/1F), \textit{HSPB1} (CMT2F), \textit{MPZ} (CMT2I/J), \textit{GDAP1} (CMT2H/K), \textit{HSPB8} (CMT2L), \textit{AARS} (CMT2N), \textit{DYNC1H1} (CMT2O), and \textit{LRSAM1} (CMT2P). Some of these genes are shown to present with recessive inheritance pattern, although most of them usually follow dominant autosomal inheritance pattern as autosomal recessive CMT2 is a very rare condition. All the disease genes of CMT2 and the subtypes caused by mutations in these genes are presented in Table 1.

Table 1. The disease genes for CMT2 and the correlating CMT2 subtypes

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{KIF1B}</td>
<td>CMT2A1</td>
</tr>
<tr>
<td>\textit{MFN2}</td>
<td>CMT2A2</td>
</tr>
<tr>
<td>\textit{RAB7A}</td>
<td>CMT2B</td>
</tr>
<tr>
<td>\textit{LMNA}</td>
<td>CMT2B1</td>
</tr>
<tr>
<td>\textit{MED25}</td>
<td>CMT2B2</td>
</tr>
<tr>
<td>\textit{TRPV4}</td>
<td>CMT2C</td>
</tr>
<tr>
<td>\textit{GARS}</td>
<td>CMT2D</td>
</tr>
<tr>
<td>\textit{NEFL}</td>
<td>CMT2E/1F</td>
</tr>
<tr>
<td>\textit{HSPB1}</td>
<td>CMT2F</td>
</tr>
<tr>
<td>\textit{MPZ}</td>
<td>CMT2I/J</td>
</tr>
<tr>
<td>\textit{GDAP1}</td>
<td>CMT2H/K</td>
</tr>
<tr>
<td>\textit{HSPB8}</td>
<td>CMT2L</td>
</tr>
<tr>
<td>\textit{AARS}</td>
<td>CMT2N</td>
</tr>
<tr>
<td>\textit{DYNC1H1}</td>
<td>CMT2O</td>
</tr>
<tr>
<td>\textit{LRSAM1}</td>
<td>CMT2P</td>
</tr>
<tr>
<td>\textit{LMNA}</td>
<td>AR-CMT2A</td>
</tr>
<tr>
<td>\textit{MED25}</td>
<td>AR-CMT2B</td>
</tr>
<tr>
<td>\textit{GDAP1}</td>
<td>AR-CMT2</td>
</tr>
</tbody>
</table>
3.2.1. Disease genes for autosomal dominant CMT2

MFN2 is a common gene underlining both AD and AR forms of CMT2. It is the most common causative gene underlining CMT type 2A2. Mutations in the MFN2 gene account for approximately 30% of all CMT2 cases. Most of the mutations identified in MFN2 are missense mutations although nonsense mutations have also been described. Over 50 mutations in MFN2 have already been reported. Most of the mutations found in MFN2 are located in the GTPase, coiled-coil domain or the conserved R3 domain of the MFN2 protein. MFN2 belongs to the group of five disease genes of CMT that are seen to cause the classical phenotype of CMT. Other genes belonging to this group include MPZ, AARS, GDAP1 as well as NEFL.

MFN2 codes for mitofusin 2 protein, which is a GTPase that is involved in the fusion of mitochondria regulating the architecture of the mitochondrial network. MFN2 is ubiquitously expressed and is present in the spinal cord, muscle, heart and peripheral nerves. MFN2 is an outer mitochondrial membrane protein, which in cooperation with the MFN1 isoform has important roles in the regulation of mitochondrial fusion and a function essential for metabolic activity in eukaryotic cells. It has also been suggested that MFN2 may be associated with maintaining mitochondrial membrane potential.

Cellular assays have been thought to be one possible way of improving the therapeutic treatments for CMT. For example, MFN1 has been seen to be able to complement MFN2 in its functions in the cells. A cellular assay could be performed, where small molecules would be identified and used to upregulate MFN1 in neurons. This way MFN1 could complement MFN2 and restore the fusion of mitochondria.

Mutations in the gene KIF1B cause subtype CMT2A1. The kinesin superfamily is responsible for microtubule-dependent transport of a variety of organelles and vesicles. KIF1B knockout mice have been shown to die at birth from apnea due to nervous system defects. All CMT2A1 patients have had a loss-of-function mutation in the motor domain of the KIF1B gene. This mutation indicates that defects in axonal transport can underlie peripheral neuropathies.

It has been reported that mutations in the ras-related protein rab-7 (RAB7A) gene cause the axonal form of CMT known as CMT2B. RAB7A is a member of the Rab family of small G proteins, which regulate intracellular vesicle traffic. RAB7 protein plays a role in the endocytic pathway through which proteins are internalized, degraded and targeted to the lysosome for degradation. This pathway is shown to be increasingly important in CMT.
Mutations in small heat shock protein genes *HSPB1* and *HSPB8* are rare causes of CMT2 subtype CMT2F and CMT2L \(^{21, 22}\). In addition to CMT2L, mutations in *HSPB8* have been reported in distal hereditary motor neuropathy type 2 (dHMNII) \(^9\). Multiple pathogenic mechanisms are possible and being investigated for these disorders based in part on the multiple actions of the heat shock proteins (Hsp). These potential mechanisms include impaired chaperon activity, formation of aggregates, disruption of the cytoskeleton and axonal transport, disrupted interactions with other Hsp members, and predisposition to apoptosis and autophagy \(^{10}\).

Mutations in genes involved in protein synthesis and turnover are also known to cause autosomal dominant forms of CMT2. Such genes include *GARS, AARS* and *LRSAM1*. Alterations in the gene *GARS*, which codes for glycyl-tRNA synthetase are known to cause subtype CMT2D, which is characterized with major motor involvement. The symptoms include atrophy and weakness of the small muscles of the hand with late involvement of the distal lower limb muscles \(^{23}\). Mutations in the gene *AARS* are known to cause subtype CMT2N \(^{24}\) and alterations in *LRSAM1* CMT2 subtype CMT2P \(^{25}\). *AARS* codes for alanyl-tRNA synthetase and mutations in this gene have been shown to cause CMT2 with classical CMT phenotype but the frequency of this gene is yet to be determined \(^{18}\). *LRSAM1* codes for E3 ubiquitin-protein ligase and the mutations identified in several families have been shown to follow both autosomal dominant and recessive inheritance patterns \(^9\).

Mutations affecting *DYNC1H1*, which encodes the cytoplasmic dynein heavy chain 1 protein, a critical subunit of dynein1 responsible for axonal retrograde transport, lead to CMT2O. Three mouse models with different mutations in *DYNC1H1* have been reported with age-related progressive loss of muscle bulk and locomotor ability. Cytoplasmic dynein is a large multisubunit motor protein complex and has a key role in retrograde axonal transport in neurons. Again, the defects in axonal transport are shown to be able to underlie peripheral neuropathies \(^{26}\).

Mutations in the gene *TRPV4* have been described as causing three different motor predominant inherited neuropathies including CMT2C. *TRPV4*, coding for transient receptor potential cation channel subfamily V member 4 is involved in calcium homeostasis \(^{27}\). A wide range of phenotypes is seen in CMT2C patients. The mutations in *TRPV4* have been shown to lead to a motor predominant CMT2 \(^{18}\).

Alterations in gene *MPZ* are known to cause CMT2 subtype CMT2I/J. *MPZ* codes for myelin protein zero and it has been shown to be involved in myelin assembly \(^{28}\). Mutations in this gene also cause
CMT1B, which is the second most common form of CMT1 \(^9\). Also intermediate forms of CMT are caused by mutations in \(MPZ\) \(^8\).

Mutations in \(NEFL\) are the cause of CMT2 subtype CMT2E/1F. These mutations cause two different kind of phenotypes, type 1F resulting in slow NCVs \(^9\). \(NEFL\) encodes for a neurofilament light polypeptide (NF-L), which is a component of the axonal intermediate filament, the neurofilament. To date, approximately 20 mutations have been identified that span all functional domains of the protein. How these mutations cause axonal damage is not known. However, it is hypothesized that the mutations prevent the neurofilaments from assembling properly, resulting in the presence of abnormal aggregates that have been identified on nerve biopsies from both patients and in cultured cells \(^10\).

### 3.2.2 Disease Genes for Autosomal Recessive CMT2

AR-CMT2 is very rare compared to AD-CMT2. Three causative genes for AR-CMT2 have been identified; lamin A/C (\(LMNA\)), mediator complex subunit 25 (\(MED25\)) and ganglioside-induced differentiation-associated protein 1 (\(GDAP1\)).

Mutations in the gene \(MED25\) are shown to cause both CMT2B2 and AR-CMT2B. It codes for mediator complex subunit 25 and belongs to a family of large transcriptional coactivator complexes related to the yeast mediator. This gene is involved in gene expression regulation in humans but its exact physiological function remains obscure \(^9\). CMT2B2 patients have presented with a classical CMT phenotype with adult onset \(^18\). AR-CMT2B patients’ symptoms are often milder than in the AR-CMT2 forms caused by mutations in \(GDAP1\) or \(LMNA\) genes \(^13\).

\(GDAP1\) encodes for a mitochondrial fission factor that is anchored to the outer mitochondrial membrane. GDAP1 is involved in the fission and fusion of the mitochondria and it is also suspected to regulate mitochondrial function \(^30\). Mutations in the gene \(GDAP1\) cause various phenotypes and are also known to adapt many different inheritance patterns. These mutations underline CMT1 and CMT2 as well as AR-CMT1 and AR-CMT2 \(^31\), \(^32\). Vocal cord paralysis has often been seen with mutations in this gene. Intermediate conduction velocities in a CMT patient can also be one of the reasons to suspect a mutation in \(GDAP1\) \(^18\).

Alterations in gene \(LMNA\) are the cause of CMT2 subtype CMT2B1 and AR-CMT2A \(^33\). Mutations in this gene were also the first ones reported with a recessive inheritance pattern \(^33\). \(LMNA\) encodes for lamin A/C protein and as the gene \(NEFL\), it is involved in neurofilament homeostasis, which is closely
in contact with axonal transport. CMT2B1 patients have presented with various phenotypes and usually in the second decade with a severe CMT phenotype. Pure forms of CMT2B1 have been reported exclusively in families originating from Algeria and Morocco and these are owing to a founder mutation in *LMNA*, which is only seen in this restricted geographical area.\textsuperscript{13}

Altogether, CMT is caused by defects in genes, which are involved in Schwann cells function or neuronal homeostasis and therefore encode proteins that are involved in a diversity of molecular pathways such as protein synthesis, mitochondrial network regulation, organelle transport, endocytosis and maintenance of the cytoskeleton (Figure 2). A common function that links all the above mentioned together is axonal transport.\textsuperscript{34} The CMT2 disease mutations thought to affect axonal transport include mutations occurring in molecular motors (KIF1B, KIF1A and DYNC1H1), in molecules that facilitate proper mitochondria positioning or dynamics (MFN2 and GDAP1), in vesicle cargo (RAB7) or in motor domains responsible for cargo binding (DYNC1H1) and in proteins of the cytoskeleton or those responsible for its organization (NEFL, LMNA, HSPB1 and HSPB8). Mutations have also been found to play an indirect role in the regulation of movement through regulation of calcium homeostasis or Schwann cell-axon crosstalk (TRPV4 and MPZ). However, there are some genes that have not yet been related to axonal transport. Mutations in ARSs, in MED25 and in LMNA suggest that undiscovered functions of these genes may be of importance for axonal homeostasis and transport.\textsuperscript{34}
How the new disease genes of CMT2 fit this profile is another story. It can be suspected by the approach described above, that a candidate gene could have a function related to axonal transport. However, it is also possible that the new disease gene has a function, which does not fit into this category. This is why eyes should be kept open for genes with various functions when identifying new disease genes for CMT2.

3.3 Next-generation sequencing

Next-generation sequencing (NGS) technologies are based on new sequencing instruments, which are capable of producing millions of DNA sequence reads in a single run. These technologies have been available for research since 2008. There has been a clear shift from searching clinically relevant alleles from specific regions to identifying variants from a genome-wide sequencing data. NGS
Platforms harness massive parallel sequencing and are therefore capable of producing data with deep coverage in a short time scale. NGS technologies can be used from categorizing genetic diversity on a population level to identifying causal variants in a single individual. This individuality aspect of whole exome sequencing (WES) can ultimately lead to directed therapy.

The price of sequencing one’s whole genome or exome has gone down significantly, making NGS techniques available for researchers without large budgets. For example, in 2008 human genome was sequenced in 5 months for $1.5 million, but in 2011 the same accomplishment was done in a couple of days for approximately $10,000. The latter was made possible by the first commercial massively parallel pyrosequencing platform, which was launched in 2005. Keeping in mind that the first human genome sequenced through the Human Genome Project in 2001 took 13 years and approximately $2.7 billion, NGS techniques have truly brought long-awaited reduces to both the price and length of whole genome sequencing processes.

### 3.3.1 Whole Exome Sequencing

WES is a powerful, fast and cost-effective next-generation sequencing technique, which concentrates only on the protein-coding portion of genome. This portion accounts only for approximately 1% of the whole human genome but the mutations found in the exome part of genome cause 85% on Mendelian diseases\(^36\). WES is used as a diagnostic tool for finding causative mutations in Mendelian disorders and nowadays also in the cases of specific rare Mendelian disorders, characterized by phenotypic and genetic heterogeneity\(^37\). It is especially used for to diagnosing those Mendelian diseases, which present with unusual symptoms because they are difficult to confirm using clinical or laboratory criteria alone or require extensive or costly evaluation.

From a single instrument run in WES it is possible to generate enormous quantities of nucleotide sequence, depending on the platform. On average, exome sequencing identifies ~24,000 single nucleotide variants (SNVs) in African American samples and ~20,000 in European American samples in comparison with the reference genome sequence\(^38\). WES and other targeted sequencing approaches have increased sequence coverage of regions of interest, such as coding exons of genes. This lowers costs and enables higher throughput compared with random shotgun sequencing methods. WES is still the tool of choice for many researchers because it enables the screening of large numbers of patients in a robust fashion\(^39\).
3.3.2 Applying Whole Exome Sequencing to Mendelian Disorders

The molecular basis for approximately 3000 Mendelian diseases is currently known and over 4500 single nucleotide polymorphisms (SNPs) have been associated with a variety of human traits and intricate disorders. However, the gene remains unknown for over 3500 Mendelian disorders \(^{38}\). As already mentioned, a large number of Mendelian diseases are caused by mutations in the exome and this is one of the main reasons why WES is such a well-suited tool for discovering the genetic cause for Mendelian disorders. By new sequencing-based studies e.g. WES it is now possible to directly identify the causal sequence of a variant in a single experiment. Karyotyping, linkage analysis, homozygosity mapping and copy number variation (CNV) analysis have previously been widely used for the search of causative genes but are now being replaced by methods, which obtain the entire data for all genes and allow directly the test of association with disease \(^{40}\). WES has proved to be an attractive method for finding causative genes in Mendelian diseases when linkage analysis has been impossible or when looking for causal de novo mutations.

The detection rate when identifying disease mutations by WES in CMT cases have been as high as 32\%, which suggests that WES can act as a highly exact, rapid, and economical molecular diagnostic tool for CMT patients who are tested for major genetic causes \(^{41}\).

3.3.3 The Workflow of Whole Exome Sequencing

3.3.3.1 Technical steps

Methods usually used for selecting a subset of the genome for sequencing include solid-phase hybridization and liquid-phase hybridization. These are at the moment the only commercially available tools for selecting the entire human exome as the target for sequencing. Sequencing methods vary more and consist of several of NGS technologies, including reversible terminator reactions, pyrosequencing, sequencing by ligation and real-time sequencing. These methods are applied to the enriched target and create millions of short sequence copies, also called reads. These reads cover the portions of the reference genome that were targeted. Multiple algorithms are created for aligning the created sequences with the reference sequence. This resulting aligned sequence is next inspected for positions that vary from the human reference genome and are identified as SNPs.
As with the alignment tools, many algorithms have been developed for high-quality identification of variants in NGS projects. These SNP discovery tools include programs such as SAMtools and the Genome Analysis Toolkit Unified Genotyper. As already mentioned, typically 20,000 to 25,000 variants are detected per exome with the variation in this number occurring from the ancestry of individuals being sequenced. The number of detected variants can be even lower depending on the exome target definitions.

3.3.3.2 Filtering

Human genome contains a great deal of genetic variation, which makes the identification of disease mutations among the thousands of variants challenging. Moreover, even a single base-pair change can be associated with disease, which makes the distinguishing of true variation from sequencing errors even harder. This challenge is emphasized in WES analysis where the goal often is to discover rare causative variants causing disease. Also the amount of false-positive variants is thought to be quite high in WES, even though WES is able to compensate this error rate with its acute coverage of target areas. These facts are to be kept in mind when filtering the variants with the aim of finding causative genes.

Many filtering methods have been developed to narrow the hunt for the causal variant from 20,000 to often a single variant or to a single gene. These methods mostly rely on three defaults: the causal variant will have an effect on the protein coding sequence, it will be extremely rare, it will show complete penetrance (100% probability of observing a genotype given the phenotype).

Variants are divided into synonymous (ones that do not alter the amino acid sequence) and non-synonymous (protein-altering) ones. Mutations can also be classified into missense variants (ones that introduce an amino acid change) and nonsense (ones that prematurely truncate proteins) mutations. According to the literature approximately 50 to 75 % of variants can be removed by excluding all the synonymous variants from the WES data. Variants have been yet again divided into different classes on the basis of the predicted effects of the protein alterations by programs e.g. SIFT, GERP, PhyloP and PolyPhen. When assumed that variants responsible for Mendelian disorders are extremely rare and therefore are not present in the public databases of human genetic variation it has been justifiable to remove variants from further consideration if they are found in HapMap, 1000 Genomes Project or dbSNP, until recently. As the amount of genome data has increased rapidly, this approach should not be used anymore. Restricting the search to nonsynonymous variants not present in these databases usually lowers the number of assumed causative variants usually to around 200 to 500.
3.3.4 THE CHALLENGES OF WHOLE EXOME SEQUENCING

WES still possesses many technical failures that are likely to decrease over the next few years. At the moment these problems cause clear delay for example to the variant identification process of WES 38. For starters, the probes used in WES only target known exons. Related to this problem is the biggest limitation of WES; it only targets exons and leaves out regulators, miRNA:s and UCEs which could all potentially carry causative mutations. Another limitation to WES is that it is unable to detect structural variations, such as copy-number variations, inversions and translocations 39.

Uneven capture efficiency across exons causes problems when the depths (meaning here the amount of reads) of specific genes’ exons are not good enough for the sufficient confirmation of a possible causative variant. This in turn can increase the false positives risk of WES technique 37. Some sequences such as GC-rich sequence stretches can be difficult to capture and some sequences can not be targeted at all 37. Also, the off-target hybridization value is notable (20 %). There are also problems with the alignment of sequences in WES; not all sequences can be aligned with the reference sequence to allow base calling. Nevertheless WES has higher base calling rates than Sanger sequencing. This increases the coverage depth and diminishes the number of false calls.

WES creates lots of background, which consists mainly of non-pathogenic polymorphisms and sequencing errors. On top of that the whole process of WES generates huge amount of data, which then has to be stored and analyzed. Filtering variants has to be done carefully so that the possible causative genes are not lost completely 37.

Although the price of WES analysis is decreasing, the re-sequencing of mutant or variant genes using conventional sequencing techniques increases the cost of the approach. This step is frequently needed for the proper validation of the variants.

3.3.5 WHOLE EXOME SEQUENCING AS A DIAGNOSTIC TOOL

WES analysis can be used for identifying disease genes in disorders with all kinds of inheritance patterns. Here, the steps of identifying new causative mutations by WES are presented as well as some facts about the recent usage of WES in discovering mutations in known and new disease genes.
3.3.5.1 Identifying recessive and dominant variants

When searching for a recessively inherited causal variant by WES analysis, only variants present in a homozygous or compound heterozygous state can be further considered. WES analysis has been successful in many cases, when identifying recessively descending variants, partly because these variants are easily detectable. Identifying these variants is especially straightforward if the patient is a child of a consanguineous marriage. Then WES analysis is likely to identify a manageable number of rare homozygous variants and further studies can be performed to investigate the functional relevance of these variants. Sometimes however, additional novel methods incorporating techniques from population and statistical genetics will be needed to further narrow the search and finally successfully identify one causative gene.

In autosomal dominant disorders, each candidate gene must show at least one change per individual. For finding dominant variants, WES analysis is not always as straightforward and fast as for identifying recessively descending mutations. The genetic analysis of these variants poses greater technical challenges and certain experiments are needed in order to find the causative variants from the huge amount of WES data. These experiments can be designed to elucidate the inheritance pattern of the variants depending on the availability of patient cohorts and the effect of the disease on reproductive fitness. If large families with the disease under research are available, WES analysis should be performed for at least two affected family members. If common variants in these two patients are found in WES analysis, it can be suspected that one of these variants is disease causing. When large affected families are not available, one can perform WES analysis on a number of unrelated patients with similar clinical manifestations and select genes that are commonly mutated in the patient cohort. If this approach is selected, the mutation burden of the entire gene set has to be taken into account. This means, that large genes are more likely to contain many rare functional variants and because of this can appear to be most interesting. If the disease phenotype is severe enough such that it affects reproductive fitness, one can perform WES analysis for two unaffected parents or siblings of the patient in addition to the proband. This is done, because de novo mutation can be this way easily spotted among the other variants. Any variant, that is not present in the unaffected parent or sibling WES data but is present in the patient data, is probably a de novo mutation.
3.3.5.2 Identifying new disease genes by whole exome sequencing

When using WES as a diagnostic tool and identifying novel causative variants in diseases, some steps are required to be sure of the validity of the mutations. The most important one is surely the re-sequencing of the suspected causative variants. This basically means, that the validity of the mutation is verified by using Sanger sequencing method \(^{43},^{44},^{41}\). Also, the cosegregation in family members is studied by the same approach. With \textit{de novo} mutations, the segregation in the parents is not under research because this mutation is not present in unaffected members of the patient’s family. When identifying and validating a new disease causing gene mutation, additional patients identified with the same mutation can help making it more probable that the mutation is causative. Also making sure that this mutation is not present in a control sample further verifies the causative nature of the variant \(^{44}\). The more conserved the area where the new causative variant is situated is, the more probable that the mutations are disease causing and deleterious for the gene’s function. If additional patients with the mutation are found, control samples do not possess the mutation and the mutated gene fits by its function to the hypotheses of being disease gene, these facts can be enough for the gene to be listed as disease gene. In these cases, it is not always necessary to provide data about the effect of the mutation to the function of the protein \(^{44},^{45}\). However, if the mutation is found in only one patient, further functional analysis of the mutated protein is usually needed for the gene to be noted as a disease gene. Currently, laboratories using WES analysis for finding novel disease genes have reported success rates as high as 50 \(\%\) \(^{46}\).

3.3.5.3 Identifying mutations in known disease genes by whole exome sequencing

Some of the steps done while identifying mutations in known disease genes by WES are the same as when identifying new causative variants in new genes. The variants should be always re-sequenced before investigated further. Also the segregation in the family has to be verified to be correct. By using a list of known disease genes for the disorder under investigation and looking at only the variants in these genes, it can be quite fast to identify mutations in known disease genes by WES. If the variant is novel, but the gene is a known disease gene for the disorder under research, functional analysis may not be needed to verify the causative nature of the variant \(^{43}\). Still, control samples can be used to verify that the mutation is truly rare. WES has previously been successfully used in identifying mutations in known diseases genes also for CMT patients, both in identifying novel disease causing heterozygous mutations
in known disease genes for CMT and for seeking previously identified variants in known disease genes
4. AIMs OF THE STUDY

The aim of this study is to identify the molecular cause of disease for four patients suffering from early-onset axonal neuropathies using WES as an analysis tool. The second aim is to improve the process of analyzing variants from whole exome sequencing data. It is particularly important to develop the analyzing steps of insertion and deletion variants further.

In addition to these aims, one of the goals is to perform functional analysis of the identified potential disease causing mutations. This way the disease causing nature of these gene mutations can be further investigated and confirmed.
5. MATERIALS AND METHODS

5.1 PATIENTS

All patients were treated in the Department of Pediatric Neurology of Helsinki University Central Hospital by pediatric neurologist Tuula Lönnqvist. Patients’ approval for WES analysis was received before starting this study. This study was approved by the Helsinki University Central Hospital. The following permission was granted for this study: Perinnöllisen polyneuropatian ja spastisen parapareesin molekyyligeneettinen tausta Suomessa, Medisiinisen eettisen toimikunnan puolto 371/13/03/01/12. The study is also a part of study project TYH2013227 "Lapsuusäänneurodegeneraation geneettinen tausta".

Patients were selected to whole exome sequencing based on their interesting phenotype and unorthodox symptoms. All patients were diagnosed with axonal neuropathy and some specified to be suffering from CMT2. Here are listed all the symptoms of the patients as well as the examinations already done before the WES analysis.

5.1.1 PATIENT 1

Patient 1 is female who had first symptoms at the age of 7 months. Intrauterine and early developmental stages were normal. The patient had loss of deep tendon reflexes and hypotonia. ENMG showed sensory focused axonal neuropathy. Nerve biopsy showed marks of axonal neuropathy as well but muscle biopsy was normal. The patient is now 3 years old and has never walked by herself. She also suffers from severe visual impairment. The patient’s brain MRI findings were normal. She also has normal cognition and is the only affected individual in this family. The patient was diagnosed with autosomal recessive axonal neuropathy.

The most common causative gene mutations of CMT were checked by DNA testing. The genes PMP22, MPZ, MFN2 and FXN were sequenced with normal findings. A heterozygous c.593G>A variant in gene OPA1 was found. Patient’s parents were also sequenced and the same variant was found in her father. Mutated amino acid Arg198 is not conserved and the fifth exon, where the mutation is situated, does not code for a functional domain. These facts lead to the conclusion that the OPA1 variant is not disease causing.
5.1.2 Patient 2

Patient 2 is female and was referred to the pediatric department on account of muscle hypotonia at the age of four years. Intrauterine and early developmental stages were normal and the patient is also the only affected individual in the family. The patient has always been physically frail and her growth is one standard deviation below the mean. The patient also has weak musculature and skeleton. ENMG showed protracted conduction velocities and nerve biopsy showed axon loss. The patient has no visual impairment. Her cognition is normal as is brain MRI. She is able to walk with the help of orthosis. The patient was diagnosed with severe autosomal recessive CMT2.

DNA testing revealed no mutations in genes PMP22 and MPZ.

5.1.3 Patient 3

Patient 3 is male and had normal intrauterine and early development stages but was notably clumsy in childhood. He learned to walk at the age of 16 months. The patient is the only affected individual in his family. First ENMG findings suggested demyelination but later findings confirmed the defect to be axonal. The patient’s disease progress has been rapid during recent years and his condition has deteriorated. He has breathing difficulties as well as osteoporosis. The patient has no visual impairment and his cognition is normal. He was diagnosed with severe autosomal recessive CMT2.

DNA testing revealed no mutations in genes GJB1, LITAF, PMP22, MPZ and MFN2.

5.1.4 Patient 4

Patient 4 is male and was diagnosed with laryngomalacia after birth but had otherwise normal early development. The patient was referred to the pediatric department at the age of 4 years on account of slow motor development. ENMG findings showed signs of axonal neuropathy. Nerve biopsy as well as muscle biopsy confirmed these findings. The patient has developed scoliosis and also suffers from learning disabilities. His brain MRI showed abnormalities in the white matter. The patient does not suffer from visual impairment. He is diagnosed with axonal neuropathy and is currently in a ventilator. The patient’s two siblings are also affected with the same disease but their parents are healthy.
suggesting autosomal recessive inheritance. DNA testing revealed no mutations in genes *PMP22, MPZ, MFN2* and *POLG*.

5.1.5 *THE COLLECTION OF PATIENT SAMPLES*

DNA samples from the four patients were collected for WES analysis. Collection of the samples was done by pediatric neurologist Tuula Lönnqvist, who contacted the patients and arranged the sample collection.

5.2 *WHOLE EXOME SEQUENCING*

5.2.1 *WHOLE EXOME SEQUENCING*

Whole exome sequencing was performed for four patients suffering from early-onset axonal neuropathies as a service at the Finnish Institute of Molecular Medicine (FIMM) in Helsinki. Exome target capture was done using the NimbleGen v2 kit and sequencing was done with the Illumina HiSeq2000 platform. The variant calling pipeline of FIMM was used to align 2x82 paired-end reads to the *hg19* assembly. The sequence capture protocol ([http://www.nimblegen.com/products/seqcap/ez/v2](http://www.nimblegen.com/products/seqcap/ez/v2)) of NimbleGen v2 kit starts with a sample of genomic DNA. An oligo pool is constructed against target regions in the genome and simultaneously a standard shot-gun sequencing library is made. The sequencing library is then hybridized to the oligo pool. Streptavidin beads are used to pull down the complex of capture oligos and genomic DNA fragments. After this, the unbound fragments are removed by washing followed by an amplification of the enriched fragment pool by PCR. The success of enrichment is measured by real-time polymerase chain reaction (qPCR) at control loci. Finally, a sequencing library enriched for target regions is ready for high throughput sequencing. The sequencing is then followed by mapping, alignment and variant calling. As mentioned before, all the steps described here were included in the service done at FIMM.
5.2.2 *ANALYZING TOOLS OF WES DATA*

The analysis of WES data was done by utilizing the Microsoft Office Excel program, dbSNP database, Ensembl Genome browser (www.ensembl.org), Integrative Genomics Viewer (IGV) and the SIFT genome tool (http://sift.jcvi.org/). The SIFT Genome-tool predicts the effect of an amino acid substitution to the protein’s function by taking account the conservation of amino acid residues in sequence alignments derived from closely related sequences. The SIFT Genome-tool also includes only the variants that are situated in the coding area of genes. By the IGV program it is possible to view the entire data collected from the patient and examine the variants by browsing through the exome.

5.2.2.1 *Homozygous SNVs*

The analysis of homozygous variants began by excluding known single nucleotide polymorphisms (SNPs) from the WES data with the help of dbSNP database. Then the non-genic variants were filtered out and the homozygous variants selected from the WES data. The variants left after these analyzing steps were entered to the SIFT Genome-tool. This step made it possible to classify the variants into the groups of non-damaging, damaging and stop codon variants (nonsynonymous and synonymous). Only the coding and nonsynonymous variants were taken into account in the next step. Finally the variants, which had either a damaging or a stop codon prediction, were selected. These account for the missense or nonsense mutations respectively. The analyzing steps are illustrated in Figure 3.

![Figure 3. The steps of analyzing homozygous SNVs.](image-url)
5.2.2.2 Compound heterozygous SNVs

The analysis of the compound heterozygous variants was begun by excluding the homozygous variants from the entire WES data of the patients. This was followed by the exclusion of known SNPs (dbSNP database) and non-genic variants. Again the variants that were left after these steps were entered to the SIFT Genome tool and the nonsynonymous variants with either damaging or stop codon prediction were selected. Eventually only genes with two heterozygous variants in them were taken into account. The steps are shown in Figure 4.

![Diagram of analysis steps](image)

Figure 4. The steps of analyzing heterozygous SNVs.

5.2.2.3 Homozygous deletions and insertions

In the beginning of the analysis of homozygous indel variants all known SNPs and non-genic indel variants were excluded followed by the selection of homozygous indel variants. This was done by using a less than or equal to 0.15- filter in the call ratio of indel variants. Indel variants which coverage depth was 10 or more were included in the analysis. Next the indel variants which variant calls were less than 10 were excluded. These steps were done to decrease the amount of all indel variants and especially the number false positive indel variants. Also the repeat indels were excluded.
The selection of genes that encode for a protein was done using the Ensembl Genome browser. The last step of this analyzing process was to confirm that the found variants were in the exons of the genes. This step restrained the variants from approximately 75 left variants to ~5 variants that actually were situated in the coding region. This verification was done using the IGV. All the analyzing steps are presented in Figure 5.

1. Exclusion of known SNPs (dbSNP)  
2. Filtering out the non-genic indel variants  
3. Selection of homozygous indel variants  
4. Exclusion of indel variants with depth less than 10  
5. Exclusion of indel variants with variant calls less than 10  
6. Exclusion of repeat indel variants  
7. Selecting protein coding genes  
8. Selection of indel variants which are in the coding area

Figure 5. The steps of analyzing homozygous deletions and insertions.

### 5.2.2.4 Compound heterozygous deletions and insertions

The analysis of compound heterozygous indel variants started with the exclusion of known SNPs and non-genic indel variants. Next the heterozygous variants were selected by using a filter in the call ratio of variants and defining it to be between 0.2 and 0.8. Indel variants which reference calls were one were excluded to decrease the noise. Next only genes that had two indel variants in them were selected. These variants were screened for being in the coding area by the IGV. All steps are illustrated in Figure 6.
5.2.2.5 Heterozygous deletions and insertion variants combined with heterozygous SNVs

The analysis of heterozygous indel variants and the combination of those indel variants with the heterozygous SNVs started with the exclusion of known SNPs and non-genic indel variants. Next the heterozygous indel variants were selected as in the previous section and the indel variants with reference calls one were discarded respectively.

The combining of heterozygous indel variants with heterozygous variants with a damaging or a stop codon prediction produced a great amount of variants. Again the selection of indel variants which were in the coding are (done by the IGV) resulted in a much more limited number of variants. The analyzing steps are shown in Figure 7.
5.3 THE CLASSIFICATION OF INTERESTING VARIANTS

Important steps in the classification the interesting variants- procedure include investigating tissue express patterns of the identified candidate genes and taking into account the function of the genes and their correlation to the phenotype of the patients. This was done by browsing through the information of interesting genes using the NCBI website (http://www.ncbi.nlm.nih.gov/). Expression patterns of the candidate genes were analyzed from the Unigene database (http://www.ncbi.nlm.nih.gov/UniGene). Existing data from the genes was investigated in the case of associations with diseases that could eliminate some of the interesting genes from being the disease gene for early-onset neuropathies. This data was also found from the NCBI website.

Anu Wartiovaara’s research group’s exome sequencing variant database was used to exclude some of the interesting variants. This database is built from the whole exome sequencing data of almost 50 patients suspected to have a mitochondrial disease. Most of the patients have a Finnish origin. Using a specific variant query-tool of the database I verified if some of our interesting variants could be found from these patients. If a positive hit was found, this gene was excluded from being a possible disease gene of autosomal recessive early-onset neuropathy because it is highly unlikely that a causative variant would be present in the patients of Anu Wartiovaara.
5.3.1 THE SCORING SYSTEM

After these steps had been done for the interesting variants, the most promising genes were selected for Sanger sequencing with the goal of verifying the authenticity of the mutations in these genes. This was done by creating and using a special scoring system for the classification of the numerous different variants. The scoring system was based on giving the variants a score from one to five depending on the characters the gene possessed.

- Score one was given to variants with strong gene expression in a wrong tissue or no gene expression in nerve and without any interesting function concerning neuropathies.
- Score two was given to variants that were in a previously known disease causing gene and with no or small gene expression data. The function of these variants did not rule out the chance of this gene to be causing a neuropathy.
- Score three was given to variants possessing a slightly interesting function and with little gene expression or with gene expression in all tissues.
- Score four was given to variants with quite strong nerve or brain gene expression or with gene expression in all tissues combined with strong nerve or brain expression. The function of these variants had to be interesting regarding neuropathies for them to obtain this score.
- Score five was given to variants that could most likely be disease genes for a neuropathy. Their function had to be very well suited or linked to neuropathies and gene expression strong in nerve or brain. Also the amino acid conservation of the mutated area was taken into account.

5.4 SANGER SEQUENCING

Sanger sequencing confirmed the authenticity of five possible candidate variants. Re-sequencing the possible causative variants identified by WES is needed due to the large amount of false positive variants found by this approach.

5.4.1 POLYMERASE CHAIN REACTION (PCR) PROTOCOL

PCR primers were designed for Sanger sequencing of the mutations. First the genomic sequence and mRNA of the genes SYNM, GTF2IRD2, TRIM2, BSN and PLEC were searched from USCS website
Open reading frames were identified from the mRNA by the help of ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). This enabled finding the right open reading frame where the translation of the gene starts. The coding exons of the genes listed above and the variant nucleotides were next identified from the genomic sequences. Primers were designed by Primer 3 Input version 0.4.0 by entering the exon parts and some intronic sequence around them to the program. The size of the selected nucleotide sequence was chosen by the size of the exon.

For the PCR of the variants in genes SYN M, GTF2IRD2, TRIM2 and BSN the PCR reaction consisted of 25 ng of genomic DNA from the patients, 2.5 µl of 10 x DynaZyme buffer (Finnzymes), 1 µl of 10 mM dNTPs (2 mM dATP, 2 mM dCTP, 2 mM dGTP, and 2 mM dTTP) (Bioline), 0.5 µl of 10 µM primers 1 to 16 (Table 2, Oligomer), 2 µl of 1.00 M magnesium sulfate (MgSO₄, SIGMA) and 0.6 U of DynaZyme 2 polymerase (Finnzymes) in a total volume of 25 µl. The amplification cycles were: initial denaturation at 95°C for 4 min, followed by 32 cycles of 95°C for 30 sec; 54°C (SYNM)/58°C (GTF2IRD2)/56°C (TRIM2)/62°C (BSN); 72°C for 1 min and final extension in 72°C for 10 min.

For the sequencing of the variants in the gene PLEC the PCR reaction consisted 50 ng of genomic DNA, 5 µl of 10 x DynaZyme buffer, 5 µl of 2 µM dNTPs (2 mM dATP, 2 mM dCTP, 2 mM dGTP, and 2 mM dTTP)(Bioline), 5 µl of 10µM primers 17 to 20 (Table 2, Oligomer) and 0.6 U of DynaZyme 2 polymerase in a total volume of 50 µl The amplification cycles were: initial denaturation at 94°C for 2 min, followed by 9 cycles of 94°C for 1 min; 59°C (ex22)/ 62°C (ex 32); 72°C for 2 min and 24 cycles of 94°C for 1 min; 59°C / 62°C for 1min; 72°C for 2 min (2 additional sec added to this step in each cycle) and final extension in 72°C for 10 min.

To analyze the success of the PCR reaction 5 µl of the PCR product was run with 6 x LB loading dye on a 1.2 % agarose gel. Ethidium bromide was added to the gel to the final concentration of 0.5 µg/ml. The samples were run alongside of a GeneRuler 100 bp DNA ladder (Fermentas) in a 0.5 x TBE buffer and the PCR products visualized and photographed by molecular imager ChemiDoc XRS+ (BioRad).

5.4.2 SEQUENCING

The PCR products were first purified from the amplification primers with Exonuclease 1 (EXO, Fermentas) and Shrimp Alkaline Phosphatase (SAP, Fermentas) enzymes. The EXOSAP reaction consisted of 5 µl of PCR product, 10 U of SAP and 1 U of EXO in a total volume of 10 µl. The EXOSAP conditions were: 37°C for 1 hour and 72°C for 15 min.
The PCR products were sequenced with a Big Dye Terminator v3.1 Sequencing kit (Applied Biosystems). The sequencing reaction consisted of 4 µl of EXOSAP PCR product, 1.8µl of 5 x BigDyl 3.1 buffer, 0.33 µl of 10 µM primers 1 to 20 (Table 2, Oligomer) and 0.35 µl BigDyl 3.1 polymerase in a total volume of 10 µl. The reaction conditions were: initial denaturation at 96° C for 1 min followed by 29 cycles of 96° C for 10 sec; 50° C for 5 sec; 60° C for 4 min.

The sequencing reactions were analyzed by the capillary method in FIMM in Helsinki.

Table 2. Primers used in Sanger sequencing of the possible causative gene variants.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Name</th>
<th>Target</th>
<th>5’→3’ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BSN_ex5_F</td>
<td>BSN genomic sequencing</td>
<td>CCTGAGCTGGAGATGGAGAG</td>
</tr>
<tr>
<td>2</td>
<td>BSN_ex5_R</td>
<td>BSN genomic sequencing</td>
<td>CAGAGTCAAGGGAGGGTGAG</td>
</tr>
<tr>
<td>3</td>
<td>BSN_ex6_F</td>
<td>BSN genomic sequencing</td>
<td>CCAAAGCTCTGGGAAAAGAGA</td>
</tr>
<tr>
<td>4</td>
<td>BSN_ex6_R</td>
<td>BSN genomic sequencing</td>
<td>TTCTCCTCCTCATCCAGCTC</td>
</tr>
<tr>
<td>5</td>
<td>BSN_ex2_F</td>
<td>BSN genomic sequencing</td>
<td>CCTGGGTAAACCAGAGAGCAG</td>
</tr>
<tr>
<td>6</td>
<td>BSN_ex2_R</td>
<td>BSN genomic sequencing</td>
<td>ACACAAATGCTGGTGCTCAA</td>
</tr>
<tr>
<td>7</td>
<td>GTF2IRD2_ex2_F</td>
<td>GTF2IRD2 genomic sequencing</td>
<td>CAGACTCTGAATTTCGGCCACTT</td>
</tr>
<tr>
<td>8</td>
<td>GTF2IRD2_ex2_R</td>
<td>GTF2IRD2 genomic sequencing</td>
<td>TGTTTATGTATCTGTCTATTTG</td>
</tr>
<tr>
<td>9</td>
<td>TRIM2_ex8_F</td>
<td>TRIM2 genomic sequencing</td>
<td>AGACTTCAGATACATGGTGTAAGGA</td>
</tr>
<tr>
<td>10</td>
<td>TRIM2_ex8_R</td>
<td>TRIM2 genomic sequencing</td>
<td>TTATCAGTTACTTCTACTTCACCAAG</td>
</tr>
<tr>
<td>11</td>
<td>TRIM2_ex5_F</td>
<td>TRIM2 genomic sequencing</td>
<td>GATGATGAAGGGCATAGAATTTT</td>
</tr>
<tr>
<td>12</td>
<td>TRIM2_ex5_R</td>
<td>TRIM2 genomic sequencing</td>
<td>TTCAACCTTGCTCTTTGCA</td>
</tr>
<tr>
<td>13</td>
<td>SYNM_ex4_F</td>
<td>SYNM genomic sequencing</td>
<td>AGCCCCGTAAGTTATGTCAGC</td>
</tr>
<tr>
<td>14</td>
<td>SYNM_ex4_R</td>
<td>SYNM genomic sequencing</td>
<td>GCATGAATTCACCGCTTCT</td>
</tr>
<tr>
<td>15</td>
<td>SYNM_ex5_F</td>
<td>SYNM genomic sequencing</td>
<td>AGCTCTCGGTGTGTCGAGACC</td>
</tr>
<tr>
<td>16</td>
<td>SYNM_ex5_R</td>
<td>SYNM genomic sequencing</td>
<td>GGAGAGAAGACTCCAGGGAAA</td>
</tr>
<tr>
<td>17</td>
<td>PLEC_ex22_F</td>
<td>PLEC genomic sequencing</td>
<td>CTCCGTGTGCTCCTGTCG</td>
</tr>
<tr>
<td>18</td>
<td>PLEC_ex22_R</td>
<td>PLEC genomic sequencing</td>
<td>CCTCCCTCCACCCAGAGAG</td>
</tr>
<tr>
<td>19</td>
<td>PLEC_ex32_F</td>
<td>PLEC genomic sequencing</td>
<td>GCCTGGTGCTACCTCTATGG</td>
</tr>
<tr>
<td>20</td>
<td>PLEC_ex32_R</td>
<td>PLEC genomic sequencing</td>
<td>CAGCTCGGTGTAAGCTGAGG</td>
</tr>
</tbody>
</table>
5.5 THE CARRIER FREQUENCY OF TRIM2 BY MINISEQUENCING

The carrier frequency of TRIM2 variants was studied by the minisequencing method from 185 control DNA samples (370 chromosomes). The carrier frequency analysis was done to analyze the frequencies of the variants in the general Finnish population. The minisequencing method is based on PCR amplification done with one biotinylated and one non-modified primer followed by affinity-capture of the biotinylated PCR product on a streptavidin-coated microtiter plate. The variation in the nucleotide sequence is detected by a primer extension reaction with labeled nucleotides. DNA polymerase with a single labeled nucleotide complementary to the nucleotide at the variable site is utilized: DNA polymerase extends a detection step primer that anneals immediately next to the nucleotide that is analyzed (Figure 8). The amount of the incorporated label is then calculated and it indicates which nucleotide is present at the variable site.

Minisequencing method is especially useful when detecting a large number of samples because it is performed on a microtiter plate with simple manipulations and the result is an interpretive numeric value.

The control DNA samples had been collected from excess blood samples of individuals who had been checked for lactose intolerance by gene testing of the LCT gene. The control DNA samples were diluted into appropriate concentration for PCR (25 ng/µl) into 96-well plates.

5.5.1 MINISEQUENCING PROTOCOL

5.5.1.1 PCR protocol

Human TRIM2 genomic sequence was taken from the UCSC website. Biotinylated primers were designed according to the instructions given. The 20 nucleotides long detection primers were designed to anneal directly next to the nucleotide to be analyzed. For the missense mutation c.761T>A the detection primer was designed as a forward primer and for the c.1779delA the detection primer was designed as a reverse primer.

PCR was performed from 185 control DNA samples with one biotinylated primer (primer 21 for the missense mutation and primer 23 for the deletion) and one non-modified primer (primer 11 for the missense mutation and primer 10 for the deletion). PCR reaction consisted from 25 ng of DNA, 5 µl of
10 x DynaZyme buffer (Finnzymes), 0.6 U of DynaZyme 2 polymerase (Finnzymes), 1 µl of 10 mM dNTPs (2 mM dATP, 2 mM dCTP, 2 mM dGTP, and 2 mM dTTP) (Bioline), 5 µl of 2 µM biotinylated primer (10 pmol) (Table 3, Metabion) and 5 µl of 10 µM non-modified primer (50 pmol) (Table 3, Oligomer) in a total volume of 50 µl. The amplification cycles were: initial denaturation at 95° C for 1 min, followed by 30 cycles of 95° C for 30 sec; 56° C for 30 sec; 72° C for 1 min and final extension at 72° C for 3 min (Figure 8 part 1).

The success of the PCR reaction was analyzed as described above.

5.5.1.2 Minisequencing steps

First 10 µl of PCR products was pipeted to a microtiter plate with streptavidin-coated wells (Thermolab). 40 µl of PBS + 0.1%Tween (20 mM sodium phosphate buffer, pH 7.5, and 0.1% (v/v) Tween-20) was then added to each well. The compound was mixed gently. Plate was covered with removable tape and incubated in a shaker 37° C for 1.5 hours (Figure 8 part 2).

After incubation plates were washed with MiniTent (40 mM Tris-HCl, pH 8.8, 1 mM EDTA, 50 mM NaCl, and 0.1% (v/v) Tween-20) (200 µl/well) 3 times. Plates were properly dried by tapping into a papertowel. PCR fragments were denaturated with 50 mM NaOH (500 µ 5 M NaOH+ 45.5 ml aqua) (100 µl/well). This mix was incubated for 5 min in RT followed by same washes as in the previous step (Figure 8 part 3).

50 µl of the detection mix was added to the wells. This detection mix consists of 0.8 U of DynaZyme 2 polymerase (Finnzymes), 5 µl of DynaZyme buffer (Finnzymes), 0.1 µl of [3H]-labeled dNTPs (Adenine and Thymine, GE Healthcare) and 1 µl of 10 µM detection primer (10 pmol) (primer 22 for the missense mutation and primer 24 for the deletion, Table 3, Oligomer) in a total volume of 50 µl. The detection mix was incubated in the wells for 10 min 50 °C followed by washes (Figure 8 part 4).

60 µl of 50 mM NaOH was added to the wells and incubated for 5 min in RT. Next the liquid was transferred to scintillation tubes. 3 ml of scintillation liquid (OptiPhase) was added to the tubes. The tubes were properly mixed. The activity of the tubes was measured with 2450 MicroBeta² LumiJet device (PerkinElmer) and the result calculated (Figure 8 parts 5 and 6).
Table 3. Primers used for minisequencing, detection primers 22 and 24 and biotinylated primers 21 and 23.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Name</th>
<th>Target</th>
<th>5’→ 3’ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>TRIM2_ex_5_R_bio</td>
<td>TRIM2 minisequencing</td>
<td>TTCAACCTTGCTCTTGTCAG</td>
</tr>
<tr>
<td>22</td>
<td>TRIM2_ex_5_det</td>
<td>TRIM2 minisequencing</td>
<td>TGTGCTGTATTATGGAATTGG</td>
</tr>
<tr>
<td>23</td>
<td>TRIM2_ex_8_F_bio</td>
<td>TRIM2 minisequencing</td>
<td>AGACTTCAGATACATGGTCTAGGAA</td>
</tr>
<tr>
<td>24</td>
<td>TRIM2_ex8_det</td>
<td>TRIM2 minisequencing</td>
<td>ATTAGTAGTTAATCTTACCT</td>
</tr>
</tbody>
</table>
Figure 8. Minisequencing steps: 1. Amplification by PCR done with one biotinylated and one non-modified primer 2. Affinity-capture of the biotinylated PCR product on a streptavidin-coated microtiter plate 3. Primer extension reaction with labeled nucleotides 4. and 5. Extending a detection step primer that anneals immediately next to the nucleotide that is analyzed by DNA polymerase 6. Calculating the amount of the incorporated label. Attached from Suomalainen et al 48.

5.6 TRIM2 SITE-DIRECTED MUTAGENESIS

Expression constructs with the c.761T>A and the c.1779delA variants of TRIM2 cDNA were made for modeling of their effect on the function of the TRIM2 protein. These constructs were created by site-directed mutagenesis. Mouse wild-type Trim2 cDNA in myc-tagged expression vector was a kind gift.
from Dr Martin Balastik (49). Primers containing the c.761T>A or the c.1779delA variant were used to construct the mutant cDNAs into the plasmid holding the wild type mTrim2 (Table 4).

5.6.1 PCR PROTOCOL

Mouse Trim 2 genomic sequence was taken from the NCBI website (http://www.ncbi.nlm.nih.gov/). 29 nucleotides long mutagenesis primers contained the desired variants c.761T>A and c.1779delA of the cDNA.

A gradient PCR was made with the PCS2- plasmid containing the wild-type mTrim (Balastik et al 49). The PCR reaction consisted of 5 ng of the plasmid, 5 µl of 5 x Phusion buffer GC, 0.5 µl of dNTPs (2 mM dATP, 2 mM dCTP, 2 mM dGTP, and 2 mM dTTP) (Bioline), 0.5 µl of primers 25 to 28 (Table 4, Oligomer) and 0.6 U of Phusion Hot Start High Fidelity DNA Polymerase (Thermo Scientific) in a total volume of 25 µl. The amplification cycles for the 6.6 kb plasmid were: initial denaturation at 98° C for 30 sec, followed by 30 cycles of 98° C for 10 sec; 58°-65° C for 30 sec; 72° C for 40 sec and final extension in 72° C for 10 min. To analyze the success of the PCR reaction 25 µl of the PCR product was run with 6 x LB loading dye on a 0.8 % agarose gel. Ethidium bromide was added to the gel to the final concentration 0.5 µg/ml. The samples were run alongside of a GeneRuler 1 kb DNA ladder (Fermentas) in a 0.5 x TBE buffer and the PCR products visualized and photographed under an UV-light.

5.6.2 TRANSFORMATION OF PLASMIDS INTO E.COLI CELLS

Successful PCR products for the mTrim2 c.761T>A and the c.1779delA variants were cut from the agarose gel with a disposable scalpel (Swann-Murton) and weight in an eppendorf tube. The PCR product was extracted from the agarose gel by NucleoSpin Extract II kit (Macherey-Nagel) and eluted into 15 µl of elution buffer provided by the kit. PCR products were treated with DpnI restriction enzyme (BioLabs), which cuts all methylated DNA (the template plasmid) from the PCR reaction. The DpnI reaction consisted of 1 µl of DpnI, 15 µl of purified PCR product and 1.6 µl of 10x FastDigest buffer (Thermo Scientific). The mix was kept in +37°C for 1 hour.

Next the mTrim2 mutagenesis c.761T>A and c.1779delA constructs were transformed into competent E.coli cells. 4 µl of constructs were added to 50 µl of competent E.coli cells on ice and kept there for 30 min. Heat-shock reaction was performed at 42° C for 30 sec. The bacteria cells were kept on ice for 2
min after which 250 µl of room temperature SOC media was added to the cells. The bacteria cells were kept at +37°C for 1 hour at 220 rpm in a shaker. The bacteria cells were plated on an ampicillin resistant plate and kept at +37°C overnight.

### 5.6.3 Miniprep Plasmid Isolation and Digestion

5 ml miniprep cultures were produced from 10 colonies for both of the constructs. 5 µl of ampicillin was added to 5 ml of LB growth medium. The bacterial cultures were kept at +37°C at 220 rpm in a shaker overnight. Also ampicillin resistant control plates were made from the colonies used in the 5 ml bacterial cultures and kept at +37°C overnight. The bacterial cultures were centrifuged at 2500 rpm for 10 min at +4°C. The plasmid DNA’s were extracted from the cultures with GeneElute HP Plasmid miniprep kit (SIGMA) at the following day. Plasmid DNA concentration was measured with NanoDrop Spectrophotometer ND-1000 (Thermo Scientific). Both the mTrim2 c.1779delA and the c.761T>A constructs were digested with XhoI and XbaI Fast Digest-restriction enzymes (Thermo Sciences). The digestion reaction consisted of 2 µl of the plasmid, 2 µl of 10 x FastDigest Green Buffer and 1 µ of the XhoI and XbaI in a total volume of 20 µl. The digestion reactions were kept at +37°C for 1 hour. The digestions were next run on a 0.8 % agarose gel in the FastDigest Green Buffer. Ethidium bromide was added to the gel to the final concentration of 0.5 µg/ml. The samples were run alongside of a GeneRuler 1 kb DNA ladder (Fermentas) in a 0.5 x TBE buffer and the PCR products visualized and photographed by molecular imager ChemiDoc XRS+ (BioRad).

### 5.6.4 Sequencing of the Constructs

The mTrim2 mutagenesis c.761T>A constructs were sequenced with primer 31 and c.1779delA constructs with primer 33 to confirm that the mutagenesis had worked. The Trim2 constructs were sequenced with Big Dye Terminator v3.1 Sequencing kit (Applied Biosystems). The sequencing reaction consisted of 300 ng (max 2 µl) of plasmid, 1.65 µl of 5x Big Dyl 3.1 buffer, 0.65 µl of 10 µM primers 29 to 34 (Table 4, Oligomer) and 0.7 µl Big Dye 3.1 polymerase in a total volume of 10 µl. The reaction conditions were: initial denaturation at 96°C for 1 min followed by 29 cycles of 96°C for 10 sec; 50°C for 5 sec; 60°C for 4 min.
Table 4. Primers used for mTrim2 mutagenesis.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Name</th>
<th>Target</th>
<th>5´ → 3´ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>mTrim2_E254V_F</td>
<td>mTrim2 mutagenesis</td>
<td>GCTCATGGAGTTGGTGTTGCAAATATGGTC</td>
</tr>
<tr>
<td>26</td>
<td>mTrim2_E254V_R</td>
<td>mTrim2 mutagenesis</td>
<td>GACCATAAGTTGACCACAACTCCATGAGC</td>
</tr>
<tr>
<td>27</td>
<td>mTrim2_del_F</td>
<td>mTrim2 mutagenesis</td>
<td>AATGATGGGAAGTTTACGACAAAAATTGGAT</td>
</tr>
<tr>
<td>28</td>
<td>mTrim2_del_R</td>
<td>mTrim2 mutagenesis</td>
<td>ATCCAATTTTTGTCTAATTCCATCATATT</td>
</tr>
</tbody>
</table>

Primers were designed for sequencing the whole mTrim2 insert from the plasmids (Table 5, Oligomer). 20 nucleotides long primers were designed at 300 bases apart from each other to sequence the whole mTrim cDNA.

The whole insert was sequenced from 4 of the mTrim2 mutagenesis c.761T>A and the c.1779delA mutagenesis constructs as described above.

Table 5. Primers used for sequencing the mTrim2 inserts from the plasmids.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Name</th>
<th>Target</th>
<th>5´ → 3´ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>mTrim2_sekvR</td>
<td>mTrim2 plasmid sequencing</td>
<td>ATGGTTGAGGCCAGAAAGAG</td>
</tr>
<tr>
<td>30</td>
<td>mTrim2_sekvF1</td>
<td>mTrim2 plasmid sequencing</td>
<td>GTGTGCGCCAGACATCCAT</td>
</tr>
<tr>
<td>31</td>
<td>mTrim2_sekvF2</td>
<td>mTrim2 plasmid sequencing</td>
<td>GATGACATCCACTCCACCTT</td>
</tr>
<tr>
<td>32</td>
<td>mTrim2_sekvF3</td>
<td>mTrim2 plasmid sequencing</td>
<td>CAGCCCATGTCCGTTACTAT</td>
</tr>
<tr>
<td>33</td>
<td>mTrim2_sekvF4</td>
<td>mTrim2 plasmid sequencing</td>
<td>AACAAAGGGAATTTACAAAA</td>
</tr>
<tr>
<td>34</td>
<td>mTrim2_sekvF5</td>
<td>mTrim2 plasmid sequencing</td>
<td>TTTGAGGTCCCCACTTTTC</td>
</tr>
</tbody>
</table>
5.6.5 MIDIPREP PLASMID ISOLATION AND GLYCEROL STOCKS

Midiprep cultures, in the volume of 50 ml, were produced from the mTrim2 mutagenesis constructs sequenced earlier. The constructs were confirmed to hold in them the complete mTrim2 cDNA sequence with the correct mutations before producing the midiprep cultures. Specific colonies were selected from the control plate made at the same time as the 5 ml bacterial cultures. Also 50 ml bacterial cultures were prepared from these colonies for the plasmid isolation. Again 50 µl of ampicillin was added to 50 ml LB growth medium. A bacterial culture in the volume of 3 ml was grown from the selected colony for approximately 7 hours at +37° C and after that 1 ml of the bacterial culture was added to 50 ml of LB growth medium and kept at +37°C overnight. The following day the bacterial culture was centrifuged at 6000 rpm for 10 min at +4° C. The plasmid DNA was extracted from the culture with NucleoBond Xtra Midi kit (Machery-Nagel) and eluted into 500 µl of sterile MQ water. The plasmid DNA concentration was measured with NanoDrop Spectrophotometer ND-1000 (Thermo Scientific).

Glycerol stocks were made at the same time from the mTrim2 mutagenesis constructs as the 50 ml bacterial culture. LB growth medium and 100 % glycerol was mixed 1:1. 1 ml of the 3 ml mini culture was added to 450 µl of the LB-glycerol mix. The glycerol stocks were stored at -80 ° C.

5.7 TRIM2 WESTERN BLOTTING

Three mTrim2 constructs presented in Table 6 were received from Dr Martin Balastik and used in western blotting. Western blotting was done to ensure that the constructs were stable and working correctly.

Table 6. Constructs received from Dr Balastik.

<table>
<thead>
<tr>
<th>The name of the construct</th>
<th>The vector used in the cloning</th>
<th>The restriction enzyme sites used in the cloning</th>
<th>The changes made in the structure of mTrim2</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTrim2</td>
<td>pCS2-myc</td>
<td>XhoI + XbaI</td>
<td>Full length mTrim2</td>
</tr>
<tr>
<td>mRBCC-Trim2</td>
<td>pCS2-myc</td>
<td>BamHI</td>
<td>Missing the RBCC domain of mTrim2</td>
</tr>
<tr>
<td>mNHL-Trim2</td>
<td>pCS2-myc</td>
<td>XhoI + SacII</td>
<td>C-terminally truncated</td>
</tr>
</tbody>
</table>
5.7.1 Transfection of the Constructs

HEK293 cells were transfected with the wild-type mTrim2, mRBCC-Trim2 or mNHL-Trim2 plasmid. The control sample was transfected without DNA.

HEK293 cells were cultured in 10 ml of DMEM F12 cell culture media (Thermo Scientific) in a 10 cm cell culture plate. A volume of 5 ml of Pen-Strep (10 000 U penicillin and 10 000 U streptomycin, Lonza), 5 ml of L- Glutamax (Lonza) and 50 ml of inactivated Fetal Bovine Serum (Gibco) were added to a full bottle of DMEM (500 ml, Lonza). Dividing the cells was done by first washing the cells with 1x PBS followed by detaching the cells from the cell culture plate with of 5x Trypsin EDTA (5 ml of 10x Trypsin EDTA, 45 ml of 1x PBS).

When the plates of the cells were 60-80 % confluent they were transfected using the JetPrime transfection kit (Polyplus) with wild type mTrim2 construct, with mRBCC-Trim2 construct or with mNHL-Trim2 construct (2 µg of DNA). Once transfected the cells were grown for one day before immunoprecipitation.

5.7.2 Protein Extraction

Proteins were extracted from the cell with the following protocol: first the cells were washed three times with ice cold 1 x PBS. Then the cells were scraped with a cell scraper (Corning Incorporated) and collected with 300 µl of RIPA buffer (150 mM sodium chloride, 1.0% NP-40 or Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS (sodium dodecyl sulphate), 50 mM Tris, pH 8.0). One complete protease inhibitor tablet (Roche) was added to 10 ml of RIPA buffer. The cell lysate was kept on ice for 30 min and centrifuged at 12 000 rpm for 20 min at +4°C. The supernatant was then moved to a fresh tube. Protein concentration was measured using the Bradford method. 200 µl of Protein Assay Dye Reagent (BioRad) was pipeted as triplets into a microtiter plate. Standard samples were made from BSA (Bovine Serum Albumin, Lyophilized, Biowest) ranging from 0.25 - 4 mg/ml. 1 µl of each sample was added as triplets to the Protein Assay Dye Reagent solution. The concentrations were measured with the SpectraMAX 190 (Molecular Devices) spectrophotometer.
5.7.3 Western Blotting

5x SDS-PAGE loading buffer was added to the protein samples (10-20 µg protein). The proteins were separated in a 12 % polyacrylamide gel in 1x running buffer (0.3 % Tris base, 1.44 % glycine, 0.06 % SDS, pH 8.3) and transferred to a Hybond-ECL membrane (Amersham Biosciences) by wet transfer in a transfer buffer (3% Tris base, 14.4 % glycine, 0.0025 % SDS, 200 ml methanol). The membranes were blocked in 5 % milk-TBST (5 % milk powder, 0.24 % Tris base, 0.8 % Sodium chloride, 1 % Tween-20) four an hour at room temperature and incubated overnight at +4° C with primary antibody (1/1000, a mouse monoclonal c-myc antibody raised against peptides EQKLISEEDL, Roche) in 3 % BSA TBST (3 % Albumin bovine serum in 1 x TBS with 0.1 % Tween-20). The next day the membranes were washed with TBST for 5 min three times, incubated with anti-mouse IgG secondary antibody (1/10 000, Jackson ImmunoResearch Europe) and for an hour at room temperature and washed again with TBST for 5 min three times. The signal was detected with ECL reaction (GE Healthcare).

After this the membranes were stripped with stripping buffer at 50 °C for 30 min (1.5 M TRIS pH 8.8, 20 % SDS). 390 µl of merkaptoethanol (Sigma) was added to 50 ml of stripping buffer before use. The membranes were blocked in 5 % milk-TBST for an hour at room temperature and washed 3 times 5 min with TBST. β-tubulin was used as the loading control of the samples. Primary antibody was a rabbit polyclonal beta-tubulin antibody (1/1000, raised against peptides LVSEYQQYQDATADEQGE, Cell signaling) in 3 % BSA TBST and the secondary antibody an anti-rabbit IgG secondary antibody (1/10 000, Jackson ImmunoResearch Europe) also in 3 % BSA TBST. Primary antibody was kept overnight and secondary only for an hour. The membranes were washed before changing the antibody and again after the secondary antibody with TBST for 5 min three times. The signal was detected with ECL reaction.

5.8 The Conservation of the Missense Mutation p.E254V in TRIM2

The conservation of the missense mutation p.E254V in TRIM2 was modeled using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and visualized with Jalview sequence alignment program. The protein sequences of human, mouse, chimpanzee, rat, cow, elephant, horse, chicken, platypus, frog, zebra and dog were found from NCBI website. The isoform 1 of TRIM2 protein was used throughout this master study (NP_056086.2, http://www.ncbi.nlm.nih.gov/gene).
6. RESULTS

6.1 THE SUCCESS OF WES ANALYSIS

In this section the results of the whole WES analysis are presented. First the coverage of the analysis is shown followed by the results of all the variants analyzed in the study.

6.1.1 THE STATISTICS OF WES ANALYSIS

The mean bait coverage describing the mean depth of sequenced targets is shown in Table 7 for each sample. This coverage was lower for patient 3 as was also the 20-fold coverage of targets. The 20-fold coverage of the four exomes being analyzed varied slightly among the patients. Overall the 20-fold coverage of targets was more than 50% but not more than 78% for all of the exomes (Table 7).

Table 7. The mean bait coverage and 20 x coverage of WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Mean bait coverage (times)</th>
<th>20 x coverage of targets (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>48</td>
<td>68</td>
</tr>
<tr>
<td>Patient 2</td>
<td>48</td>
<td>62</td>
</tr>
<tr>
<td>Patient 3</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>Patient 4</td>
<td>63</td>
<td>78</td>
</tr>
</tbody>
</table>
6.1.2 WES ANALYSIS RESULTS

6.1.2.1 Homozygous SNVs

Three homozygous candidate variants were found among the WES data of the studied four patients. One of these variants, GTF2I repeat domain containing 2 (GTF2IRD2) gene, found in patient 1 was interesting and the authenticity of this mutation was verified by Sanger sequencing. This variant received a score 4 in the scoring system described above. The other two homozygous variants were ruled out by reasons described in “The classification of interesting variants”-section. The WES analysis steps of identifying these variants and the number of variants in those different steps are presented in Table 8.

Table 8. The number of remaining variants after the steps of identifying homozygous variants from WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of SNVs</th>
<th>Selection of homozygous variants</th>
<th>Exclusion of known SNPs</th>
<th>Exclusion of non-genic variants</th>
<th>Selection of nonsynonymous variants</th>
<th>Selection of variants which have a damaging or a stop codon prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>571204</td>
<td>173951</td>
<td>1230</td>
<td>63</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Patient 2</td>
<td>178245</td>
<td>57725</td>
<td>1289</td>
<td>33</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patient 3</td>
<td>386923</td>
<td>120378</td>
<td>1686</td>
<td>56</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Patient 4</td>
<td>216322</td>
<td>67677</td>
<td>1356</td>
<td>43</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
6.1.2.2 Compound heterozygous SNVs

The amount of compound heterozygous variants was much higher than the amount of homozygous variants (Table 9). These variants were all individually valuated and finally three genes were selected as candidate genes of early-onset neuropathies using the guidelines described in “The classification of interesting variants”-section. The authenticity of the mutations was verified by Sanger sequencing. These genes were Synemin (SYNM) in patient 2, Plectin (PLEC) in patient 3 and Bassoon (BSN) in both patients 3 and 4. The scores that the three variants received were for PLEC and SYNM five and for BSN four.

Table 9. The number of remaining variants after the steps of identifying compound heterozygous variants from WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of SNVs</th>
<th>Selection of heterozygous variants</th>
<th>Exclusion of known SNPs</th>
<th>Exclusion of non-genic variants</th>
<th>Selection of nonsynonymous variants</th>
<th>Selection of variants which have a damaging or a stopcodon prediction</th>
<th>Two heterozygous variants in a same gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>571204</td>
<td>375033</td>
<td>73233</td>
<td>2810</td>
<td>429</td>
<td>159</td>
<td>6</td>
</tr>
<tr>
<td>Patient 2</td>
<td>178245</td>
<td>120520</td>
<td>27224</td>
<td>2495</td>
<td>777</td>
<td>355</td>
<td>19</td>
</tr>
<tr>
<td>Patient 3</td>
<td>389265</td>
<td>266545</td>
<td>77601</td>
<td>4504</td>
<td>1006</td>
<td>444</td>
<td>15</td>
</tr>
<tr>
<td>Patient 4</td>
<td>216322</td>
<td>148645</td>
<td>40394</td>
<td>3398</td>
<td>787</td>
<td>333</td>
<td>15</td>
</tr>
</tbody>
</table>
6.1.2.3 Homozygous deletions and insertions

No interesting homozygous indel variants were identified for any of the four patients by the approach used in this study and the guidelines described in “The classification of interesting variants”-section (Table 10).

Table 9. The number of remaining variants after the steps of identifying compound heterozygous variants from WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of del/ins</th>
<th>Exclusion of known SNPs</th>
<th>Exclusion of non-genic del/ins</th>
<th>Selection of homozygous del/ins</th>
<th>Exclusion of del/ins which depth is less than 10</th>
<th>Exclusion of del/ins which variants calls are less than 10</th>
<th>Exclusion of repeat del/ins</th>
<th>Selection of genes that code for a protein</th>
<th>Selection of del/ins which are in the coding area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>47016</td>
<td>25214</td>
<td>1460</td>
<td>396</td>
<td>237</td>
<td>219</td>
<td>187</td>
<td>84</td>
<td>4</td>
</tr>
<tr>
<td>Patient 2</td>
<td>15610</td>
<td>8447</td>
<td>1054</td>
<td>263</td>
<td>207</td>
<td>182</td>
<td>157</td>
<td>70</td>
<td>6</td>
</tr>
<tr>
<td>Patient 3</td>
<td>26465</td>
<td>14332</td>
<td>1088</td>
<td>325</td>
<td>198</td>
<td>155</td>
<td>135</td>
<td>53</td>
<td>4</td>
</tr>
<tr>
<td>Patient 4</td>
<td>16792</td>
<td>9056</td>
<td>1242</td>
<td>356</td>
<td>288</td>
<td>244</td>
<td>215</td>
<td>93</td>
<td>7</td>
</tr>
</tbody>
</table>
### 6.1.2.4 Compound heterozygous deletions and insertions

A small number of compound heterozygous indel variants were identified but none of them was found to be a possible disease gene for any of the four patients by the approach used in this study and the guidelines described in “The classification of interesting variants”-section (Table 11).

Table 11. The number of remaining variants after the steps of identifying compound heterozygous indel variants from WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of del/ins</th>
<th>Exclusion of known SNPs</th>
<th>Exclusion of non-genic del/ins</th>
<th>Selection of heterozygous del/ins</th>
<th>Exclusion of del/ins which reference calls is 1</th>
<th>Taking account only genes that have two del/ins in them</th>
<th>Selection of compound heterozygous del/ins which are in the coding area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>47016</td>
<td>25214</td>
<td>1460</td>
<td>657</td>
<td>645</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Patient 2</td>
<td>15610</td>
<td>8447</td>
<td>1054</td>
<td>456</td>
<td>445</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Patient 3</td>
<td>26465</td>
<td>14332</td>
<td>1088</td>
<td>474</td>
<td>468</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td>Patient 4</td>
<td>16792</td>
<td>9056</td>
<td>1242</td>
<td>521</td>
<td>516</td>
<td>36</td>
<td>1</td>
</tr>
</tbody>
</table>
6.1.2.5 Heterozygous indel variants combined with heterozygous SNVs

When the heterozygous indel variants were combined with heterozygous SNVs and further analyzed by the steps described in “The classification of interesting variants”- section, two variants in the gene Tripartite motif containing 2 (TRIM2) were found to be interesting in patient 2 (Table 12). The variants received a score 5 in the scoring system. These variants were also verified by Sanger sequenced.

Table 12. The number of remaining variants after the steps of identifying heterozygous indel variants combined with heterozygous variants from WES data.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of del/ins</th>
<th>Exclusion of known SNPs</th>
<th>Exclusion of non-genic del/ins</th>
<th>Selection of heterozygous del/ins</th>
<th>Exclusion od del/ins which reference calls is 1</th>
<th>Combining the heterozygous variants with the heterozygous del/ins</th>
<th>Selection of del/ins which are in the coding area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>47016</td>
<td>25214</td>
<td>1460</td>
<td>657</td>
<td>645</td>
<td>1215</td>
<td>2</td>
</tr>
<tr>
<td>Patient 2</td>
<td>15610</td>
<td>8447</td>
<td>1054</td>
<td>456</td>
<td>445</td>
<td>1143</td>
<td>5</td>
</tr>
<tr>
<td>Patient 3</td>
<td>26465</td>
<td>14332</td>
<td>1088</td>
<td>474</td>
<td>468</td>
<td>1207</td>
<td>2</td>
</tr>
<tr>
<td>Patient 4</td>
<td>16792</td>
<td>9056</td>
<td>1242</td>
<td>521</td>
<td>516</td>
<td>1217</td>
<td>2</td>
</tr>
</tbody>
</table>

6.1.3 The five candidate genes

The classification of interesting variants was done taking into account several different aspects of the variants analyzed in this step. The number of interesting variants varied greatly between the four patients. Five different genes were selected as promising candidates and the authenticities of the base changes were verified by Sanger sequencing. Interesting variants and the expression patterns of the corresponding genes are described in Table 13. Also the scores, which a specific variant received in the scoring system used earlier when identifying the interesting variants, are illustrated in the same table.
Table 13. The interesting variants selected to re-sequencing. Also the gene expression patterns, mutation types and amino acid changes are presented

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of interesting variants</th>
<th>The sequenced interesting variants</th>
<th>The strongest expression (s) of the gene</th>
<th>The mutation type (s)</th>
<th>Amino acid change (s)</th>
<th>Score given in the scoring system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>16</td>
<td>GTF2IRD2</td>
<td>Not in a specific tissue</td>
<td>Homozygous SNV</td>
<td>E31K</td>
<td>GTF2IRD2: 4</td>
</tr>
<tr>
<td>Patient 3</td>
<td>26</td>
<td>BSN, PLEC</td>
<td>BSN: In nerves and brain PLEC: In all tissues and in nerves</td>
<td>Both compound hetero SNV</td>
<td>BSN: P205H and P2902Q PLEC: Q3921L and W1020L</td>
<td>BSN: 4 PLEC: 5</td>
</tr>
<tr>
<td>Patient 4</td>
<td>25</td>
<td>BSN</td>
<td>In nerves and brain</td>
<td>Compound hetero SNV</td>
<td>S1009I and Q1084H</td>
<td>BSN: 4</td>
</tr>
</tbody>
</table>

When sequenced, it was noticed that the homozygous variant c.93T>C in the gene GTF2IRD2 in the patient 1 did not segregate correctly in the family.

For the patient 2 two genes seemed promising, SYNM and TRIM2. The compound heterozygous variants in the gene SYNM were c.2502A>C and c.3738T>G. These variants were not authentic.

The heterozygous c.761T>A variant and the heterozygous c.1779delA in TRIM2 turned out to segregate correctly in the family.

Both the patient 3 and 4 had compound heterozygous variants in the gene BSN, which were not authenticated when verified by Sanger sequencing. For the patient 3 the variants were c.615A>C and c.8706A>C and for the patient 4 the variants were c.3027T>G and c.3252T>G. The patient 4 had a second interesting candidate variant, PLEC, for which Sanger sequencing could not be completed because of poorly working primers. The compound heterozygous variants in PLEC were c.11763T>A and c.3060A>C.
6.1.4 **SANGER SEQUENCING THE INTERESTING VARIANTS VERIFIED THE AUTHENTICITY OF THE TRIM2 VARIANTS**

All interesting variants were re-sequenced to verify their authenticity. None of the candidate variants identified by WES proved to exist except the two TRIM2 variants in patient 2 (Figures 9 and 10).

![Figure 9. The missense mutation c.761T>A in patient 2.](image)

![Figure 10. The deletion c.1779delA in patient 2.](image)

The DNA samples of the parents of patient 2 were sequenced to confirm the correct segregation of the missense mutation and the deletion. The mother was a heterozygous carrier of the mutation c.761T>A (Figure 12) and the father a heterozygous carrier of the c.1779delA (Figure 11). This confirmed that the missense mutation and deletion are situated in different alleles in the patient 2.
Figure 11. The deletion c.1779delA in the parent samples (on the top row mother, on two rows in the middle father and on the lowest row mother samples).

Figure 12. The missense mutation c.761T>A in the parent samples (on the top row mother, on the two middle rows father and on the lowest row mother samples).
6.1.5 The TRIM2 Variants Are Not Present in the General Population

It was assumed that if the TRIM2 variants are pathogenic, they should be extremely rare in the Finnish population due to the small number of patients suffering from early-onset neuropathies in Finland. Thus the carrier frequencies of the variants were determined.

183 control DNA samples (366 chromosomes) from people with a Finnish origin were analyzed for the deletion in TRIM2. 187 control DNA samples (374 chromosomes) from people with a Finnish origin were analyzed for the missense mutation in TRIM2. As a result, no carriers of either variant were identified.
6.2 *TRIM2*, A POSSIBLE NOVEL DISEASE GENE FOR EARLY-ONSET NEUROPATHIES

The 1-bp deletion in the position c.1779 in exon 8, identified by WES analysis in the gene *TRIM2*, causes a premature stop codon that truncates the TRIM2 protein (p.K594Rfs7X). This results in a shortened version of the TRIM2 protein, as the mutated TRIM2 is only 587 amino acids long. The non-mutated TRIM2 protein’s length is 772 amino acids. The variant T>A in the position c.761 in exon 5 changes the glutamic acid in this position into valine (p.E254V).

TRIM2 protein comprises of several different domains. The mutation p.E254V locates in the coiled-coil domain of TRIM2 and lies on a greatly conserved area (Figure 13). The SIFT genome tool predicted that the amino acid change from glutamic acid to valine is significant enough to cause changes to the function of the protein.

![Figure 13: The conservation of the missense mutation p.E254V in TRIM2.](image-url)
6.3 **mTRIM2 SITE-DIRECTED MUTAGENESIS**

The site-directed mutagenesis was done to be able to investigate the effect of c.761T>A and c.1779delA on the function of TRIM2 in future studies (Figure 14). Therefore, the constructs are believed to benefit the future research by helping in the functional analysis of *TRIM2* as a causative gene. Both of the mTrim2 mutagenesis constructs turned out to be successful. The mTrim2 constructs were verified by sequencing the whole insert and making sure that there were no unintentionally induced mutations in the constructs and that the deletion and missense mutation were successfully planted into the constructs.

![Diagram of mTRIM2, mRBCC-Trim2, mNHL-Trim2, c.761T>A, mTRIM2-c.761T>A, mTRIM2-c.1779delA, c.1779delA](image)

**Figure 14.** Site-directed mutagenesis constructs and constructs received from Dr Balastik, modified from Balastik et al. 49.
6.4 Western blotting of mTrim2 constructs

Western blotting was done to set up the conditions for functional studies with mutant mTrim2 constructs. Western blot was performed using the protein lysates from mTrim2, mRBCC-Trim2 and mNHL-Trim2 transfected HEK cells and it presents the sizes of these constructs (mTrim2 approximately 150 kDa, mRBCC-Trim2 and mNHL-Trim2 approximately 90 kDa) (Figure 15). Beta-tubulin was used as the loading control.

Figure 15. The western blot from protein lysates of mTrim2 transfected cells.
7. DISCUSSION

In this master study, whole exome sequencing analysis was performed on four patients diagnosed with early-onset axonal neuropathies. As a result, a good candidate gene for early-onset neuropathies was identified. The use of whole exome sequencing in the identification of new disease genes for CMT is justified in this study by its fast and cost-effective nature and its particular suitability for diseases with a recessive inheritance pattern. Whole exome sequencing has already been used successfully in the molecular diagnosis of CMT \(^{43, 41}\) and in the identification of genes for other recessive syndromes \(^{50, 45}\). Whole human genome sequencing has also been used in the identification of mutations in known genes in recessive forms of CMT but it is significantly more expensive than whole exome sequencing \(^{51}\). In the following section, the fulfillment of the aims set out for this study is discussed.

7.1 IMPROVING THE PROCESS OF ANALYZING VARIANTS FROM WES DATA

One of the aims of this study was to improve the process of analyzing variants from WES data. For this aim to be realized, a systematic analysis technique was needed to be set up and all the steps in this analyzing process to be reevaluated. The most crucial steps in this procedure are further discussed and their success evaluated in the next sections.

7.1.1 ANALYZING WES DATA

7.1.1.1 The analyzing steps of WES

The analysis of WES data was initiated by examining for homozygous causative variants. After this the compound heterozygous variants were investigated. The analysis continued by exploring the indel variants (homozygous, compound heterozygous and heterozygous indel variants combined with heterozygous variants). The first steps in the analysis of all the variants were the same: first the known SNPs and the non-genic variants were excluded from the data and the homozygous or heterozygous variants/indels selected. After this some variation was introduced to the steps depending on the nature of the analyzed variants.
The indel variant analysis process continued with additional steps, which were made with the aim of decreasing the overall number of variants. After that, the repeat indel variants were the excluded and this step was then followed by the selection of protein coding genes. For the compound heterozygous indel variants only genes with two indel variants in them were taken into account. For the heterozygous indel variants combined with heterozygous variants this step was done after the ones described earlier. The last step done for all the indel variants analyzed in this study was the selection of variants, which are in the coding area.

The analysis of homozygous variants and compound heterozygous variants continued with the prediction of damaging variants, filtering out the synonymous variants and selecting the variants with a damaging or a stop codon prediction. For the compound heterozygous variants the final step was to take into account only genes with two different mutations in them.

The analyzing steps used in the study can be described as being effective and they were performed in a rational order. The analyzing process of indel variants, both homozygous and compound heterozygous included more steps than the analysis of homozygous and compound heterozygous variants. The number of indel variants is typically smaller than the amount of single nucleotide variants (SNVs) which is consistent with the WES analysis of this Master study.

7.1.1.2 The success of WES analysis

The target capture of this master study’s WES analysis was poor resulting in low sequencing coverage. To overcome the problem, the sequencing was repeated but still the coverage was not as high as hoped. An optimal 20-fold sequence coverage would have been at least 95 %, which was not even nearly accomplished. The re-sequencing done to repair this problem did not help because of poor target capture, which instead resulted in a high amount of variants from the non-coding areas to be included in the WES data.

These obstacles obviously had their effects on the analysis steps of WES data and on the overall success of the WES analysis. Firstly, the number of SNVs was significantly higher than what is normally seen in WES data (usually 100 000 SNVs but for our patients the number of SNVs ranged from 178 000 to 580 000). This was due to the poor target capture as already mentioned. The problem with the target capture had its effects especially on the analyzing steps of indel variants and resulted in a number of steps that had to be done to be successful in decreasing the number of variants, to be finally able to identify good candidates from the pool of variants. The steps done to insure this goal when analyzing
homozygous indel variants were “Exclusion of indel variants with a depth less than 10” and “Exclusion of indel variants with variant calls less than 10” (Figure 5). When analyzing compound heterozygous indel variants and heterozygous indel variants combined with heterozygous variants the step done to fulfill this goal was “Exclusion of variants which reference call is 1” (Figures 6 and 7). The poor coverage resulted in a high number of variants with a low number of reads, making it extremely difficult to predict if these variants were authentic or not.

Although the capture and coverage of this master study’s WES were poor, interesting variants were found and one candidate gene identified. This shows that even through bad capture it is possible to find causative variants. However, it is noted that the disease gene was identified for the patient having the least amount of SNVs in the beginning. Also, when it is known that the capture of the WES analysis has not been successful it is important to check that the variants are in the coding area when identifying the indel variants of exome sequencing data using for example IGV. By doing this, a large number of variants were identified as false.

In this study the polymorphisms listed in the dbSNP database were filtered out. It was assumed that any allele found in the dataset could not be disease causing. This approach has its flaws. First, the dbSNP database is “contaminated” by known pathogenic alleles. Secondly, because of investigating a recessive disease the segregation frequency of the pathogenic alleles is expected to be very low because it is not likely that a causative variant would be a common one. This makes the assumption of not filtering out pathogenic alleles by using dbSNP database or 1000 Genomes database as a filter set problematic. Also in recessive diseases the carrier status of a pathogenic allele will not result in a phenotype that might not exclude an individual from a control population, such as 1000 Genomes Project. Recently the use of 1000 Genomes frequency 0.005 as a filter instead of dbSNP database has been successfully used in WES analysis. This could be one way of improving the analysis process even more and it could ultimately result in identifying disease genes for all the patients participating in this study.

Moreover, trusting to the SIFT genome tool in the analyzing steps of WES data in terms of the deleteriousness of the variants leaves a chance for bias. The authenticity of this verification is much dependent on in the later steps of the analysis of homozygous and compound heterozygous variants as well as heterozygous indel variants combined with heterozygous variants. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related species. SIFT predictions are widely dependent on when performing for example WES analysis, but it should be remembered that these predictions should not be trusted blindly and variants not possessing a damaging or a stop codon prediction are possible causative variants too, at least in theory.
The use of IGV tool was proven to be crucial when the target capture is poor. Many interesting indel variants were excluded with the help of IGV when found out that they are not in the coding area of the gene. Using IGV in this step proved to be indispensable because WES data always contains some variants from the noncoding areas.

One of the aims of this Master study was to develop the analyzing steps of WES data and especially advance the analysis of indel variants. It can be claimed that this aim was accomplished as a candidate disease gene was found through an indel analysis of the WES data and by looking at heterozygous indel variants combined with heterozygous variants. The result of this study further proves that the analysis of insertion and deletion variants is an important feature in the analysis of WES data.

Though, it is also noted that three patients were left without a candidate disease gene and for this, it can be speculated if remodeling this phase of the study would help to give answers to these patients also. However, the number of SNVs in the WES data of these patients was extremely high (especially for patient 1) which makes the identification of causative variants very challenging.

7.1.1.3 The classification of the interesting variants

The classification of interesting variants is one of the most important steps in the analysis of WES data and at the same time also one of the most sensitive steps for making errors. Losing an interesting variant into the huge amount of WES data can result in misplacing one of the possible causative genes. Thorough consideration is required for getting rid of the non-pathogenic variants without losing the interesting variants at the same time.

When the scoring system was put into practice, the amount of interesting variants had already been cut down significantly and many of these variants had a very low coverage, which suggested that the variants were not authentic. The motive for using this scoring system was to quickly identify the most promising candidate genes. The scoring system was based on important facts about the genes under investigation and it can be seen as an effective tool as well as well suited to be used at this stage of the study. The scoring system is one for the phases of the WES analysis, that could be further improved in the future and also one of the steps, if done differently, could result in finding candidate disease genes for all patients.
7.2 IDENTIFYING THE MOLECULAR CAUSE OF DISEASE FOR FOUR PATIENTS SUFFERING FROM AXONAL NEUROPATHIES USING WES

7.2.1 THE RE-SEQUENCING OF THE VARIANTS

The re-sequencing of the interesting variants was successful for all variants except those in the PLEC gene. The primers designed to target this gene did not work as they should have and because of time restrictions the sequencing was aborted. This can have resulted in not identifying one of the possible disease genes, but that is highly unlikely because of the small number of variant calls of the PLEC variants in the WES data. During the re-sequencing, it became obvious that all variants, whose variant calls were two or three, were not authentic. This was the case with the BSN and SYNM mutations. Based on this fact, it can be suggested that “variant calls under 10” should be eliminated in variant identification in the future.

7.2.2 THE MISSING CAUSE FOR DISEASE FOR PATIENTS 1, 3 AND 4

The reasons for not identifying a disease gene for all the patients who participated in this study are complex. First of all, the unsuccessful target capture and poor sequence coverage increased the amount of SNVs and the amount of variants outside the coding area. Some variants might have been lost due to the unsuccessful capture. This made the planning of the analyzing steps even harder. Of course, the aim was to eliminate false variants and end up with finding a causative variant. However, not always the taken steps are heavy enough or gentle enough to be able to keep all the interesting variants aboard the analysis and at the same time to be able to decrease the number of variants so that the interesting ones can be identified from the pool.

Secondly, looking at the characteristics of the interesting variants to be able to decide which variants should be further investigated and re-sequenced was a very challenging step. In this point a possible mistake could have been made, when excluding the variants from re-sequencing based on the function of the gene and the tissue specific expression patterns. Moreover, the disease genes of CMT are very heterogeneous with multiple different functions and expression patterns. This is why it was difficult to positively exclude a gene from being the disease gene of CMT, when there was no specific kind of gene function that would unquestionably suggest that the gene is a possible disease gene. It is possible, that the disease genes for these three patients lie in the list of genes, which got a score of 1–4 with the scoring
system. These genes could be first sequenced before starting the whole analysis of WES data over for the remaining three patients.

7.2.3 **IDENTIFYING CANDIDATE DISEASE GENE FOR PATIENT 2**

By WES analysis a probable disease gene was identified for one of the four patients participating in this study. The possible gene, *TRIM2* encodes for an E3 ligase protein that ubiquitinates NF-L\(^49\). The facts, which initiated and later strengthened the suspicion of the causative nature of these *TRIM2* mutations were the function and expression of the gene, the effect of the amino acid changes to the proteins function, the segregation of the missense mutation and the deletion, the conservation of the area where the found missense mutation is located, the rareness of the missense mutation and the deletion and the already existing mouse model for *TRIM2*.

*TRIM2* is strongly expressed in neurons and brain and is known to affect the polarization of axons, which is an important step in the proliferation of axons\(^52\). One of the targets of *TRIM2* is NF-L, which is part of the axoskeleton and functionally maintains the neuronal caliber. Mutations in the gene *NEFL* coding for NF-L also cause Charcot-Marie-Tooth disease types 1F and 2E. This further reinforces the idea of TRIM2 being a disease gene for early-onset axonal neuropathy. Moreover, the gene *LITAF* (previously called *SIMPLE*), which is shown to cause CMT1C is known to react with an E3 ligase protein Nedd4\(^53\). Therefore, it is not farfetched that mutations in a gene encoding an E3 ligase could be the cause of CMT type 2. Especially, when considering that *TRIM2* is shown to interact with proteins, which are highly involved in axon development and function\(^49,52\).

The SIFT genome tool predicted the amino acid changes caused by the missense mutation in *TRIM2* to be harmful for the function of the protein. The missense mutation is located in the coiled-coil domain in TRIM2, which is a highly conserved area in the protein. These facts further fortify the suspicion of *TRIM2* being a disease gene.

The missense mutation and deletion were shown to segregate correctly in the patient’s family, which was required for them to be able to be located in different alleles.
7.2.3.1 TRIM2 belongs to E3 ligase protein family

TRIM2 is a member of the TRIM super family of E3 ligase proteins. These proteins function as ubiquitin ligases (E3s), those in the combination of E2 ubiquitin-conjugating enzymes attach ubiquitin to a lysine of the target protein via an isopeptide bond. The enzyme families involved in the ubiquitination reaction are E3s, ubiquitin activating enzymes and ubiquitin conjugating enzymes. E3s’ task in this process is to ensure the correct localization, timing and specificity of the ubiquitination reaction. Ubiquitination of a protein can influence the protein’s activity and stability or its interactions and intracellular localization.

TRIM2 or tripartite motif protein 2 contains a cysteine-rich RING (really interesting new gene) - finger domain, a B-box and a coiled-coil structure (Figure 16). It has been characterized based on its conserved modular structure. In spite of the conserved modular structure, no general biological role has yet been found for the TRIM proteins. Tripartite RING finger proteins are actually a subgroup of RING finger E3s, which are also the most abundant E3 proteins.

![TRIM2 protein structure](image)

Figure 16. TRIM2 protein structure. R, RING-finger domain; B-Box, B-Box domain; CC, coiled-coil structure. Picture modified from.

7.2.3.2 Previous investigations about TRIM2

The role of TRIM2 in the ubiquitination of Bcl-2-interacting mediator of cell death (Bim) has been previously investigated using a proteomics approach. It has been shown that Bim is ubiquitinated and degraded followed by neuroprotection episode observed in rapid ischemic tolerance. In this study TRIM2 was shown to bind to Bim when it is phosphorylated by p42/p44 MAPK and ubiquitinate it in rapid ischemic tolerance. Also, mice deficient in Trim2 have been shown to have increased levels of neurofilament light chain (NF-L) in axons. In addition NF-L filled axonal swellings were found from the mice’s cerebellum, retina, spinal cord and cerebral cortex. This particular study stated that TRIM2 regulates the ubiquination of NF-L and not only Bim as shown above. In this study also an
The polarization of a single axon is greatly dependent on the changes in the dynamics of the neuronal cytoskeleton. These molecules are still weakly understood. This is why a study has been performed where the genes, that are differentially regulated in developing neurons, were studied with the aim of eventually identifying these proteins of interest. The gene L-TRIM of the snail Lymnaea stagnalis was found to be up-regulated in postnatal brain development. TRIM2, TRIM3 and TRIM32 are the mammalian orthologues of L-TRIM. In this paper it was shown that TRIM2 regulates the polarization of mouse hippocampal neurons by the ubiquitination of NF-L. Also the over-expression of TRIM2 increases the specification of multiple axons whereas the suppression of TRIM2 results in the loss of neuronal polarity. TRIM2 was thereby shown to have an important function for axon outgrowth during development.

All of these studies support the hypotheses of TRIM2 being a candidate disease gene for axonal neuropathy. The mice deficient in TRIM2 and therefore suffering from axonopathy particularly strengthen this suspicion. They showed that TRIM2 deficiency is associated with neurofilament accumulation and axonal degeneration, which is due to deficient ubiquitination and thus inefficient degradation of the NF-L subunit of neurofilament, a substrate of TRIM2-mediated ubiquitination. This kind of lack in the regulation of the intermediate filament network can be suspected to have a major impact in the axons. Furthermore, when TRIM2 is shown to have a function in the specification of axons and in the axon polarization, it is even more probable that deficiencies in this gene can be the cause of a neuropathy.

7.2.4 The Possible Effects of the Missense Mutation and Deletion on the Function of TRIM2 Protein

As the SIFT Genome Tool predicted, the missense mutation found in TRIM2 is likely to have a damaging effect on the protein function. The effects that the c.761T>A mutation could have on the TRIM2 protein include disruption of its structure and instability of the whole protein. The missense
mutation found in TRIM2 is located in a highly conserved area in the protein. This high degree of conservation implies that this area holds important functions in mediating specific protein interactions. If this area is mutated, it could be disallowed from performing its tasks and this could lead to the instability of the protein. The missense mutation could this way prevent TRIM2 from functioning properly as an E3 ubiquitin ligase.

The deletion in TRIM2 is very likely to have some major effects on TRIM2, because it causes a premature stop codon to the protein. One of these effects could be the degradation of mRNA containing the c.1779delA deletion. If the missense mutation would cause the TRIM2 mutant protein to be unstable and the deletion would cause the mRNA containing it to be degraded, the patient cells could show a significant deficiency in TRIM2. Furthermore, TRIM2 is known to ubiquitinate NF-L, whereupon it can speculated that the deficiency of TRIM2 protein could have some effects in the neurofilament structures of the cells.

7.2.5 TRIM2 CARRIER FREQUENCY

As stated before, the carrier frequency studied by minisequencing showed that the allele frequency was 0.00 for both the mutation c.761T>A and the deletion c.1779delA, meaning that these mutations are not present in the general population. This result was quite expected and it reinsures the causative nature of the TRIM2 mutations, because the disease causing mutations were not expected to be common due to the small number of patients suffering from early-onset neuropathies in Finland. It has been previously noticed that the disease causing mutations can be extremely rare in CMT and this for its part explains for example why TRIM2 mutations were not found in the other patients participating in this study. Also the high number of disease causing genes in CMT supports the possibility of TRIM2 adding to this group of disease genes.

Adding more items to the carrier frequency sample could have increased the reliability of the carrier frequency in some extent. If for example 400 chromosomes were studied instead of less than 300 the soundness of these results could have been a little bit higher.
7.2.5 The Limitations of Verifying the Causative Nature of TRIM2 Mutations

Patient cells could have helped to further confirm the causative nature of TRIM2 missense mutation and deletion. These cells could have been used to investigate the effect of the missense mutation and the deletion to the function and stability of the TRIM2 protein. The mTrim2 autoubiquitination assay described in 49 could also prove to be beneficial and the effects of the missense mutation and deletion to the ubiquitination of NF-L could be further investigated. These results could help to further see how big effects the missense mutation and the deletion actually have on the function of the TRIM2 protein.

7.3 Performing Functional Analysis of Identified Potential Disease Mutations

Because of time restrictions, this master study succeeded only in starting up the functional analysis of TRIM2. The functional analysis was initiated by performing a western blot with the aim of setting up the conditions for subsequent functional studies with mutant TRIM2 constructs.

7.3.1 mTrim2 Mutagenesis and Western Blotting

The mutagenesis of both TRIM2 c.761T>A mutation and the TRIM2 c.1779delA deletion were successful. These constructs are believed to benefit the future research by helping in the functional analysis of TRIM2 as a causative gene. These constructs could possibly be used in the future in an autoubiquitination assay discussed before. In addition, these mutation constructs could be used when investigating the stability of the mutant TRIM2 protein.

A small caution should be kept in mind when using the TRIM2 mutagenesis constructs in future research. The mTrim2 site-directed mutagenesis was done with mouse Trim2 constructs for a future use in an autoubiquitination assay and in TRIM2 antibody testing. Using the mouse TRIM2 in the TRIM2 constructs was not the ideal situation since mouse TRIM2 differs from the human one. Two isoforms of TRIM2 exist for both human and mouse. One is shorter (744 amino acids for human) than the other (771 amino acids for human) in both human and mouse. The shorter form of mouse TRIM2 was used in the constructs of Dr Balastik. This form is 99.2 % identical with the shorter isoform of the human TRIM2. However the longer form of the human TRIM2 is the one that is assumed as the main isoform. This
creates a small hesitation of mTrim2 constructs actually showing the effect of the missense mutation and deletion in human TRIM2. This is although very unlikely.

The western blotting done using the protein lysates from mTrim2, mRBCC-Trim2 and mNHL-Trim2 transfected HEK cells was also successful. From looking at the western blotting it can be seen that the constructs are stable. This shows that western blotting was successful and these constructs are ready for future use when investigating TRIM2 as a disease gene.

7.4 Subsequent research on verifying the causative nature of the TRIM2 missense mutation and deletion

After the work done in this Master study, this project was continued and it finally resulted in a publication of TRIM2 as a new disease gene for early-onset CMT2. The study’s results suggested that loss-of-function mutations in TRIM2 are a cause of axonal neuropathy, which was propose to develop as a consequence of axonal accumulation of neurofilament, secondary to lack of its ubiquitination by TRIM2.

Cultured fibroblasts were obtained from a skin biopsy of the patient and TRIM2 protein levels were measured. Approximately 10% of TRIM2 was present in the patient cells. The low level of TRIM2 protein in patient fibroblasts suggested that the TRIM2 mutant protein was unstable. To further assess this possibility, HEK cells with expression vectors for the wild-type and p.E254V mutant TRIM2 were transfected and subjected to a time course cycloheximide stability assay. Hence, the construct with the c.761T>A mutation generated by mutagenesis in this master study was used in this particular assay. The result of this assay suggested reduced stability for the mutant protein in comparison with the wild-type TRIM2. The mutant TRIM2 protein was suggested to be degraded through other pathways than the proteasome, either by lysosomal and autophagic degradation.

The influence of TRIM2 deficiency in the architecture of intermediate filaments other than the neurofilament was investigated, but no effects in the patient fibroblasts was identified. Also a heterogeneous international cohort of 87 early-onset CMT patients was investigated for mutations in TRIM2. These patients had been diagnosed with axonal, demyelinating and intermediate forms of CMT. No putative mutations were identified in this cohort indicating that TRIM2 has a low mutation frequency and is therefore rare cause for CMT2. Moreover, it was stated that common to the heterogeneous CMT disease, defects in TRIM2 might cause highly variable clinical outcomes. In conclusion, this study
resulted in taking one step closer to knowing the genetic defects leading to early-onset recessive axonal neuropathies.

7.5 FUTURE PROSPECTS AND ETHICAL CONSIDERATIONS OF WES

7.5.1 WHAT THE FUTURE HOLDS FOR WES

Improvements of the technical, statistical and bioinformatic methods of WES analysis are needed in the future. The technical aspects of WES analysis that should be paid attention in the future include the automation of the algorithms for annotating variants and improving the approaches for characterizing the functional impact of rare and novel variants. Also the rate of false-positive and false-negative variant calls should be decreased. Accomplishing these improvements would mean that the diagnosis of various diseases could be made more efficient. Also standards and guidelines for WES testing and reporting in clinical laboratories will need to be established. Moreover, issues related to the data storage of WES data should be addressed. The amount of analysis data will definitely increase in the future and restoring it might prove to be tricky.

This study has taken one step closer in finding the causative genes of different CMT forms. In future, increasing amount of causative genes for Mendelian diseases will be found using WES analysis. When technical problems can be solved and the whole method made reliable, WES will become even faster and more effective method than it is at the moment. Also the price of a single WES run will decrease significantly. WES is depicted to improve genetic diagnosis procedures or even replace traditional genetic testing together with whole genome sequencing in the future. WES can be especially useful for CMT patients, because the exact diagnosis of CMT is very important for the management and treatment of CMT. In the future genetic diagnostics may play an important role for realizing personalized therapy. Identifying a candidate gene (or genes) for every Mendelian recessive disorder using as few as one affected individual per disorder is a realistic goal at present when using WES as the main analysis tool. Solving all dominant disorders will be more difficult, but the enhancements added to WES analysis procedure in the future will continue to make this goal more achievable.
7.5.2 THE ETHICAL CONSIDERATIONS OF WES ANALYSIS

The ethical considerations of next-generation sequencing methods are to be taken seriously. One of main issues when performing a WES analysis study is the management of individual research results. There are several practices to minimize the return of research results to the participants. The results must have been validated and they must have been decided to be clinically useful and to be actionable. Unfortunately, the details of how the terms useful and actionable are defined remain under dispute. The results also need to be identified in the course of routine research analysis. WES approaches also increase the chance of uncovering clinically useful results that are unrelated to the primary aim of the study. This is why it is important to confirm that the patient is truly aware of what kind of information will be available after participating in a study comprising of WES analysis.\textsuperscript{38}
8. ACKNOWLEDGEMENTS

This study was carried out in the group of Henna Tyynismaa, in Biomedicum, Helsinki, Faculty of Medicine, Research Program of Molecular Neurology. Financial support for this study was provided by the Arvo ja Lea Ylppö foundation.

First and foremost I would like to thank my supervisors Henna Tyynismaa and Emil Ylikallio for all the advices and support you have given me.

Members of Wartiovaara and Battersby research groups are thanked for creating an exhilarating atmosphere and providing support throughout this project. Brendan Battersby is also thanked for his participation in this project.

Dr Balastik is also thanked for his collaboration with this study.

My family is thanked for believing in me and always being there for me. I would like to especially thank my older sister Maria who has been an irreplaceable supporter in all of my projects during my university studies.
9. REFERENCES


46. Robinson, P. N., Krawitz, P. & Mundlos, S. Strategies for exome and genome sequence


