GENETICS OF HEREDITARY SENSORY AND AUTONOMIC NEUROPATHY (HSAN)

by

Seden Tezel

B.S., Molecular Biology and Genetics, Middle East Technical University, 2016

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2020

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my thesis supervisor Prof. Esra Battaloğlu for her wonderful guidance, support, and patience throughout the study. Being her student and learning from her was a great experience. I would like to thank my jury members, Prof. Müge Türet and Prof. Yeşim Parman for allocating their time to evaluate my thesis and their valuable criticism.

I would like to thank my one and only lab member Ayşe Candayan for helping me, guiding me, giving her valuable insights from the very beginning and extend my thanks to Harun Niron for his friendship. I would like to express my sincere gratitude to all academic staff in the department for their contribution to my improvement.

I thank my parents and my dearest sister for being there for me and supporting me no matter what. They are the most wonderful family anyone could ever have. My deepest appreciation goes to Özcan Gündeş because of not only his valuable contributions to this study in data analysis but also his incredible support, wonderful personality and making my life much happier. He is really the brightest wizard of our age.

I would also like to share my gratitude to my road companion Başak Kalfa, voice of my conscious, my ultimate study partner and much more to me, we started this journey together and I am incredibly happy to end it together.

My last thanks go to my best friends Yasemin Başkaya, Ezgi Kurtulmuş and Serhat Say for bringing so much joy into my life. I am grateful for having you as my second family.

This study was supported by Boğaziçi University Research Fund (BAP11940).

ABSTRACT

GENETICS OF HEREDITARY SENSORY AND AUTONOMIC NEUROPATHY (HSAN)

Hereditary Sensory Neuropathy (HSN), predominantly affects sensory nerves and is characterized by sensory loss in addition to muscle weakness, wasting and painless injuries. Autonomic symptoms are commonly associated with HSN, thus, it is also known as Hereditary Sensory and Autonomic Neuropathy (HSAN). A Turkish family initially diagnosed with HSAN was investigated in this study to identify the gene responsible for the disorder. The consanguineous parents had three affected and one unaffected children suggesting autosomal recessive inheritance. Along with HSN, cough and gastroesophageal reflux (GER) have been reported as additional symptoms in family. The disease was mapped previously to chromosome 3p22-24 in patients with this phenotype, however, the gene could not have been identified, yet. We have identified the causative candidate gene variants using whole exome sequencing (WES) data in the HSN family based on autosomal recessive, autosomal dominant, and X-linked inheritance. WES data were further evaluated using HOMWES program for homozygosity mapping. The causative role of each variant was evaluated by segregation analysis. Unfortunately, these analyses gave inconclusive results with exclusion of 12 different candidate variants. A high rate of false positive results was observed when WES results were compared with that of direct sequencing. In last month of the study, upon hospitalization, the patient was re-diagnosed as cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome (CANVAS). The haplotype analysis showed the presence of CANVAS associated core haplotype among the affected members of the family. This finding indicated that repeat expansion of the $(AAGGG)_{11}$ unit in *RFC1* gene is responsible for the disorder in this family.

ÖZET

HEREDİTER DUYSAL VE OTONOMİK NÖROPATİ (HSAN) GENETİĞİ

Kalıtsal Duysal Nöropati (HSN), ağırlıklı olarak duyu sinirlerini etkiler ve en bilinen semptomları, kas güçsüzlüğü, zayıflama ve ağrısız yaralanmalara ek olarak duyu kaybıdır. Otonomik semptomlar genellikle HSN ile birlikte seyreder, bu nedenle hastalık, Kalıtsal Duyusal ve Otonom Nöropati (HSAN) olarak da bilinir. Bu çalışmada, hastalıktan sorumlu geni belirlemek için başlangıçta HSAN tanısı alan bir Türk ailesi araştırıldı. Akraba evliliğine sahip ebeveynler, otozomal çekinik kalıtımı düşündüren üç hasta ve bir sağlıklı çocuğa sahiptir. HSN ile birlikte öksürük ve gastroözofageal reflü (GER) ailede ek semptomlar olarak bildirilmiştir. Hastalık, bu fenotipe sahip hastalarda daha önce kromozom 3p22-24 bölgesine haritalanmış, ancak sorumlu gen henüz tanımlanamamıştır. Otozomal çekinik, otozomal baskın ve X'e bağlı kalıtıma dayalı olmak üzere, ailede tüm ekzom dizileme (WES) yöntemi kullanılarak aday gen varyantlarını belirlenmiştir. WES verileri ayrıca homozigotluk haritalaması için HOMWES programı kullanılarak da incelenmiştir. Her aday varyantın rolü, segregasyon analizi ile değerlendirilmiştir. Ne yazık ki, bu analizler 12 farklı aday varyantın dışlanması ile sonuçlanmış, WES sonuçları doğrudan dizileme ile karşılaştırıldığında yüksek oranda yanlış pozitif sonuç gözlenmiştir. Çalışmanın son ayında, hastaneye yatışının ardından hastaya serebellar ataksi, nöropati ve bilateral vestibüler arefleksi sendromu (CANVAS) tanısı konulmuştur. Yapılan haplotip analizi ile, ailenin hasta üyelerinde CANVAS haplotipinin varlığı ortaya çıkmıştır. Bu bulgu, RFC1 genindeki $(AAGGG)_{11}$ diziliminin çok sayıda tekrarlanması ile genişlemesi sonucunda bu ailedeki bozukluktan sorumlu olduğunu işaret etmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Charcot-Marie-Tooth (CMT) Disease

Charcot-Marie-Tooth (CMT), also named as hereditary motor and sensory neuropathy, is an inherited neuromuscular disease which affects peripheral nervous system. The prevalence of CMT is around 1:2500 that makes this disorder the most common inherited neuropathy worldwide (Timmerman et al., 2014). It is genetically heterogeneous with more than 80 disease-associated genes and over a thousand different mutations (Pareyson et al., 2017). CMT is also clinically heterogeneous and is characterized by progressive degeneration of peripheral nerves that leads to muscle weakness and wasting in distal limbs such as hands and feet. Most of CMT patients suffer from sensory loss, muscle atrophy extending to foot deformities and walking disabilities (Nam and Choi, 2019).

Cellular processes that are critical for axonal health and maintenance include myelin structural integrity, anterograde and retrograde axonal transport, dynamics and functions of mitochondria, trafficking between membrane and vesicles, integrity of the axonal cytoskeleton, protein translation and turnover. When a malfunction occurs in these cellular processes, a neuropathy condition occurs (Bussmann and Storkebaum, 2017). A visual representation highlighting the neuropathy genes and the respective cellular processes affected by them, is given in Figure 1.1.

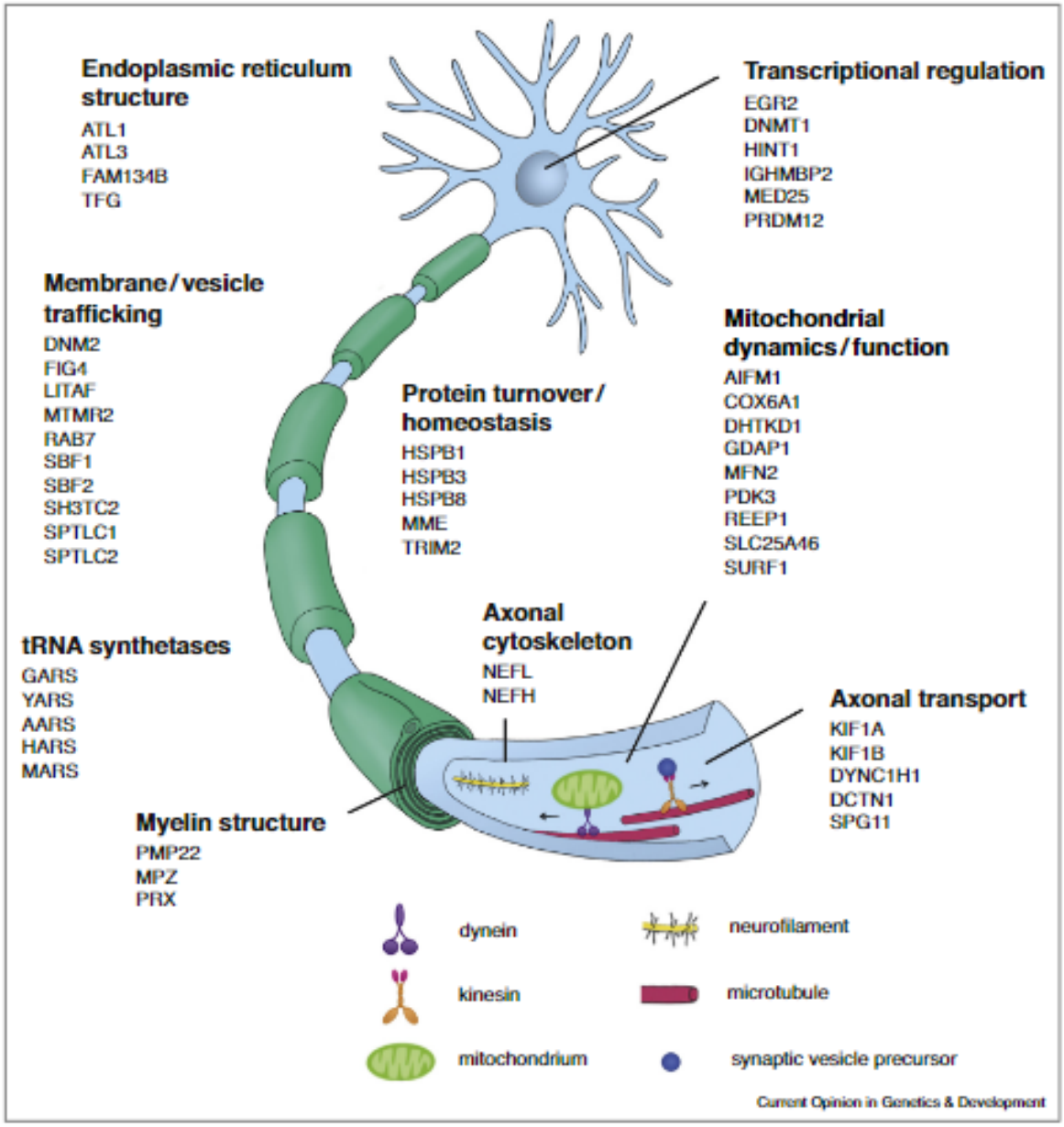

Figure 1.1. Neuropathy genes and their associated cellular processes (Bussmann and Storkebaum, 2017).

CMT disorder has two further subtypes that are described according to the affected part of the peripheral nervous system; in hereditary motor neuropathy (HMN) the motor nerves degenerate predominantly and in hereditary sensory neuropathy (HSN) the degeneration is observed in sensory features. The family investigated in

this study was initially diagnosed as having HSN and then as having a type of cerebellar ataxia with neuropathy known as CANVAS phenotype. Both HSN and CANVAS phenotypes are described with their genetic features in the following sections.

1.2. Hereditary Sensory Neuropathy (HSN)

HSN is a slowly progressing disease with clinical and genetical heterogeneity. Autonomic nervous system can also be affected along with sensory system in patients with HSN, this is why HSN type of disease is also named as hereditary sensory and autonomic neuropathy (HSAN). The exact prevalence of the disease remains unknown, but it is known to be very low (Auer-Grumbach, 2008). It can be inherited autosomal recessively or autosomal dominantly. Symptoms of autosomal recessively inherited HSN disease is generally congenital, present from birth and the disease condition is more severe compared to autosomal dominant cases. HSN with autosomal dominant inheritance has an adult onset and the symptoms may appear between second and forth decade of life. Sensory loss primarily affects distal limbs, feet and legs followed by muscle weakness and wasting. Painless injuries lead to foot ulcers (Kok et al., 2003). Autonomic features are usually in the form of sweating disturbances that are observed without variation (Auer-Grumbach, 2008).

Twenty-four genes and seven chromosomal loci have been associated with HSN so far. These genes, the associated clinical findings, and the types of HSN are given in Table 1.1 that was obtained from Neuromuscular Disease Center website of Washington University (NeuroMuscular, "n.d").

Most commonly observed type of HSN is Type I (HSN1) which is inherited as an autosomal dominant disease. Subunit 1 of the enzyme serine palmitoyltransferase (*SPTLC1*) has been identified as the causative gene for HSN Type I and this subtype is defined as HSN1A. The mutations in *SPTLC1* is known to alter the sphingolipid synthesis (Dawkins et al., 2001), *SPTLC2* is also responsible for HSN1 and this subtype is defined as HSN1C. Sphingolipid synthesis is crucial to neural cells as they have

a role in modulating myelin properties and affecting the parameters of the action potential (Saba et al., 2020). Mutations in sphingolipid synthesis results in accumulation of abnormal enzyme products in place of sphinganine. Early targets of sphingolipid pathways are mitochondria, affected by membrane potential loss and abnormal Ca^{2+} loading, and ER, affected by a leakage of Ca^{2+} from membrane channels. Mechanism is still largely unknown (Wilson et al., 2018).

Kok et al. (2003) described a subtype of HSN1 that is associated with chronic cough and gastroesophageal reflux (GER) in 2003 and this subtype is defined as HSN1B. The causative gene was mapped to chromosome 3p22-24 but itself could not be identified (Kok et al., 2003). They studied two large Australian families in which the disease was inherited autosomal dominantly. Clinics of these two families resembles with our P1403 family in a way of having sensory neuropathy with cough and GER; however, differentiates in mode of inheritance. Because parents were first degree cousins, had three affected and one unaffected children in our family that suggested autosomal recessive inheritance. The index case had congenital hearing loss, cough and gastroesophageal reflux while the affected sisters had additional retinal detachment (cataract).

HSN Type II is the recessive form of the disease and has similarities to Type I in the way of sensory impairment. High frequency of foot ulcers is the major characteristic. Foot ulcers is the main reason that this phenotype is considered as subtype of HSNII. *FAM134B* is the gene that causes HSNIIB, its protein product is critical in survival of ganglion neurons (Kurth et al., 2009).

Table 1.1. Types of Hereditary Sensory Neuropathy (HSN), the associated genes and clinical findings.

HSN Type	**Gene**	**Locus**	**Inheritance**	**Age of Onset**	**Clinical features**
IA	SPTLC1	9q22	Dominant	>20 years	Pan-sensory loss
IB	Not known	3p24	Dominant	20 to 40 years	Pan-sensory loss Cough; GE reflux
IC	SPTLC2	14q24	Dominant	>20 years	Acromutilation
ID	ATL1	14q11	Dominant	Early adult	Sensory loss Acromutilation
IE	DNMT1	19p13	Dominant	16 to 35 years	Sensory loss Deaf; Dementia
IF	ATL3	11q13	Dominant	14 to 35 years	Sensory loss
IIA	WNK1 /HSN2	12p13	Recessive	Congenital, or Early childhood	Sensory loss Acromutilation
IIB	FAM134B	5p15	Recessive	Childhood	Sensory loss Acromutilation
IIC	ATSV (KIF1A)	2q37	Recessive	6 to 15 yrs	Sensory loss Acromutilation Weakness, Distal
IID	SCN 9A	2q24	Recessive	Congenital	Absence of pain Autonomic
III	IKBKAP	9q31	Recessive	Congenital	Riley-Day
IV	NTRK1	1q21	Recessive	Congenital, or Early childhood	Sensory neuropathy Anhidrosis
V	NGF-β	1p13	Recessive	Early childhood to Adult	Absence of pain No anhidrosis

Table 1.1. Types of Hereditary Sensory Neuropathy (HSN), the associated genes and clinical findings (cont.)

VI	Dystonin	6p12	Recessive	Congenital	Absent tears
VII (Absent pain) Episodic pain	SCN 11A	3p22	Dominant	Congenital	Absence of pain Hyperhidrosis; GI Δ
Absent pain, Congenital	Not known	Not known	Recessive	Congenital, or Early childhood	Absence of pain No anhidrosis
VIII Absent pain & temperature Congenital	PRDM12	9q33	Recessive	Congenital	Absent pain Sweat reduced
Insensitivity to pain	ZFHX2	14q11	Dominant	Childhood	Absent pain Sweat reduced
Absent pain, Congenital	CLTCL1	22q11	Recessive	Congenital	Absent of pain & Light touch
Erythromelalgia	SCN 9A	2q24	Dominant	Childhood	Pain, distal Episodic
Erythromelalgia	NMNAT2	1q25	Recessive	Childhood	Pain, distal Episodic
Biemond ataxia	Not known	Not known	Dominant	19 to 30 years	Sensory loss
Ataxic Neuropathy	PLD3	19q13	Dominant	After 4th decade	Sensory loss Ataxia
Ataxic neuropathy 2	Not known	16q22	Dominant	After 2nd decade	Sensory loss Ataxia
Ulcero-Mutilation	Not known	Not known	Dominant	5 to 30 years	Acromutilation
Spastic paraparesis	CCT5	5p15	Recessive	1 to 5 years	Acromutilation

Table 1.1. Types of Hereditary Sensory Neuropathy (HSN), the associated genes and clinical findings (cont.)

Thermoanalgesia	Not known	Not known	Dominant	5th decade	Absent pain & temperature; Ataxia
PCARP	FLVCR1	1q31	Recessive	Early childhood	Proprioceptive loss Visual loss
SNAX1	RNF170	8p11	Dominant	3rd to 8th decade	Proprioceptive loss
HSAN + CNS	Prion	20p13	Dominant	Early adult	Autonomic & Sensory loss
CMT 2V: SN + Pain	NAGLU	17q21	Dominant	Adult	Pain; Ataxia
Autonomic	TECPR2	14q32	Recessive	Congenital	Autonomic
Scoliosis	PIEZO2	18p11	Recessive	Congenital	Sensory

1.3. Cerebellar Ataxia with Neuropathy and Bilateral Vestibular Areflexia Syndrome (CANVAS)

While we were focused on finding a novel gene for HSN using the samples of our family members by WES analysis, the index patient was hospitalized because of worsening of his condition. He was re-diagnosed as having cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome, known also as CANVAS. CANVAS is a slow progressive sensory ganglionopathy with a cerebellar impairment, affecting the dorsal and vestibular root ganglia (Burke and Halmagyi, 2018). Figure 1.2 shows the mentioned ganglion nerves and a difference in ganglions of normal and CANVAS-affected patient. A visual representation of clinical spectrum of CANVAS is given in 1.3 that can also explain the initial clinical diagnosis being as sensory neuropathy. Cerebellar integration of proprioceptive information received from visual system by sensory neurons is a crucial necessity for a successful motor coordination. When one component of this system fails, an ataxia is generated. Therefore, imbalance and falls are among the

observed symptoms in the patients due to loss of this coordination by cerebellum. Orthostatic hypotension, constipation, impotence, impaired sweating and chronic cough are part of the symptoms in patients with autonomic involvement (Cortese et al., 2019). The exact prevalence of CANVAS is estimated to be very low as one review reported 51 cases over a decade (Szmulewicz et al., 2015) and another review estimates 6-7 new cases per year (Ahmad et al., 2018). Mode of inheritance is suggested to be autosomal-recessive; however, autosomal-dominant inheritance with an incomplete penetrance is also possible (Rafehi et al., 2019; Szmulewicz et al., 2014). Ahmad et al. (2018) , for example, presented a British family with presumed autosomal dominant inheritance of CANVAS with incomplete penetrance and variable expression. In that family, a missense mutation in *ELF2* gene (which also resides in chromosome 4 just like *RFC1* gene, see below) increased *ATXN2* and reduced *ELOVL5* gene expression, the causal genes of type 2 and type 38 spinocerebellar ataxias. Therefore, it provides an additional link to network of ataxias (Ahmad et al., 2018).

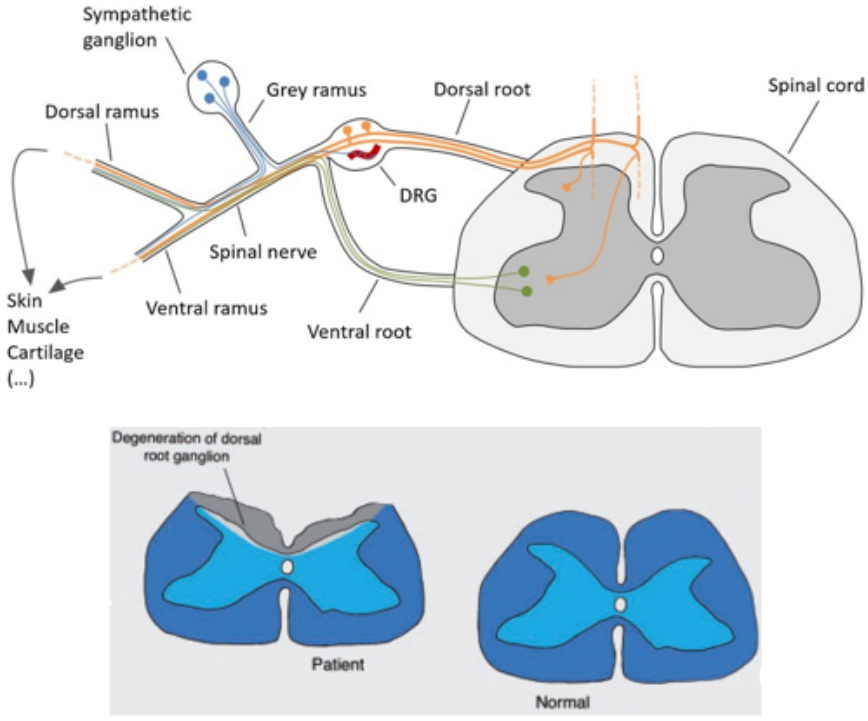

Figure 1.2. Affected dorsal and vestibular root ganglia nerves in CANVAS along with the comparison of ganglia of normal and CANVAS-affected patient. Bottom part of the figure is taken from the work of Paisán-Ruiz and Jen (2020).

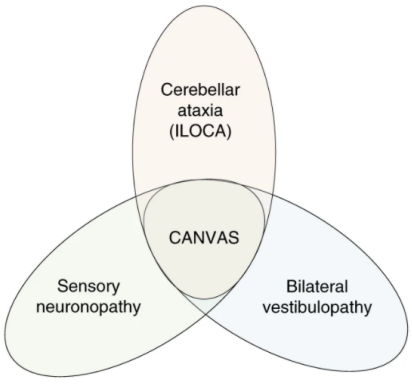

Figure 1.3. Clinical spectrum of idiopathic late-onset ataxia from isolated cerebellar, vestibular, and sensory variants to CANVAS (Cortese et al., 2019).

Up until recently, genetic mechanism underlying CANVAS remained unknown but suspected to result from pathogenic repeat expansions. It turns out to be correct as the two newly published papers by Rafehi et al. (2019) and Cortese et al. (2019), associates an intronic repeat expansion of AAGGG pentamer in the replication factor C subunit 1 (RFC1) gene as the cause of CANVAS. *RFC1* gene has a role in DNA replication by activating DNA polymerases and coordinating the synthesis of new DNA strands during replication after a DNA damage (Boesch and Nance, 2020). The expansion differs from eleven repeats of AAAAG allele in reference genome by copy number and nucleotide sequence. The expansion size is generally stable within families and differs from 400 to 2000 while most of cases are around 1000 copies. It is inherited as an autosomal-recessive disease. An ancestral haplotype study is conveyed based on WES data (Rafehi et al., 2019) and interestingly most of the patients were found to have a common core haplotype around the *RFC1* gene.

Molecular mechanisms of CANVAS have not been fully unraveled yet. However, pathogenic repeat expansion resides in Alu element, altering its activity. Dysregulation of Alu elements are known to disrupt the homeostasis of mitochondria in brain leading to neuronal stress and death (Paisán-Ruiz and Jen, 2020).

1.4. Whole Exome Sequencing (WES)

Genetic diagnosis of CMT is as important as its clinical diagnosis since neuropathies can also be acquired due to metabolic conditions, injury or even chemotherapy. Most important of all, most of the CMT types overlap with each other in terms of phenotypic features. Hence, it is generally hard to guess the CMT gene responsible for the disease in each patient directly by evaluating the clinical findings. Targeted Sequencing (TS) is the method of choice for genetic diagnosis of CMT that can screen almost all CMT genes at once. Whole Exome Sequencing (WES) or Whole Genome Screening (WGS) are used for identification of novel CMT genes when pathogenic mutation cannot be identified in known CMT genes by TS.

We have used WES approach in this study to identify the causative gene in our HSAN family. WES covers the protein-coding regions of the human genome that constitutes about 1% of the whole genome (Bamshad et al., 2011). It enables the coverage of all causative genes for determining the type of CMT in addition to the identification of novel causative gene variants. Next generation sequencing was first introduced in mid-2000s. Since it offers a considerable speed as well as reduced cost, it had a revolutionary effect in disease studies. WES with an exome coverage of 95% (Lelieveld et al., 2015), can detect not only the previously identified disease gene variants, but also identify new causative genes. Novel genes are discovered in an accelerated way with WES approach since its first application in 2008. Technical difficulties of initially introduced WES was resolved until 2011. This approach has a diagnostic yield of 45% if clinical features of a patient are also available for analysis (Pipis et al., 2019). CMT disease is one of the diseases that benefits most from WES as the number of genes associated with CMT is continuing to increase in time by use of this technique (Pareyson et al., 2017). WES contributes to personalized medicine with its integration into healthcare. Nowadays, it has a multipurpose usage ranging from genetic diagnosis to decisions for treatment (Bertier et al., 2016).

1.5. Homozygosity mapping

Occurrence of recessive traits is more often in the siblings of consanguineous marriages than in general population. A common disease-causing allele from a recent ancestor is inherited to the inbred children from both of his parents. Such phenomenon, called homozygosity by descent (HBD) is used as the basis of a powerful technique of genetic mapping to successfully identify the disease-causing gene in consanguineous families (Kruglyak et al., 1995). This technique is known as homozygosity mapping and is used to determine the molecular defect by finding regions of HBD. These regions of chromosomes are shared between affected siblings of consanguinity that lead to rare diseases with autosomal recessive inheritance. The frequency of disease-causing mutations found in blocks of homozygosity is directly correlated with the percentage of inbreeding and changes from 0.4% (third-cousin parents) to 12.5% (double first cousin

parents) (Alkuraya, 2010). With classification of the homozygous regions in affected individuals, homozygosity mapping allows limiting the number of candidate loci and allows researcher to focus on small number loci in rare diseases with autosomal recessive inheritance. Homozygosity mapping based on WES data (HOMWES) was performed in this study for our family where the consanguineous parents were second cousins, using the PLINK bioinformatics tool.

2. PURPOSE

The main purpose of this study is to identify a novel disease-causing gene responsible for Hereditary Sensory Neuropathy in a family from Turkey.

HSN predominantly affects sensory nerves and is a very rare disease. More than 24 causative gene variants and seven chromosomal loci have been associated with HSN indicating its high genetic heterogeneity. HSN Type 1 (HSN1) was localized to chromosome 3 previously in two Australian families, however, the causative gene has not been identified, yet. The family, investigated in this study presents cough and gastrointestinal reflux associated with HSN that have also been reported in the Australian families. These findings indicate that the family can be very informative and helpful for the identification of the gene responsible for HSN1. The only difference between these families was that the reported the disease was autosomal dominant but presence of consanguinity and three affected siblings suggests autosomal recessive inheritance in our family.

We have used whole exome sequencing (WES) and homozygosity mapping methods in the family samples to identify the causative gene.

In terms of its benefit to the families, the genetic diagnosis is highly critical for neurological disorders since they provide evidence for its heritability and determines the choice of medical treatment. Besides, it is critical for differential diagnosis because clinical symptoms overlap highly in those disorders complicating the clinical diagnosis.

In terms of its contribution to literature, identification of novel genes might enlighten the underlying genetic mechanisms that cause the disease. One more minor step will be achieved in the way of unraveling the disease mechanisms and the ultimate goal is to provide an insight to the possible future treatments of the disease.

3. MATERIALS

3.1. Project Family

The family P1403, investigated in this particular study was diagnosed as hereditary sensory neuropathy (HSN). Along with HSN, cough and gastroesophageal reflux (GER) have been reported as additional symptoms in the clinical findings. Pedigree of this family shown in Figure 3.1 suggested autosomal recessive inheritance with variable expression.

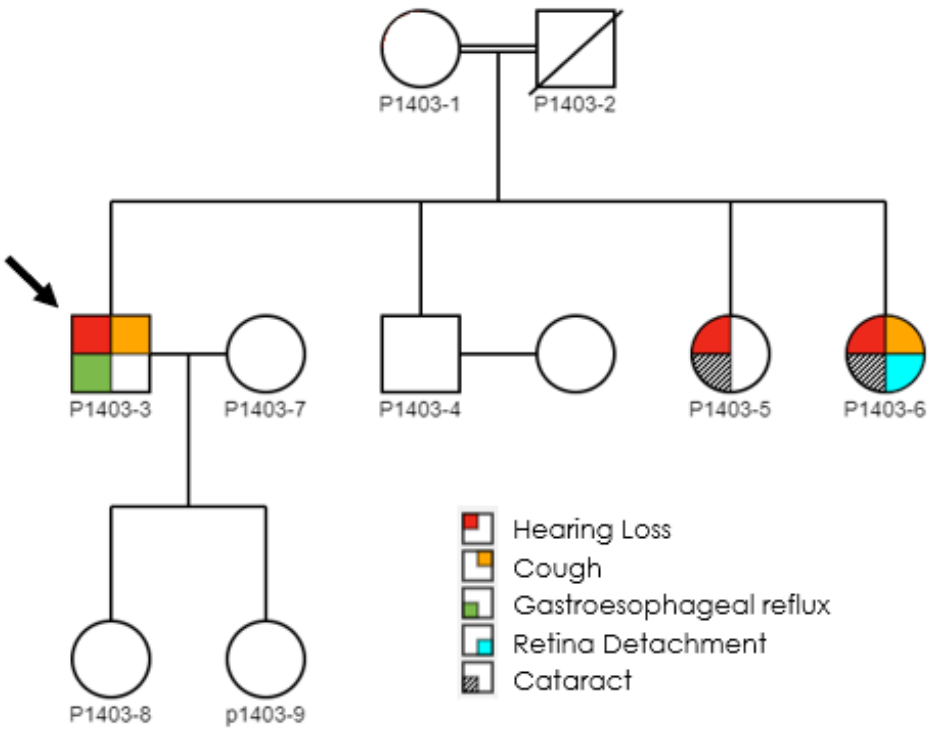

Figure 3.1. Pedigree of the project family suggests variable expression. Arrow indicates the index case.

3.2. Primers

Primer sequences for candidate genes, selected from WES data of the family, were designed using Primer3 software and are given in the Table 3.1.

Table 3.1. Primer sequences used for amplification of the candidate genes in this study.

Candidate gene	**Exon**	**Primer**	**Sequence of forward (F) and reverse (R) primers**
ARSD	Exon 5	ARSD_ex5_F	GCGACGCTGGAACTGTATC
		ARSD_ex5_R	GCACCTGGGCTAAAAGATCC
CLASP2	Exon 30	CLASP2_ex30_F	ACCTAATGTTTGTGTGGCTTCC
		CLASP2_ex30_R	TGGAATTTACAGCCAAGGGT
FCRL2	Exon 4	FCRL2_ex4_F	TGTCACTGCTTTGGGGACTA
		FCRL2_ex4_R	TCAAGGACAGGAGTTGAACAA
FMO2	Exon 6	FMO2_ex6_F	GGAACAGTGGAGGAGAACATT
		FMO2_ex6_R	TGCCAGTTGAAATTCTCTTGACT
FMO2	5'UTR	FMO2_5UTR_F	TGTTGACCTCAGAGCACGAG
		FMO2_5UTR_R	CACAGCTACCTTCTTTGCCA
MPZL1	Upstream	MPZL1_upst_F	AGACAATGAAGAGGACCACG
		MPZL1_upst_R	TGCTTGGGGAAGAGGCTT
MYH7B	Exon 17	MYH7B_ex17_F	CTTTCCCCATCTGCGAGAG
		MYH7B_ex17_R	CTCAAAGTGACCTTGCCGC
MYH14	Intron	MYH14_Int_F	GCTGAGCAGCAAGGATGAC
		MYH14_Int_R	GGCTCTCCTTACTTTCTATGCG
POGK	Downstream	POGK_downst_F	TCCTAAGCTCCTTACCACCTC
		POGK_downst_R	AATTTCATCCCCAGCTGTTTAA
RBMX	Exon 8	RBMX_ex8_F	AAAACCTCCCACCCAGTTG
		RBMX_ex8_R	TGGAGGGGAACTTAACAGGG
RFC1	Intron	CANVAS_RFC1_3F	ACTGACAGTGTTTTTGCCTGT
		CANVAS_RFC1_3R	GGCTGAGGCAGGAGATTCAC

Table 3.1. Primer sequences used for amplification of the candidate genes in this study. (cont.)

Candidate gene	**Exon**	**Primer**	**Sequence of forward (F) and reverse (R) primers**
SLC25A5	Exon 2	SLC25A5_ex2_F	TGGCTTCCTTCCTGTCTGTT
		SLC25A5_ex2_R	AAGTAGGCGGCTCGGTAGAT
TAB3	Exon 7	TAB3_ex7_F	TGAAGACTACGAAGGGGCTC
		TAB3_ex7_R	GACCACAAAGCCATTCCTC

3.3. Kits, Databases and Softwares

Whole Exome Sequencing (WES) was performed using Illumina NextSeq 500 and Agilent SureSelect V6 with UTR regions Capture kits with a minimum coverage of 50X. For the purpose of analyzing exome data efficiently, Genome Analysis Toolkit (GATK, "n.d") program, SUSPECTS ("n.d"), ToppGene ("n.d") and Endeavour ("n.d") algorithms were used. Allele frequency of the variants in the population were investigated through Genome Aggregation Database gnomAD ("n.d") and Exome Aggregation Consortium ExAC ("n.d") databases. For reaching the sequences of specific gene candidates whilst designing primers, Ensembl Genome Browser (Ensembl, "n.d") was governed. To be able to understand the effects of mutations on protein functions, SIFT ("n.d") and PolyPhen ("n.d") softwares were used. Mutation Taster (MutationTaster, "n.d") online program was used to determine the possible effect of a variant on its function. For splicing region variants, Human Splicing Finder (HSF, "n.d") algorithm was used for identifying their potential impacts on transcription. SnapGene Viewer (version 4.3.5) was used for analyzing Sanger sequencing results.

3.4. Buffers and Solutions

The buffers, gels and solutions used are listed in the Tables 3.2, 3.3 and 3.4, respectively.

Table 3.2. Buffers used in this study.

Buffer	**Recipe**
Nuclei Lysis Buffer	400 mM NaCl 2 mM EDTA (pH 7.4) 10 mM Tris-HCl (pH 8.0) Diluted to 1L with dH_2O
Red Blood Cell (RBC) Lysis Buffer	155 mM NH_4Cl 10 mM $KHCO_3$ 1 mM EDTA (pH 7.4) Diluted to 1L with dH_2O
Tris – Boric Acid – EDTA (TBE) Buffer (10X)	20 mM EDTA (pH 8.3) 0.89 M Tris-Base 0.89 M Boric Acid
Tris – EDTA (TE) Buffer	20 mM Tris-HCl (pH 8.0) 1 mM EDTA (pH 8.0)

Table 3.3. Gels used in this study.

Gel	**Recipe**
Agarose Gel (1%)	1% (w/v) Agarose in 0.5 X TBE Buffer
Polyacrylamide Gel (10%)	30% Acrylamide:Bisacrylamide (29:1) 12% 5X TBE 1% TEMED 0.1% APS

Table 3.4. Solutions used in this study.

Solution	**Recipe**
Acrylamide Solution (30%)	29% acrylamide 1% N, N'-methylenebisacrylamide
$AgNO_3$ solution	0.1% $AgNO_3$ in dH_2O
Ammonium Persulfate (APS)	10 percent APS (w/v) in dH_2O
Denaturing Loading Dye	95% formamide 20 mM EDTA 0.05% Xylene Cyanol 0.05% Bromophenol Blue
Deoxyribonucleotide Triphosphates (dNTPs)	100 mM of each dNTP in dH_2O
0.5 M EDTA	9.3 g EDTA in 50 ml dH_2O
Ethidium Bromide (EtBr)	10 mg/ml in dH_2O
0.25 M HCl	41.66 ml HCl to dH_2O for 2 L in total
NaCl solution (5 M)	292.2 g NaCl in 1 L dH_2O
0.25 M NaOH	20 g NaOH in 2 L dH_2O
Proteinase K	20 mg/ml in dH_2O
Sodiumdodecylsulfate (SDS) (10%)	10% SDS (w/v) (pH 7.2)

3.5. Chemicals and Disposables

Chemicals in the form of liquid and solid and disposable materials that were used in this study are listed in Table 3.5.

Table 3.5. Chemicals and disposables used in this study.

Chemical-Disposible	**Supplier Firm**	**Catalog Number**
Acrylamide	Sigma-Aldrich	A3553
AgNO3	Merck Millipore	101512
Ammonium peroxodisulphate (APS)	Fluka	9914
Boric Acid	Sigma-Aldrich	B6768
Bromophenol Blue (BPB)	Sigma-Aldrich	B5525
Dimethylsulfoxide (DMSO)	Sigma-Aldrich	M81802
DNA Ladder (100 bp/1kb)	ThermoFisher Scientific	SM0241/SM0311
dNTP	ThermoFisher Scientific	10297018
EDTA	Riedel-de Haen	34549
Formaldehyde solution	Sigma-Aldrich	15512
K2EDTA blood tube	BD Vacutainer	367844
KHCO3	Sigma-Aldrich	S5761
N, N'-methylenebisacrylamide	Sigma-Aldrich	T7024
NaCl	Sigma-Aldrich	S7653
NaOH	Sigma-Aldrich	6203
NH4Cl	Santa-Cruz Biotechnology	254134
Pasteur Pipettes	Isolab	
Pipette Tips (10 ul / 200 ul / 1000 ul)	Axygen Scientific	4826 / 4845 / 4713
Sodium Dodecyl Sulphate (SDS)	Sigma-Aldrich	L3771
Taq Polymerase	ThermoFisher Scientific	EP0402
Xylene Cyanol	Sigma-Aldrich	X4126

3.6. Equipment

All equipment used in the study are given in Table 3.6.

Table 3.6. Equipment used in this study.

Equipment	**Supplier Firm**
Centrifuges	Eppendorf Centrifuge 5415R, Beckman Coulter Allegra X-22R Centrifuge
Deep Freezers	-20°C (Bosch), -20°C 2021D Arçelik
Documentation Systems	GelDoc Documentation System, Bio-Rad
Incubator	Nüve EN400
Micropipettes	Gilson
Microwave oven	Arçelik
Power supply	PowerPac Basic, Bio-Rad
Thermocycler	Bio-Rad
Refrigerator	Arçelik
Shaker	SL350, Nüve
Spectrophotometer	Nanodrop ND-1000 Spectrophotometer, ThermoFisher Scientific
Vortex	Nuvemix NM110, Nüve
Water Purification	WA-TECH Ultra-Pure Water Purification System, WA-TECH

4. METHODS

4.1. Pedigree Analysis

The focus of this study was a family with the clinical diagnosis of HSN along with cough and GER. All these symptoms were compatible with the symptoms of HSAN1B form of the disease for which the locus was known but the disease-causing gene has not been identified, yet. Parents were first degree cousins, had three affected and one unaffected children (Figure 3.1) that suggested autosomal recessive inheritance opposed to reported dominant HSAN1B. The index case P1403-3 had congenital hearing loss, cough and gastroesophageal reflux while the affected sisters P1403-5 and P1403-6 not only had congenital hearing loss but also retinal detachment (cataract). Additionally, one of the sisters, namely P1403-6, had cough just like the index case and seemed to have a more severe phenotype than sister P1403-5. The index case also described dizziness whilst standing up and progressive imbalance problem. The distribution of different symptoms in different affected individuals in the family indicated that the disease gene may have a variable expression.

4.2. DNA Extraction from Peripheral Blood

Blood samples of ten ml were taken into a K_2EDTA tubes from family members. It was transferred into sterile fifty ml centrifuge tubes in the laboratory. Thirty ml of red blood cell (RBC) lysis buffer was added and each tube was vortexed to have a homogenous mix. It allows preferential lysis of RBCs from whole blood and concentrating nucleated white blood cells (WBC). After 4°C incubation for 25 minutes, the blood-buffer sample mixes were centrifuged at 5000 revolution per minute (rpm) for 10 minutes at 4°C. The supernatant was discarded, and ten ml of RBC lysis buffer was added to the pellet. After dissolving the pellet, centrifugation was repeated at 4°C and 5000 rpm for 10 min. Resulting supernatant of the centrifuged samples was discarded, and pellet was washed with RBC lysis buffer several times without harming

the integrity of the pellet. The pellet was re-suspended with 5 ml of nuclei lysis buffer, 40 μl Proteinase K and 50 μl of 10% SDS solution. Samples were incubated at 37°C overnight. Next day, 10 ml of 2.5 M NaCl was added to the overnight solution and vortexed. Samples were centrifuged at 5000 rpm at room temperature (around 22°C) for 30 minutes. Supernatant of the centrifuged samples were transferred into a new sterile 50 ml tube and ice-cold absolute ethanol of two volumes was added for precipitation of DNA. Precipitated DNA was fished out and taken into 1.5 ml centrifuge tube. 200-300 ml of Tris-EDTA (TE) buffer was added to air dried DNA. Samples were left in a shaker at the slowest rate for overnight at room temperature to dissolve the DNA completely. Genomic DNA samples were stored at -20°C.

4.3. Quantitative Analysis of Extracted DNA

DNA concentration of the samples were measured using NanoDrop ND 1000 spectrophotometer at 260 nm wavelength. To better understand the sample purity, absorbance at the wavelengths of 230 and 280 nm were also measured. Whenever the A260/A230 ratio was higher than 2, and the ratio of A260/A280 was higher than 1.8 but lower than 2, the DNA sample was accepted as pure and ready for further studies. DNA concentration of 100 ng/μl was optimal for polymerase chain reaction (PCR) and 50 ng/μl samples were used for whole exome sequencing (WES).

4.4. Polymerase Chain Reaction (PCR)

For designing of the primers used in this study (Table 3.1), Primer3 ("n.d") online software was used. Gene Tool's SNPCheck ("n.d") algorithm was utilized to analyze whether the primer sequence coincides with any single nucleotide polymorphism (SNP) region and the primer hit any regions with 100% or partial accuracy other than the targeted region. 100% accuracy of the primer hit region should be only for the DNA region of interest, partial hits were acceptable unless it was close to 100 or more. Multiple Primer Analyzer of ThermoFisher ("n.d") Scientific was used for detecting primer dimer or self-dimer presence and compare between melting temperatures (T_m) of

forward and reverse primers. Maximum of 2°C between T_m's of primers was considered to be suitable. *In-silico* PCR tool of USCS Genome Browser (Genome, "n.d") was used for checking the virtual results of PCR to be only in the targeted region of the correct chromosome and the melting temperatures.

PCR reactions were prepared using 4 μl of 100 ng DNA sample, 5 μl 10X polymerase buffer,4 μl of 25 mM $MgCl_2$, 1 μl of 10 mM dNTPs, 1 μl of 10 pmol forward primer, 1 μl of 10 pmol reverse primer, 0.2 μl of Taq Polymerase, and 33.8 μl dH_2O, to a total volume of 50 μl. These reactions were carried out in a thermal cycler using the following program; an initial denaturation step at 95°C for 5 minutes, then 20 cycles of three short steps; 95°C for 30 seconds, 30 seconds at the annealing temperature (64°C, by decreasing of 0.5°C at every cycle), 30 seconds at 72°C. After that, 30 cycles of three short steps of 95°C for 30 seconds, 30 seconds at 54°C, and 30 seconds at 72°C and a final step of elongation at 72°C for 10 minutes was used. Finally, there was a cooling step to 4°C for 10 minutes. PCR products are stored at 4°C. A volume of 5 μl PCR product was loaded to 1% agarose gel to confirm PCR product size and amplification.

4.5. Whole Exome Sequencing (WES)

4.5.1. Initial Analyses

For whole exome sequencing (WES), samples were prepared with a concentration of 50 ng/μl and a volume of at least 10 μl. These samples were outsourced to an appropriate sequencing company. This analysis was repeated for five family members analyzed in the study, namely, P1403-1 (unaffected mother), P1403-3 (index case), P1403-4 (unaffected brother of index patient), P1403-5 (affected sister) and P1403-6 (affected sister). WES was performed using Agilent SureSelect V6 with UTR regions capture kit for the index case P1403-3 with a minimum coverage of 50X of 100 bp length reads. This sequencing was outsourced to Macrogen in South Korea. For the rest of the family members (P1403-1, P1403-4, P1403-5 and P1403-6) the samples were outsourced to DNA Laboratuvarları, İstanbul. Exome capture was performed

with Illumina NextSeq 500 Capture kit and sequencing of 100 bp read lengths was achieved with a minimum 50X coverage. BWA (Burrows-Wheeler Alignment Tool) was used for sequence alignment. GATK software was used for variant calling. For variant annotation, SnpEff software was utilized. Initial analyses were performed by respective service providers. Results were delivered to the laboratory in the excel, vcf and fastq forms.

4.5.2. Data Analysis

For the first step of WES data analysis, variants in known neuropathy genes that were previously reported in Neuromuscular Disease Center page of Washington University NeuroMuscular ("n.d") were investigated for each individual. Five % minor allele frequency (MAF) at maximum was the only filtering criteria for identifying the known variants in known genes. Variants were then classified according to their inheritance patterns. For a recessive inheritance, the variant should be in homozygous state in the affected individuals and in heterozygous state in the parents. The unaffected sibling can be heterozygous or homozygous for the wild type allele. For a dominant inheritance, heterozygosity for the variant in affected individuals throughout the family and homozygosity for the wild type allele for the unaffected individuals was expected to confirm the responsibility of the candidate variant.

When previously reported gene variants could not be identified in members of the project study, search for a novel disease-causing variant in known genes was targeted. Filtering criteria were as follows; variant to be missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant, transcript type to be coding with a minor allele frequency (MAF) of 1% at maximum in the population and read depth to be 30 at minimum. If a variant could not be identified by these criteria, filtering was made more flexible; MAF in the population was reduced to 5% at maximum and read depth was decreased to 10 at minimum. SIFT ("n.d") and PolyPhen ("n.d") prediction softwares were used to prioritize the variants. Predictions of variants should be deleterious (D) (SIFT) or possibly/probably damaging (PsD and

PrD, respectively) (PolyPhen). Candidate variants identified after filtering were analyzed exclusively on Genome Aggregation Database, gnomAD ("n.d"), by entering the gene name first and finding that specific variant after. In addition to the population frequencies, number of homozygous individuals in various populations were also searched in the same databases. If homozygous individuals were reported previously, variant of interest was eliminated since it is highly unlikely to have other HSAN patients in databases of normal individuals and HSAN1B has an extremely low frequency in the population.

Due to consanguinity among parents, data analysis was initially performed according to autosomal recessive inheritance. For this purpose, candidates were selected that showed homozygosity in the affected members (P1403-3, P1403-5 and P1403-6) and heterozygosity in unaffected mother (P1403-1) while the unaffected sibling (P1403-4) can be either heterozygous or have the wild type allele for the variant in question. A coding algorithm was established for selecting the common homozygous variants in the affected patients, namely P1403-3 and P1403-6, through Phyton coding program. The algorithm also considered the variants to be heterozygous or not present in the unaffected brother P1403-4. Since variable expression, incomplete penetrance, and compound heterozygosity were also suspected, autosomal dominant inheritance for a possible disease-causing variant was analyzed. In compound heterozygosity more than one variant reside in the same gene and can cause the disease. Such variants were investigated in a way to be present in heterozygous state in affected patients (P1403-3, P1403-5 and P1403-6) and not present at all in the unaffected ones (P1403-1 and P1403-4).

4.6. Sanger Sequencing

For confirmation of segregation of selected candidate variants traditional Sanger sequencing was used. For this purpose, primers were designed as forward and reverse primers and used for amplification of the variant DNA regions in all available members of the family. After confirming the size of the PCR product by running on a 1%

agarose gel, PCR products of 50 μl volume were outsourced for direct sequencing to Macrogen, Amsterdam. Sequencing results were delivered as ab1 files and analyzed using SnapGene Viewer program version 4.3.5.

4.7. Homozygosity Mapping based on WES analysis (HOMWES)

Homozygosity mapping was used to identify the shared homozygous chromosomal regions among affected individuals (P1403-3, P1403-5 and P1403-6) on the basis of the hypothesis that the chromosomal region that is carrying the responsible variant was inherited from a common ancestor in the family. It is an effective approach for identifying molecular defects especially in the families with consanguinity. By identifying the homozygous regions in affected individuals, it allows limiting the number of candidate loci and allows researcher to focus on those small number loci. WES data can be used to perform homozygosity mapping with the help of some bioinformatics tools. PLINK program was utilized for HOMWES, using the vcf files of WES data. Homozygous regions defined as runs of homozygosity (ROH) which can be up to dozens of Centimorgans are identified separately for affected individuals. Results were obtained as excel files. While analyzing the data, ROH lengths were considered to be appropriate for gene identification when equal or higher than 6 Mb and satisfactory when around 5 Mb and unlikely when 4 Mb or lower. If there is less than 1000 kilobases between two ROHs, regions can be combined and considered as one homozygous region with longer length. This filtering was performed for each pair of patients. The identified ROHs, after use of these filtering criteria, were paired between patients and common regions were analyzed deeply in WES data results. The filtering criteria used to identify the causative variants in these shared ROHs in the patients' WES data were as follows; the variant should be homozygous, its minor allele frequency should be 5% at maximum in the population, variant function can be missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant and the type of transcript analyzed should be coding. Previously mentioned SIFT and PolyPhen algorithms, and gnomAD were used for the same purposes as described before to identify the causative variant.

5. RESULTS

5.1. Known Neuropathy Gene Screening using WES

Genomic DNA samples of five individuals from the HSAN family (P1403-1, P1403-3, P1403-4, P1403-5 and P1403-6) were investigated using whole exome sequencing (WES). A summary of WES results indicating number of different variant types in the index patient P1403-3 is given in Table 5.1. A similar description of the WES results for other investigated individuals are given in the Appendix A. We have screened the WES data for variants in genes known to be responsible for neuropathy, initially. For this purpose, a total of 202 genes were screened that cause neurological disorders including CMT, HSN, HMN, Spastic Paraplegia, and Ataxias. Filtering criteria used in this initial analysis was not strict and we have eliminated only the variants with at most 5% minor allele frequency (MAF). Variants remained after elimination of the data with this only criterion are listed in Table 5.2.

Table 5.1. Number of different variants identified in WES data of patient P1403-3.

Coding variants:	135,434
Non-coding variants:	6,595
Intron variants:	72,586
3'UTR variants:	25,934
Synonymous variants:	11,230
Missense variants:	11,086
Intergenic variants:	6,118
5'UTR variants:	5,551
Upstream variants:	5,316
Downstream variants:	4,724
Splice region variants:	2,576
Non-coding exom variants:	1,984
Disruptive inframe insertion & deletion variants:	246

Table 5.1. Number of different variants identified in WES data of patient P1403-3. (cont.)

Splice acceptor & donor variants:	166
Inframe insertion & deletion variants:	125
Stop gained variants:	99
Frameshift variants:	42
Stop lost variants:	40
Start lost variants:	22
Exon-loss variants:	1
Total number variants:	148,147

Table 5.2. Candidate gene variant list after filtration of WES data of patient P1403-3 for genes related to neurological diseases.

Gene	Chromosome	Zygosity	Annotation	Nucleotide Change	Amino acid Change	dbSNP
HK1	chr10	HET	Intron	c.2622-88delA	.	rs541953079
EBF3	chr10	HET	Intron	c.1345+192A>T	.	rs72837145
SBF2	chr11	HET	Upstream	c.-139C>T	.	rs532433256
SPTLC2	chr14	HOM	3'UTR	c.*4131G>A	.	rs45577834
SPTLC2	chr14	HET	Intron	c.1304-14365C>G	.	rs147204783
VRK1	chr14	HET	Intron	c.1159+43C>G	.	rs45457797
ZFHX2	chr14	HET	Missense	c.4078C>G	p.Gln1360Glu	rs566304076
SLC12A6	chr15	HET	Intron	c.2436+142T>G	.	rs111617676
SLC12A6	chr15	HET	Intron	c.2042+45C>A	.	rs112899095
SLC12A6	chr15	HET	Intron	c.1592-50A>G	.	rs17236791
MYH14	chr19	HOM	Intron	c.4752+197G>A	.	rs111403438

Eight of these eleven variants shown in Table 5.2 were excluded from the study since they were in introns. The segregation of the other three gene variants, *SBF2*, *SPTLC2* (3'UTR variant) and *ZFHX2* in the family, were analyzed according to their segregation pattern in the family. *SBF2* gene variants were known to cause the disease when inherited autosomal recessively. Since our patient P1403-3 was heterozygous for

the variant, we speculated that it cannot be responsible from the symptoms observed in the family. *SPTLC2* (3'UTR) gene variants were known to cause the disease when inherited autosomal dominantly. The patient was homozygous for the variant, but, the other members of the family, including other affected members, were carrying the wild type allele leading to exclusion of the variant. Conditions for *ZFHX2* gene variant was the same with *SPTLC2* 3'UTR variant except that the patient was heterozygous. As a summary, unfortunately, we could not identify any causative variant in genes responsible for neurological disorders. Hence, a novel variant search began to unravel the genetic cause of the symptoms in the family.

5.2. Search for a Novel HSAN Gene

5.2.1. Investigations for a Possible Autosomal Recessive Inheritance

The clinical symptoms in our family were compatible with HSAN Type 1B (HSAN1B) for which the causative gene has not been identified, yet. However, Kok et al. (2003) studied two Australian families with similar clinical symptoms and mapped the responsible gene to chromosome 3p22-p24. They also stated presence of autosomal dominant mode of inheritance in their families. Since parents in our family were consanguineous and had three affected siblings, autosomal recessive inheritance was more likely than dominant inheritance. Therefore, WES data results were analyzed primarily according to autosomal recessive inheritance. Affected members of the family should be homozygous for the candidate variant. Unaffected mother should be heterozygous because she must have passed on one of the causative alleles. The unaffected brother was expected to be homozygous for the wild type allele or have that variant in heterozygous state. We have used a coding algorithm (designed for this purpose by data scientist, Özcan Gündeş) for selecting only the variants matching this inheritance criterium. We used further filtering criteria for the selected items; variants that were missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant, assigned to a coding transcript type, maximum 5% minor allele frequency (MAF) and a minimum read depth of 30 were considered for analyses. The

variants that matched this filtering criteria are listed in Table 5.3.

Table 5.3. Candidate gene variant list after filtration of WES data of five family members for autosomal recessive inheritance. (Zygosity of the variants in the index patient P1403-3.)

Gene	**Chromosome**	**Zygosity**	**Annotation**	**Nucleotide Change**	**Amino acid Change**	**dbSNP**
FMO2	Chr 1	HOM	Missense	c.1172G>C	p.Arg391Thr	rs28369899
GREB1	Chr 2	HOM	Missense	c.3428G>A	p.Gly1143Asp	rs145454387
TLR10	Chr 4	HOM	Missense	c.1573C>T	p.Arg525Trp	rs11466658
TLR1	Chr 4	HOM	Missense	c.914A>T	p.His305Leu	rs3923647
HGFAC	Chr 4	HOM	Missense	c.640G>T	p.Gly214Cys	rs41264743
SLC15A3	Chr 11	HOM	Splice region & Intron	c.1108-5C>T	.	rs76312198
MYH7B	Chr 20	HOM	Frameshift Deletion	c.1612_1613del	p.Met538fs	rs571047145

As previously mentioned in Section 4.5.2, variants were prioritized using several more tools and algorithms such as SIFT and PolyPhen predictions and gnomAD database. The minor allele frequencies (MAFs), SIFT, PolyPhen, and Mutation Taster predictions, number of previously reported homozygous individuals (taken from gnomAD database) of these variants can be found in the following table (Table 5.4).

Table 5.4. Prioritization of autosomal recessively inherited candidate gene variants by using different algorithms and tools. (D: deleterious; T: tolerated; B: benign; PrD: probably damaging; PsD: possibly damaging.)

Gene	**Chromosome**	**SIFT**	**Polyphen**	**Mutation Taster**	**MAF Values**	**No of Reported Homozygous Individuals (gnomAD database)**
FMO2	Chr 1	D	PrD	Disease causing	0.011582	0
GREB1	Chr 2	T	PsD	Polymorphism	0.010982	18
TLR10	Chr 4	T	B	Polymorphism	0.027756	62
TLR1	Chr 4	D	B	Polymorphism	0.029153	60
HGFAC	Chr 4	D	PrD	Polymorphism	0.005591	17
SLC15A3	Chr 11	-	-	-	0.010383	14
MYH7B	Chr20	-	-	Disease causing	0.001211	0

While searching for a novel variant that may cause a rare disease, it is important to consider the fact that any possible variant responsible for this disease should also be rare in the population. Minor allele frequency (MAF) is an important indication in this sense that gives the frequency of the second most common allele in human population. A low MAF value strengthens the possibility that the identified variant can be responsible from the rare disease. Since HSAN is an extremely rare disease, it was expected in our study that MAF values of candidate variants should be below 1% in the population. None of the candidate variants listed in Table 5.4 except two (*FMO2* and *MYH7B*) met this criterion since we have initially used a maximum MAF value of 5%, not to miss any causative variants during our filtration. It is also of critical importance to find out the number of previously reported homozygous individuals for the candidate variants in the healthy human population, in human genome databases, especially in gnomAD. If the variant of interest is responsible from the disease, we expect these individuals to present the symptoms of the disease and they should not appear in genome databases of healthy human population. Only two candidate gene variants matched that measure, *FMO2* and *MYH7B*. SIFT, PolyPhen and Mutation Taster predictions also supported that these variants may cause a disease situation. Flavin-containing monooxygenase 2 (*FMO2*) gene codes for a NADPH-dependent enzyme which catalyzes N-oxidation of primary alkylamines. It has a major expression in

lungs, fat and esophagus and a minor expression in brain. Myosin heavy chain 7B (*MYH7B*) gene encodes a heavy chain of myosin II protein, a member of the motor-domain superfamily, which has a role in interaction with actin. Since this protein is involved with muscle contraction, it is thought that it might have a relation with the symptoms observed in family. We performed Sanger sequencing to confirm their segregation among family members. Therefore, primers were designed for *FMO2* and *MYH7B* variants and segregation studies were performed. Sanger chromatograms for *FMO2* gene variant are shown Figure B.1 and Sanger sequencing results are given in Table 5.5. Segregation chromatograms for *MYH7B* variant are listed in Figure B.2 and results are shown in Table 5.6.

Table 5.5. Sanger Sequencing Results for *FMO2* gene c.1172G>C variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***FMO2* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, +	var, +
P1403-3 (Index case)	var, var	var, var
P1403-4 (Unaff. brother)	var, var	var, +
P1403-5 (Aff. sister)	var, +	var, +
P1403-6 (Aff. sister)	var, var	var, var
Conclusion:	Excluded	

FMO2 gene variant (c.1172G>C) was eliminated due to differences between WES and Sanger sequencing results of the unaffected brother. This analysis revealed that unaffected brother P1403-4 also has the variant in homozygous state. This segregation pattern within the family led to exclusion of the *FMO2* gene variant from the study.

Table 5.6. Sanger Sequencing Results for *MYH7B* gene c.1612_1613del variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***MYH7B* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, +	var, +
P1403-3 (Index case)	+, +	var, var
P1403-4 (Unaff. brother)	var, +	+ , +
P1403-5 (Aff. sister)	var, var	var, var
P1403-6 (Aff. sister)	var, var	var, var
Conclusion:	Excluded	

For the *MYH7B* gene variant (c.1612_1613del), Sanger sequencing results of the index patient P1403-3, showed that he is homozygous for the wild type allele and does not carry the variant allele at all. There was also a difference in results of unaffected brother P1403-4, indicating that he is heterozygous for the variant. In conclusion, *MYH7B* gene variant cannot be causative since it was not segregating with the disease in the family.

In the interest of not missing any variant while considering appropriate family segregation, WES data of index patient P1403-3 was investigated independently. Filtering his data and following the steps mentioned in Section 4.5.2, revealed absence of a strong candidate gene.

We have also investigated the chromosome 3p22-p24 region in detail since it was previously reported as the potential responsible locus for HSAN Type 1B. In this targeted region there was a variant of *CLASP2* gene which met the filtering criteria but failed to segregate with the disease phenotype among family members. The gene variant of *CLASP2* was present in heterozygous state in index case and one of the

affected sisters (P1403-5). However, the other affected sister (P1403-6) had native alleles in that position, excluding this variant. Protein product of cytoplasmic Linker Associated Protein 2 (*CLASP2*) is involved in microtubule binding by enhancing the stabilization of dynamic microtubules. A similar stabilization function of *CLASP2* is at the kinetochore where the chromosomes align on the mitotic spindle. Most importantly, gene expression is highest in the brain. In order not to ignore the *CLASP2* gene variant because of misfit in just one individual, direct sequencing was performed for the affected sister P1403-6. Results of direct sequencing matched with WES data, and in conclusion, *CLASP2* gene variant was excluded. Chromatograms for *CLASP2* variant was given in Figure B.3. The data for the *CLASP2* variant and Sanger results are listed in Table 5.7. With this last investigation, analyses of above-mentioned gene variants were finalized and studies focusing on autosomal dominant segregation started.

Table 5.7. Variant information and Sanger Sequencing Result for CLASP2 gene c.3347C>T variant.

CLASP2 gene:	**Chromosome**	**Annotation**	**Nucleotide change**	**Amino acid change**	**dbSNP**	**Minor Allele Frequency**
	3	Missense	c.3347C>T	p.Thr1116Ile	rs187113660	0.0033
	Sanger Sequencing Result			**WES Data Result**		
P1403-6 (Affected sister)	+, +			+, +		
Conclusion:	Excluded					

5.2.2. Investigations for a Possible Autosomal Dominant Inheritance

Variable expression of clinical phenotypes that were present among family members were explained in detail in Section 4.5.2. The variable expression was observed at the level of both clinical symptoms and the severity of disease. Incomplete penetrance and compound heterozygosity could be other exceptional properties that could be present in the family. To unravel these possibilities, autosomal dominant segregation was investigated as the next step. Furthermore, autosomal dominant segregation was observed in two previously reported Australian HSAN1B families (Kok et al., 2003). The same algorithm used for autosomal recessive inheritance, was re-designed

to screen the WES data of the family for the variants that segregate in autosomal dominant fashion. For any possible disease-causing variant to be responsible from HSAN Type 1B, affected individuals (P1403-3, P1403-5 and P1403-6) should be heterozygous for that variant while the unaffected individuals should have the wild type allele. In this case, we assume that the mother is heterozygous for the causative variant and still unaffected due to incomplete penetrance. The variants were selected based on the following criteria; variant to be missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant, correlated with a coding transcript type, presenting maximum 5% minor allele frequency (MAF) and a minimum read depth of 30. The variants that matched these filtering criteria are listed in Table 5.8.

Table 5.8. Candidate gene variant list after filtration of WES data of the whole family for autosomal dominant inheritance. (Zygosity of the variants belong to the index patient P1403-3.)

Gene	**Chromosome**	**Zygosity**	**Annotation**	**Nucleotide Change**	**Amino acid Change**	**dbSNP**
ANKRD54	chr22	HET	Missense	c.704G>A	p.Arg235His	rs150257798
BPIFC	chr22	HET	Splice Region & Intron	c.1218-7C>T	.	rs55807703
DSG2	chr18	HET	Missense	c.877A>G	p.Ile293Val	rs2230234
ENPP3	chr6	HET	Splice Region & Intron	c.277+7G>A	.	rs189132143
FAM169B	chr15	HET	Missense	c.38T>G	p.Phe13Cys	rs144765676
FHOD3	chr18	HET	Missense	c.646G>A	p.Val216Ile	rs551483382
ITIH2	chr10	HET	Missense	c.1705C>G	p.Leu569Val	rs7084817
KIAA2026	chr9	HET	Missense	c.1510G>T	p.Val504Leu	rs41314612
MCM9	chr6	HET	Missense	c.2272A>G	p.Thr758Ala	rs79670608
SLC5A4	chr22	HET	Stop Gained	c.415G>T	p.Glu139*	rs62239058

Additionally, SIFT and PolyPhen predictions as well as presence in gnomAD database were considered as the variants are prioritized. Human Splicing Finder algorithm was used for identification of the potential impact on transcription of variants in splicing regions. The minor allele frequencies (MAFs), SIFT, PolyPhen, Mutation Taster and Human Splicing Finder predictions of these variants are listed in Table 5.9.

Table 5.9. Prioritization of autosomal dominantly inherited candidate gene variants by using different algorithms and tools. (D: deleterious; T: tolerated; B: benign; PrD: probably damaging; PsD: possibly damaging; Pr. no impact: probably no impact.)

Gene	**Chromosome**	**SIFT**	**Polyphen**	**Mutation Taster**	**Human Splicing Finder**	**MAF values**
ANKRD54	chr22	D	PrD	Disease causing	-	0.00044
BPIFC	chr22	-	-	-	Pr. no impact	0.02535
DSG2	chr18	D	PsD	Polymorphism	-	0.0234
ENPP3	chr6	-	-	-	Pr. no impact	0.00500
FAM169B	chr15	D	PrD	Disease causing	-	0.00090
FHOD3	chr18	D	B	Disease causing	-	0.00008
ITIH2	chr10	D	PsD	Polymorphism	-	0.03900
KIAA2026	chr9	D	B	Polymorphism	-	0.00130
MCM9	chr6	T	B	Polymorphism	-	0.02700
SLC5A4	chr22	-	-	Disease causing	-	0.01135

Ankyrin repeat domain 54 (*ANKRD54*) gene has a role in regulating protein kinase activity. Its expression is mainly in testis, but present in twenty-five other tissues including brain. In addition, a family member of *ANKRD54*, *AnkrG*, was recently associated with neuropathy (Wang et al., 2018). BPI fold containing family C (*BPIFC*) gene encodes for a protein related with lipid and lipopolysaccharide binding which has a main expression in skin tissue. Although its relation to neuropathy was weak, analysis of *BPIFC* gene variant still performed in order not to ignore any data. Desmoglein 2 (*DSG2*) gene is a member of desmoglein gene family which are calcium-binding transmembrane glycoprotein components of desmosomes. *DSG2* gene has no expression in brain but its analysis was done with the same reason of *BPIFC*.

Ectonucleotide Pyrophosphatase/Phosphodiesterase 3 (*ENPP3*) gene performs the hydrolysis of extracellular nucleotides and has a minor expression in brain while most of its expression remains in small intestine. Family with sequence similarity 169, member B (*FAM169B*) is known to code for a protein whose function is yet to be identified. *KIAA2026* gene also has the same situation. Thus, it might be important to analyze variants of these genes. Protein product of formin homology 2 domain containing 3 (*FHOD3*) plays a role in actin filament polymerization, additionally, expression of the gene is present in brain. Inter-alpha-trypsin inhibitor (*ITIH2*) gene expresses its protein product in brain, also. This gene is involved in extracellular matrix stabilization as well as prevention of tumor metastasis. Inclusion of *ITIH2* gene was with the aim of not missing any possibly important variant. Minichromosome maintenance 9 homologous recombination repair factor (*MCM9*) gene is a part of the gene family whose involvement is critical for initiation of eukaryotic genome replication. Binding of *MCM9* protein to the chromatin is a prerequisite for MCM2-7 helicase recruitment to DNA replication origins. Last of the candidate gene variants, solute carrier family 5 member 4 (*SLC5A4*) is a sodium glucose transporter, as the name implies.

All above listed candidate gene variants were excluded based on several lines of evidence. The variants should have deleterious, (probably/possibly) damaging, or disease-causing effects as predictions from SIFT, PolyPhen and MutationTaster, respectively, to be considered as a strong candidate. Therefore, a prediction result of 'benign' for the effect of nucleotide change on protein function, decreases the possibility that the variant in question can be responsible for the disease. Unfortunately, SIFT, PolyPhen and Mutation Taster algorithm results were indicating benign nature for some of the variants and still some were already reported as common polymorphisms. This led to elimination of the variants in genes *DSG2*, *FHOD3*, *ITIH2*, *KIAA2026* and *MCM9*. These *in silico* tools cannot be used to predict the effect of nucleotide changes in splice sites variants. Thus, for unraveling their effects on splicing, Human Splicing Finder bioinformatics tool was used (HSF, "n.d"). *BPIFC* and *ENPP3* gene variants were tested by this bioinformatics tool and they were found not to affect the splicing with a high probability. Human Splicing Finder outcomes of *BPIFC* and *ENPP3* gene

variants are given in Figure C.1 and Figure C.2, respectively.

Even though all these evidences helped us to exclude these variants, the strongest evidence came from their presence in healthy human population. Our purpose was to find a variant that was as rare as possible that is in accordance with the rare nature of HSAN1B. Thus, these variants should not be previously reported in population databases. On the contrary to this expectation, they were all reported to be present in the healthy population in heterozygous state. These heterozygous individuals were obviously do not present the symptoms of the disease. Furthermore, all of the candidate variants, except *ANKRD54* gene variant, listed in Table 5.8. were previously reported in homozygous state among healthy individuals. As a result, the candidate gene variants obtained after filtering of WES data of the family members for autosomal dominant segregation were all excluded to be disease causing.

5.2.3. Investigations for a Possible X-linked Inheritance

A possible X-linked inheritance was also considered when the previous inheritance investigations came out inconclusive. For this type of inheritance pattern, a male index patient P1403-3 might have a possible disease-causing variant on X chromosome while the other two female siblings, P1403-5 and P1403-6, could become a carrier by having the variant on one of their two X chromosomes. Segregation such as mentioned, would also explain the milder symptoms that the sisters had than the index patient. Filtering of WES data was performed for selecting the variants based on the following criteria; variant to be missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant, correlated with a coding transcript type, presenting maximum 5% minor allele frequency (MAF) and a minimum read depth of 30, SIFT and PolyPhen predictions to be damaging. The variants that matched these filtering criteria are listed in Table 5.10.

Table 5.10. Candidate gene variant list after filtration of WES data of the whole family for X-linked inheritance. (Zygosity of the variants belong to the index patient P1403-3.)

Gene	**Chromosome**	**Zygosity**	**Annotation**	**Nucleotide Change**	**Amino acid Change**	**dbSNP**
ARSD	chr X	HET	Missense	c.845C>A	p.Ala282Asp	rs78034736
SLC25A5	chr X	HET	Missense	c.217G>A	p.Gly73Ser	rs143413528
SLC25A5	chr X	HET	Missense	c.230A>C	p.Asn77Thr	rs148294496
SLC25A5	chr X	HET	Missense	c.235A>T	p.Ile79Phe	rs141428607
SLC25A5	chr X	HET	Missense	c.352G>A	p.Ala118Thr	rs199707714
SLC25A5	chr X	HET	Missense	c.361G>T	p.Gly121Cys	.
SLC25A5	chr X	HET	Missense	c.413G>A	p.Arg138His	rs200550329
SLC25A5	chr X	HET	Missense	c.518T>C	p.Leu173Pro	rs200606066

Arylsulfatase D (*ARSD*) gene encodes a member of sulfatase family which are essential for correct composition of bone and cartilage matrix. Its major expression resides in kidney with many other tissues including brain. Solute carrier family 25 member 5 (*SLC25A5*) belongs to a mitochondrial carrier subfamily of solute carrier genes. The product of this gene functions as a gated pore that regulates the translocation of ATP/ADP between cytoplasm and mitochondrial matrix. All of the candidate genes have a low, yet present expressions in brain. As explained before, all of them pursued because they should not be unnoticed if they really are a disease-causing variant responsible from the symptoms. Primers were designed for each of them. One forward and reverse primer set was designed covering for seven different variants in *SLC25A5* gene. Sanger sequencing results of *ARSD* and *SLC25A5* (c.217G>A) gene variants are listed in Table 5.11 and 5.12, respectively. Chromatograms generated after segregation analyses can be found for *ARSD* in B.4, for each of seven different variants which belong to *SLC25A5* from B.5 to B.11. Only the first (c.217G>A) of *SLC25A5* gene variants is shown in this section to avoid repetition, tables of remaining six variants can be found together with their chromatograms in Appendix B.

Table 5.11. Sanger Sequencing Results for *ARSD* gene c.845C>A variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***ARSD* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	+, +	+, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

Table 5.12. Sanger Sequencing Results for *SCL25A5* gene c.217G>A variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SCL25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	var	var
P1403-4 (Unaff. brother)	+	var
P1403-5 (Aff. sister)	+, +	var, +
P1403-6 (Aff. sister)	+, +	var, +
Conclusion:	Excluded	

Up to this point, WES results changed every time when samples were sent for direct sequencing. By recognizing this fact, all of the candidate gene variants of X-linked inheritance investigation were sent for Sanger sequencing regardless of their segregation in WES data within family. Results were again, unsurprisingly, differentiated from WES data for each variant but show that the gene variant could not be responsible from disease. As a result, all the candidate gene variants were excluded from the study since a suitable segregation matching with HSN could not be found.

5.3. Homozygosity Mapping based on WES data (HOMWES)

Since the WES data did not reveal any candidate variants/genes for the disease in the investigated family we decided to use another powerful approach known as homozygosity mapping. Homozygosity mapping can be used effectively in search for candidate loci of autosomal recessively inherited disease in families with consanguinity, such as our project family. In this study, PLINK bioinformatics tool was used for performing homozygosity mapping based on WES data as described in Section 4.7. The approach is known as HOMWES technique. Filtering of homozygous regions shared between affected individuals was performed for the index patient P1403-3 and two affected sisters, P1403-5 and P1403-6. Shared homozygous regions between patients P1403-3 and P1403-5 are given in Table 5.13 and for patients P1403-3 and P1403-6 are given in Table 5.14 . Nucleotide positions of the homozygous regions were given as appeared in the file of the patient P1403-3. Columns in the given tables/figures represents the chromosome number (A), beginning (B) and end (C) nucleotide positions of the homozygous region in the chromosome, length (D) of the homozygous region, identified SNP number on the region (E), density of SNPs (F), SNP percentage of homozygosity (G) and heterozygosity (H) in the region, respectively. A graphical representation of homozygous region distributions of the affected patients P1403-3, P1403-5 and P1403-6 are shown in Figures 5.1, 5.2 and 5.3, respectively.

Table 5.13. Shared homozygous regions of patients P1403-3 and P1403-5, indicated by HOMWES analysis.

CHR	**Begin (P1403-3)**	**End (P1403-3)**	**KB**	**NSNP**	**DENSITY**	**PHOM**	**PHET**
1	40554929	77530127	36975.2	95774	0.386	1.000	0.000
2	214727221	224635213	9907.992	32301	0.307	0.999	0.001
5	6020418	7885791	1865.373	2835	0.658	0.990	0.010
5	7885795	9190404	1304.609	719	1.814	0.990	0.010
5	9190598	10236927	1046.329	467	2.241	0.983	0.017
5	13914832	16711363	2796.531	5477	0.511	0.994	0.006
5	16711367	21491658	4780.291	1763	2.711	0.990	0.010
6	18467259	19840095	1372.836	288	4.767	0.986	0.014
6	19840492	24489652	4649.16	3529	1.317	0.991	0.009
10	18968894	20506326	1537.432	1472	1.044	0.987	0.013
10	20506418	21782842	1276.424	1421	0.898	0.996	0.004
10	21783634	25312689	3529.055	7553	0.467	0.995	0.005
12	1945499	7069421	5123.922	24214	0.212	0.999	0.001
12	7069425	8359128	1289.703	7940	0.162	0.999	0.001
12	53343266	69090575	15747.31	62006	0.254	0.999	0.001
12	99639946	105514479	5874.533	20540	0.286	0.999	0.001
16	58546297	60392555	1846.258	2332	0.792	0.992	0.008
16	60392994	61681372	1288.378	37	34.821	0.946	0.054
16	61682109	64977608	3295.499	454	7.259	0.993	0.007

Table 5.14. Shared homozygous regions of patients P1403-3 and P1403-6, identified by HOMWES analysis.

CHR	Begin (P1403-3)	End (P1403-3)	KB	NSNP	DENSITY	PHOM	PHET
1	156354349	161641045	5286.696	31623	0.167	0.999	0.001
1	161641237	168025538	6384.301	13457	0.474	0.999	0.001
1	168025628	200089875	32064.247	65428	0.490	0.999	0.001
6	165108670	169632219	4523.549	7842	0.577	0.999	0.001
11	20959482	22400099	1440.617	1404	1.026	0.996	0.004
11	22851409	26656370	3804.961	1634	2.329	0.992	0.008

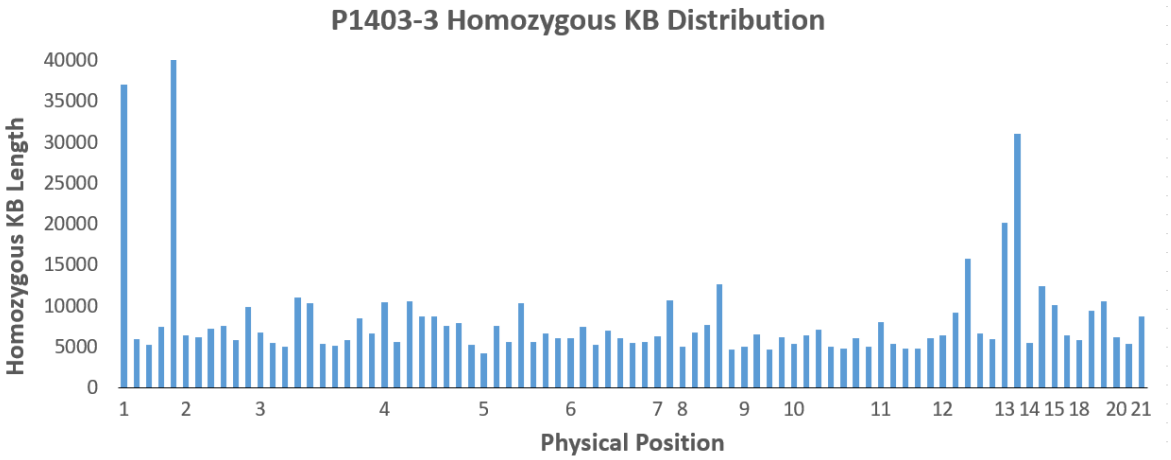

Figure 5.1. Distribution of homozygous regions for patient P1403-3.

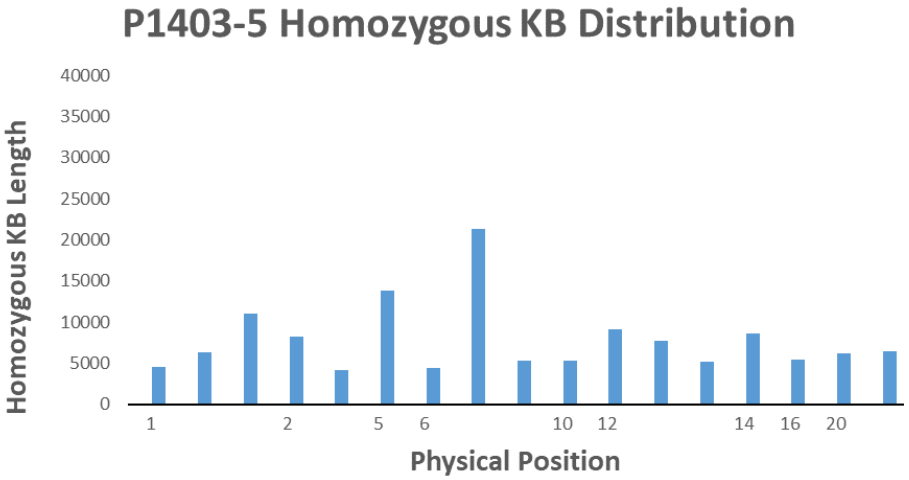

Figure 5.2. Distribution of homozygous regions for patient P1403-5.

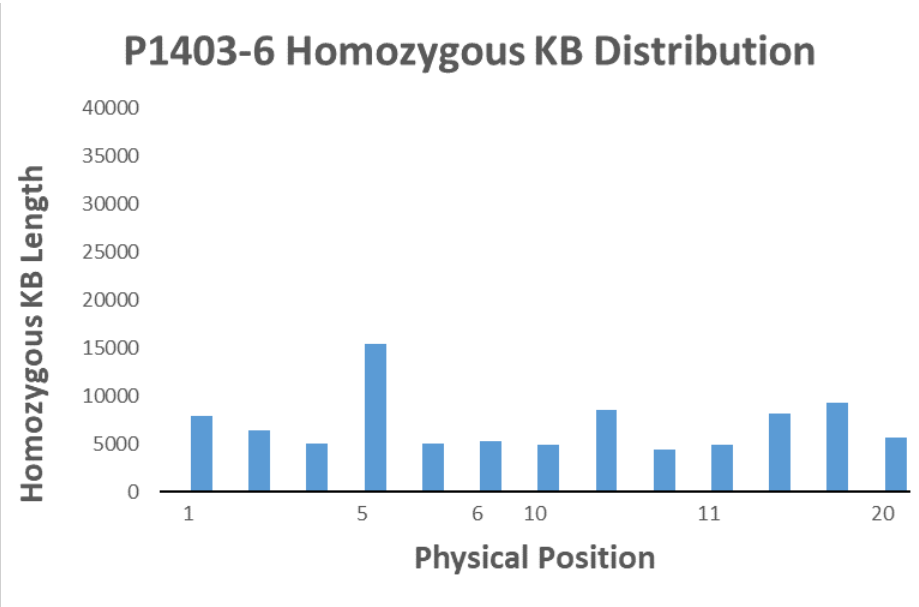

Figure 5.3. Distribution of homozygous regions for patient P1403-6.

The analyses revealed that a significant region in chromosome 1 (chr1:153412391-168025538) was shared between all 3 patients. Location coordinates of the shared region was analyzed exclusively and more deeply in the WES data results. Variants residing in the shared region, were filtered according to the criteria: the variant to be

missense, disruptive inframe deletion, frameshift, downstream, upstream, intergenic or splice region variant, assigned to a coding transcript type, a maximum MAF of 5% and a minimum read depth of 30. Read depth criteria was relaxed when the length of the homozygous region was lower than 1 kb. The variants that fulfilled the filtering criteria are given in Table 5.15. All other shared homozygous regions were investigated independent from their length in order not to miss any strong variants.

Table 5.15. Candidate gene variant list after filtration for HOMWES.

Gene	**Chromosome**	**Zygosity**	**Annotation**	**Nucleotide Change**	**Amino acid Change**	**dbSNP**
FCRL2	chr1	HOM	Missense	c.478C>T	p.Pro160Ser	rs148305794
APCS	chr1	HOM	Splice region & Intron	c.65-5C>T	.	rs200668921
KCNJ10	chr1	HOM	Missense	c.53G>A	p.Arg18Gln	rs115466046
PEA15	chr1	HOM	Splice region & Intron	c.235+3G>A	.	rs8192602
CD244	chr1	HOM	5'UTR	c.-155G>A	.	rs181763130
ITLN2	chr1	HOM	Missense	c.502T>C	p.Ser168Pro	rs79889663
C1orf111	chr1	HOM	Splice region & Intron	c.104-6A>C	.	rs164182
POGK	chr1	HOM	Missense	c.494G>C	p.Ser165Thr	rs143520580
POGK	chr1	HOM	Downstream	c.*1950G>A	.	rs140649503
GPA33	chr1	HOM	Missense	c.355G>T	p.Val119Phe	rs72689400
MPZL1	chr1	HOM	Upstream	c.-203T>C	.	rs3738229
ADCY10	chr1	HOM	Splice region & Intron	c.4482+3G>A	.	rs75209820
F5	chr1	HOM	Splice region & Intron	c.730+7C>T	.	rs6023
FMO2	chr1	HOM	5'UTR	c.-87C>T	.	rs28369797
FMO2	chr1	HOM	Missense	c.1172G>C	p.Arg391Thr	rs28369899
FMO2	chr1	HOM	Stop lost	c.1414T>C	p.Ter472Glnext*?	rs6661174

Fc receptor like 2 (*FCRL2*) gene encodes members of immunoglobulin receptor family and proved to be a prognostic marker for chronic lymphocytic leukemia (Van Bockstaele et al., 2009). Although it does not have an expression in brain and no relation with neuropathy, it was still analyzed with the aim of not missing any variant. Amyloid P component, serum (*APCS*) gene codes for a glycoprotein which is thought to control degradation of chromatin. The gene has a low expression in brain. Potassium voltage-gated channel subfamily J member 10 (*KCNJ10*) gene could be an important gene to follow because it might be responsible for potassium buffering action of glial cells in brain. Moreover, it has a biased expression in brain. Proliferation and apoptosis adaptor protein (*PEA15*) gene is responsible for encoding negative regulator of apoptosis, protein product of *PEA15* is endogenous substrate for protein kinase C. Its expression is broad in brain.*CD244* and intelectin 2 (*ITLN2*) genes are related with immune system, former encoding cell surface receptors in natural killer and some T cells, and latter playing a role against pathogens. Their involvement in study is the same with *FCRL2* gene variant. *C1orf111* gene and its protein product is uncharacterized; however, in the WES data files of patients except P1403-3, this gene appeared as *SPATA46* gene which has a role in spermiogenesis and fertilization. Function of a protein encoded by pogo transposable element derived with KRAB domain (*POGK*) gene, is not known but protein has two domains: KRAB domain for protein-protein interactions and transposase domain. *POGK* gene expressed in brain. Glycoprotein A33 (*GPA33*), on the other hand, has no brain expression.*GPA33* is known to have a role in cell to cell recognition and signaling. Myelin protein zero like 1 (*MPZL1*) gene is related with adhesion and tyrosine kinase pathways. It is expressed generally including brain. Adenylate cyclase 10 (*ADCY10*) codes for adenylyl cyclase which is insensitive for G protein. Even though it is expressed only in testis, it was included in further analyses, just to be sure. Protein product of coagulation factor 5 (*F5*) is a cofactor of blood coagulation cascade, as the name implies. It circulates in plasma and converts prothrombin into thrombin. Expression in brain of the protein is very low. Function of *FMO2* gene is explained in Section 5.2.1.

SIFT, PolyPhen and Human Splicing Finder prediction algorithms or gnomAD databases were used for prioritization of variants. SIFT, PolyPhen predictions, zygosity of the variants in the unaffected brother P1403-4, Human Splicing Finder results, the minor allele frequencies (MAFs) and the number of homozygous individuals (taken from gnomAD database) of the filtered variants, can be found in Table 5.16. Variants of the same gene were listed in the table by indicating functional annotation to avoid confusion.

Table 5.16. Prioritization of HOMWES candidate gene variants by using different algorithms and tools. (D: deleterious; T: tolerated; B: benign; PrD: probably damaging; PsD: possibly damaging; Pr. no impact: probably no impact.) (Hom: homozygote; Het: heterozygote; WT: wild type.)

Gene	**Chromosome**	**SIFT**	**Polyphen**	**Zygosity in P1403-4**	**Human Splicing Finder**	**MAF values**	**Reported Homozygous Individuals**
FCRL2	Chr1	T	B	HET	-	0.0037	12
APCS	Chr1	.	.	HOM	Pr. no impact	0.00059	0
KCNJ10	Chr1	T	B	HOM	-	0.0043	17
PEA15	Chr1	.	.	HOM	Pr. no impact	0.0089	65
CD244	Chr1	.	.	HOM	-	0.0011	-
ITLN2	Chr1	D	PrD	HOM	-	0.032	309
C1orf111	Chr1	.	.	HOM	Potential alteration of splicing	0.037	1187
POGK (Missense)	Chr1	T	B	HOM	-	0.0013	2
POGK (Downstream)	Chr1	.	.	WT	-	0.0013	-
GPA33	Chr1	D	D	HOM	-	0.0057	27
MPZL1	Chr1	.	.	WT	-	0.045	-
ADCY10	Chr1	.	.	HOM	Pr. no impact	0.015	28
F5	Chr1	.	.	HOM	Pr. no impact	0.032	645
FMO2 (5'UTR)	Chr1	.	.	WT	-	0.011	-
FMO2 (Missense)	Chr1	D	PrD	HET	-	0.011	0
FMO2 (Stop lost)	Chr1	.	.	HOM	-	0.046	191

HOMWES analyses successfully pinpointed the long stretches of homozygosity that may indicate the position of the candidates responsible for the disease. Since the unaffected brother should not carry the allele in homozygous state, candidate gene variants that were homozygous in P1403-4, were excluded in the study. Nine of the candidate gene variants including *KCNJ10, PEA15, ITLN2, C1orf111, POGK* (missense), *GPA33, ADCY10, F5* and *FMO2* (stop lost) were eliminated based on the finding that they were previously reported in homozygous state among healthy individuals. Human Splicing Finder tool predicted that *C1orf111* gene variant can have a potential effect on splicing; however, presence of 1187 healthy individuals homozygous for this variant in gnomAD database significantly decreased the possibility that it can cause a rare disease. Human Splicing Finder prediction tool outcomes for the *APCS, PEA15, C1orf111, ADCY10* and *F5* splice region variants are given in Figures from C.3 to C.7, respectively. Eleven out of the 16 candidate gene variants were not pursued furthermore due to these negative results. One of the remaining five variants, *FMO2* (missense, c.1172G>C), was further investigated by Sanger sequencing. The results of *FMO2* can be found in Table 5.5 and Figure B.1. Other four gene variants, *FCRL2, POGK* (downstream), *MPZL1* and *FMO2* (5'UTR) were also further analyzed by Sanger sequencing for verification. Sanger chromatograms for *FCRL2, POGK* (downstream), *MPZL1* and *FMO2* (5'UTR) gene variants are given in Figures from B.14 to B.17, respectively. Comparison of Sanger sequencing and WES data results of these candidate genes are given in Tables 5.17, 5.18, 5.19 and 5.20.

Table 5.17. Sanger Sequencing and WES data results for *FCRL2* gene c.478C>T variant. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***FCRL2* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, +	var, +
P1403-3 (Index case)	var, var	var, var
P1403-4 (Unaff. brother)	var, var	var , +
P1403-5 (Aff. sister)	var, +	var, var
P1403-6 (Aff. sister)	var, var	var, var
Conclusion:	Excluded	

Homozygosity mapping program scans a sliding window of twenty SNPs on a chromosomal region to map homozygous regions. The program labels a region as homozygous when the homozygous SNP number exceeds ten. Therefore, even though a homozygous region was shared among patients, they can be heterozygous in single gene variants. This is the situation for WES data in sister P1403-5. Sanger sequencing results for the *FCRL2* gene variant (c.478C>T) showed differences from WES data. First, affected sister P1403-5 appeared as homozygous for the variant in WES data but was actually heterozygote for the variant. Secondly, and more importantly, unaffected brother P1403-4 seemed to be heterozygous while he turned out to be homozygous for the variant in direct sequencing. Unaffected brother was expected to have a different genotype than the patients by carrying wild type alleles or being heterozygous for the variant for a suitable disease segregation. In conclusion, *FCRL2* gene variant was excluded to be causative.

Table 5.18. Sanger Sequencing and WES results for *POGK* gene c.*1950G>A variant. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***POGK* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	+, +	+, +
P1403-3 (Index case)	+, +	var, var
P1403-4 (Unaff. brother)	+, +	+ , +
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

POGK gene variant (c.*1950G>A) was first thought to be a strong candidate because of its low frequency in healthy population and pursued due to unreliability of WES data results. However, direct sequencing results showed that all individuals in the family were carrying the wild type allele, indicating a false positive result. Therefore, this variant was also excluded.

Table 5.19. Sanger Sequencing and WES results for *MPZL1* gene c.-203T>C variant. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***MPZL1* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, var	+, +
P1403-3 (Index case)	var, var	var, var
P1403-4 (Unaff. brother)	var, var	+ , +
P1403-5 (Aff. sister)	var, var	+, +
P1403-6 (Aff. sister)	var, var	var, var
Conclusion:	Excluded	

MPZL1 gene (c.-203T>C) variant results also indicated false positives in WES results in three patients, P1403-1, P1403-4 and P1403-5. They all appeared to have the wild type alleles in WES results, but direct sequencing revealed that all individuals of the family were actually homozygous for the variant including the unaffected ones. As a result, *MPZL1* gene (c.-203T>C) variant was also excluded from the study.

Table 5.20. Sanger Sequencing and WES results for *FMO2* gene c.-87C>T variant.(Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***FMO2* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, +	+, +
P1403-3 (Index case)	var, var	var, var
P1403-4 (Unaff. brother)	var, var	+ , +
P1403-5 (Aff. sister)	var, +	+, +
P1403-6 (Aff. sister)	var, var	+, +
Conclusion:	Excluded	

Sanger sequencing results for *FMO2* gene c.-87C>T variant were almost completely different from what WES results had shown. Unaffected mother P1403-1 and affected sister P1403-5 were heterozygous for the *FMO2* gene variant even though they seemed to inherit the wild type allele according to WES results. Unaffected brother P1403-4 and other affected sister P1403-6 turned out to be homozygous for the variant; in WES data they appeared as homozygous for wild type allele. Only WES results of the index case were same in different approaches. Therefore, *FMO2* gene variant was no longer a candidate for our disease due to segregation exclusion in family. All the candidate gene variants coming from HOMWES study were eliminated in conclusion. Upon observation of too many conflicts between the WES data and direct sequencing results, we have chosen a gene variant that previously appeared to have very odd segregation within the family and outsourced for direct sequencing. This gene variant, namely, *MYH14* gene c.4752+197G>A , had been identified in filtering for known neu-

ropathy genes. Chromatograms of *MYH14* gene variant is given in B.18. and Sanger results of the variant are shown in Table 5.21.

Table 5.21. Sanger Sequencing and WES results for *MYH14* gene c.4752+197G>A variant. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***MYH14* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	var, +	var, var
P1403-3 (Index case)	var, +	var, var
P1403-4 (Unaff. brother)	var, +	+ , +
P1403-5 (Aff. sister)	var, +	var, +
P1403-6 (Aff. sister)	+, var	+, +
Conclusion:	Excluded	

In WES data results for this *MYH14* gene c.4752+197G>A variant, the mother P1403-1 had the variant in homozygous form; however, the brother P1403-4 and sister P1403-6, somehow did not show any of the mother's alleles and were homozygous for the wild type allele. Sanger sequencing revealed the real genotypes; The unaffected mother P1403-1 and index patient P1403-3 were heterozygous for the *MYH14* variant. Unaffected brother P1403-4 appeared as he was carrying wild type alleles; however, Sanger sequencing results showed that he was indeed heterozygous. Overall conclusion was not affected since *MYH14* gene variant was excluded again; however, by this analysis we have shown that the reliability of the WES data was very weak.

5.4. Haplotype analysis for CANVAS diagnosis

While reaching the end of the project, symptomatic conditions of index patient P1403-3 worsened and he was hospitalized. During his time in hospital, he was re-diagnosed with cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome (CANVAS). It is caused by an intronic repeat expansion of pentamer

$[(AAGGG)_{exp}]$ in the Replication Factor C1 (*RFC1*) gene. In the reference genome, there is $(AAGGG)_{11}$ short tandem repeat present in *RFC1* gene second intron. This pentamer repeat of eleven copies in reference is replaced by pathogenic repeat expansion from 400 to 2000 copies, most of them around 1000 copies, in individuals with CANVAS. It is inherited recessively; thus, biallelic expansion is a necessity for CANVAS. In a 2019 paper that suggested the pathogenic repeat expansion by investigating twenty-two CANVAS-affected families, a core ancestral haplotype was identified (Rafehi et al., 2019). They performed linkage analysis in all of the families and detected a founder mutation which is estimated to have arisen thousands of years ago in Europe. Due to crossing-over, small changes happened over the time, but a core ancestral haplotype remained that can be identified using WES data. Our index patient P1403-3, also the affected siblings P1403-5 and P1403-6 were investigated for the presence of this haplotype. Results of haplotype analysis are given in Table 5.22 and a similarity percentage with core haplotype for the individuals studied in 2019 paper as well as our patients are listed in Table 5.23.

Table 5.22. Haplotype analysis of P1403-3, P1403-5 and P1403-6.

position	REF	ALT	genes	region	SNP	MAF	Core Haplotype	P1403-3	P1403-5	P1403-6
38972691	T	C	TMEM156	exonic	rs140693293	0.0038		0/0	0/0	0/0
38972793	A	G	TMEM156	intronic	rs3733268	0.0197		0/0	0/0	0/0
38990334	C	T	TMEM156	intronic	rs3821986	0.4426		1/1	1/0	1/1
38995310	A	G	TMEM156	intronic	rs11721954	0.3605		1/1	1/0	1/1
38995374	T	C	TMEM156	exonic	rs10212770	0.88	C	1/1	1/0	1/1
39000305	A	G	TMEM156	exonic	rs11542133	0.23	A	1/1	1/0	1/1
39064126	G	C	KLHL5	UTR5	rs2711942	0.74	C	1/1	1/1	1/1
39064162	A	C	KLHL5	exonic	rs2711941	0.74	C	1/1	1/1	1/1
39077567	G	GT	KLHL5	intronic	rs201873328	0.02	G	0/0	0/0	0/0
39114929	G	A	KLHL5	intronic	rs3796510	0.53	G	0/0	0/0	0/0
39116775	A	T	KLHL5	intronic	rs10026775	0.53	A	0/0	1/0	0/0
39116911	T	C	KLHL5	exonic	rs3733276	0.53	T	0/0	1/0	0/0
39117103	G	A	KLHL5	intronic	rs3733277	0.53	G	0/0	1/0	0/0
39205365	C	T	WDR19	intronic	rs1451817	0.97	T	1/1	1/1	1/1
39216221	C	T	WDR19	exonic	rs2167494	0.29	C	0/0	0/0	0/0
39216240	G	A	WDR19	exonic	rs75964850	0.04	G	0/0	0/0	0/0
39217779	C	T	WDR19	exonic	rs199765304	0	A	0/0	0/0	0/0
39229771	G	A	WDR19	intronic	rs11730558	0.31	G	0/0	0/0	1/1
39242111	C	A	WDR19	intronic	rs9998591	0.64	A	0/0	1/0	0/0
39254828	A	C	WDR19	exonic	rs187546086	0	A	0/0	0/0	0/0
39255496	A	G	WDR19	intronic	rs33997556	0.08	G	0/0	0/0	0/0
39271541	A	G	WDR19	intronic	rs3733280	0.31	A	0/0	0/0	0/0
39276623	A	AG	WDR19	intronic	rs11096989	0.5	AG	1/1	1/0	1/1
39279724	T	C	WDR19	intronic	rs12648082	0.5	C	1/1	1/0	1/1
39279907	A	G	WDR19	intronic	rs2276888	0.46	A	0/0	1/0	0/0
39297211	C	T	RFC1	intronic	rs41547922	0	C	0/0	0/0	0/0
39301605	G	C	RFC1	intronic	rs2066788	0.51	C	1/1	1/0	1/1
39302029	T	C	RFC1	exonic	rs2066786	0.45	T	0/0	1/0	0/0
39303925	A	G	RFC1	exonic	rs2066782	0.1	G	1/1	1/0	1/1
39329102	G	A	RFC1	intronic	rs11096992	0.47	A	1/1	1/0	1/1
39353122	C	T	RFC1	intronic	rs4975007	0.98	T	1/1	1/1	1/1
39353137	C	T	RFC1	intronic	rs374311239	0	C	0/0	0/0	0/0
39439264	T	C	KLB	intronic	rs4975015	0.1707	C	0/0	0/0	0/0
39448529	G	A	KLB	exonic	rs17618244	0.1528	G	0/0	0/0	0/0
39448542	C	G	KLB	exonic	rs7685429	0.7326	G	0/0	1/0	0/0
39448586	C	T	KLB	exonic	rs35372803	0.0375	C	0/0	0/0	0/0
39450229	C	A	KLB	exonic	rs4975017	0.2881		0/0	1/0	0/0
39450403	TGA	T	KLB	UTR3	rs112327399	0.0603		0/0	0/0	0/0
39457857	T	C	RPL9	intronic	rs1015450	0.1489		0/0	0/0	0/0
39457870	T	TG	RPL9	intronic	rs3216720	0.6465		0/0	1/0	0/0

Table 5.23. A percentage of similarity with core haplotype for previously reported CANVAS patients (CNV1-18) (Rafehi et al., 2019) and patients of this study (P1403-3, P1403-5, P1403-6). (Same: Same changes with core ancestral haplotype of CANVAS. Not Same: Reverse of the Same results.)

COUNT	CNV1	CNV3	CNV4	CNV7	CNV8	CNV9	CNV12	CNV18	CNV14	P1403-3	P1403-5	P1403-6
same	27	30	29	28	20	31	28	27	18	26	23	25
not_same	5	2	3	4	12	1	4	5	14	6	9	7

PERCENT	CNV1	CNV3	CNV4	CNV7	CNV8	CNV9	CNV12	CNV18	CNV14	P1403-3	P1403-5	P1403-6
same	84.4%	93.8%	90.6%	87.5%	62.5%	96.9%	87.5%	84.4%	56.3%	81.3%	71.9%	78.1%
not_same	15.6%	6.3%	9.4%	12.5%	37.5%	3.1%	12.5%	15.6%	43.8%	18.8%	28.1%	21.9%

Out of thirty-two variants in the core haplotype, twenty-six was present in index patient P1403-3 which equals to a percentage of 81.3%. Most of the CANVAS-affected families had similar percentages or even less, in similarity with core ancestral haplotype. P1403-5 and P1403-6 had 71.9% and 78.1% similarity that also fit the severity of symptoms observed in each of them. This analysis partially confirmed the CANVAS diagnosis for index patient at the molecular level. To confirm the presence of the repeat expansion of 400 to 2000 copies linked to that core haplotype a PCR analysis is planned to be performed in the near future. The amplification limit in conventional PCR is about 10 kilo base pairs. Pathogenic repeat expansions in CANVAS, even in the 400 copies, far exceeds this limit. Therefore, the targeted region must not be amplified with PCR in samples of CANVAS patients and observed as absence of the band in agarose gel corresponding to the repeat region. The reference pentamer repeat of eleven copies is expected to be amplified and observed as a 250 bp fragment on the gel. Primers corresponding to five internal control regions will be used in the same PCR amplification with the repeat flanking primers to confirm the successful amplification. The same analysis will be repeated for all family members for which the genomic DNA samples were available. Unfortunately, this analysis is not enough to define whether the individual is heterozygous or homozygous for absence of the repeat expansion when an amplification product appears on the gel and also for determination of copy number of the repeat expansion. A repeat primed PCR will be performed for members of the family to gain more perspective of genetic mechanisms of the CANVAS in the future.

6. DISCUSSION

A family of five members was analyzed in this particular study in order to identify a novel gene causing their rare disease predicted to be HSAN Type 1B (HSAN1B). Whole Exome Sequencing (WES) and homozygosity mapping based on WES data (HOMWES) were used that are both widely accepted methods in searching for genes responsible from autosomal recessive diseases. Research of the novel gene variant causing HSAN symptoms started with investigating the previously reported genes known to be related with neuropathy. WES results were screened for variants in CMT, HSN, HMN, Spastic Paraplegia, and Ataxia genes. Each candidate variant was picked based on the gene's known inheritance pattern and segregation was cross-checked in WES results of all family members. None of the neuropathy genes had a causative variant that could be responsible for the symptoms observed in the family.

In the interest of unraveling the reason behind those symptoms, we have started investigation for a novel gene variant. Analyses performed can be classified as four distinct approaches in accordance with the mode of inheritance; autosomal recessive, autosomal dominant, X-linked and homozygosity mapping based on WES results (HOMWES). The nature of symptoms and consanguinity of parents first led us to focus on autosomal recessive (AR) segregation of the causative variant. Analyses of WES data for AR variants revealed two candidate genes, *FMO2* and *MYH7B*. Flavin-containing monooxygenase 2 (*FMO2*) gene codes for a member of flavin containing enzymes, taking part in oxidation of NADPH and molecular oxygen. These enzymes have been reported to be involved in drug metabolism (Geier et al., 2015). Flavin monooxygenases have multiple isoforms. The *FMO2* is majorly expressed in lungs. This localization of the gene expression was thought to be related with chronic cough observed in the index patient. Another interesting feature of *FMO2* gene is that its major allele encodes a catalytically inactive truncated polypeptide (Dolphin et al., 1998). Unfortunately, direct sequencing for *FMO2* in family members revealed that segregation was not exactly as it appeared in WES data results. This led to elimination

of our strongest candidate variant.

The other candidate for autosomal recessive inheritance was *MYH7B* gene that codes for heavy chain of myosin functioning in muscle contraction. Family of genes coding for myosin heavy chains regulate the muscle gene expression with the help of their expressed intronic microRNAs. *MYH7B* and its microRNA, miR-499 are both expressed in skeletal muscle and brain (Yeung et al., 2012). This gene variant is considered to be a strong candidate because knockdown of *MYH7B* was reported to cause disruption of the critical synaptic structure (Rubio et al., 2011). Segregation of the *MYH7B* variant among the family members was investigated by Sanger sequencing and the results were again different from that of WES data causing exclusion of our second strongest candidate.

A locus residing in chromosome 3 had been identified by Kok et al. (2003) to be responsible for the HSAN Type 1B. Therefore, we have investigated that locus in the WES data exclusively. Even though there were no strong candidate gene variantS in the particular region, *CLASP2* (Cytoplasmic Linker Associated Protein 2) gene variant attracted our attention. Actually, one member of the family did not fit the expected AR segregation pattern. *CLASP* gene family is associated with microtubules and they guide, stabilize or link microtubules with actin filaments (Tsvetkov et al., 2007). To avoid false negative results, the variant was tested with direct sequencing and excluded afterwards.

When all the options for variants with AR inheritance were eliminated, a possible autosomal dominant (AD) inheritance was considered. We hypothesized that the AD inheritance of the candidate variant may also explain the incomplete penetrance and variable expression observed in our family. In addition, two families that were studied in Kok et al. (2003), were presenting autosomal dominant inheritance. A candidate gene variant list was generated consisting of ten genes after a careful filtration of WES data. All of these candidate gene variants were previously reported in healthy population databases. Since rare nature of the disease required a gene variant not to

be previously reported in healthy population, candidate gene variants of autosomal dominant inheritance were excluded from the study.

As a next step, X-linked variants were filtered from the WES data. X-linked inheritance could be expected in the family since the symptoms were severe in the male index patient and the course of disease was milder in female patients. A mutant gene variant on the single X chromosome expressed in all cells of the male index case may cause more serious symptoms while in his two heterozygous sisters, the effect of the mutant allele on one X can be compensated by the other X-chromosome carrying the normal allele. Dependent on X-chromosome inactivation pattern in different cells a milder form of the disease in possible in the heterozygous females. This analysis revealed four candidate gene variants: *ARSD, TAB3, SLC25A5* and *RBMX*. *ARSD* gene functions as arsenic metallochaperone by producing increased efflux from the cell; therefore, cells can resist to environmental concentrations of arsenic (Lin et al., 2007). *TAB3* gene has a critical role in NF-κB pathway along with TAK-1 which is an activator of NF-κB nuclear factor. This gene contributes to embryonic development as well as immunity, cell survival and many other cellular processes (Zhao et al., 2018). *SLC25A5* is responsible from the ADP/ATP exchange between the cytoplasm and mitochondria. Interestingly, this gene has also been reported to have a role in post-synaptic density of central nervous system which is critical for cognitive processes (Laumonnier et al., 2007). The last one, the *RBMX* gene encodes for RNA binding protein which contributes to alternative splicing and DNA damage response by positively regulating homologous recombination (Adamson et al., 2012). Sanger sequencing of the relevant exon of these genes revealed that all members of the family were carrying the wild type allele. None of them presented a susceptible segregation pattern, and hence, excluded.

When all of the previous approaches were inconclusive, we decided to use homozygosity mapping based on WES data. This method unravels the shared homozygous regions between affected members of the family. It has been used by our group before and proved itself effective. HOMWES analysis revealed a region located in chromosome 1 which was shared between three affected siblings. This region was analyzed

deeply, and four gene variants were considered as strong candidates, *FCRL2, POGK, MPZL1* and *FMO2* (5'UTR). *FCRL2* (FC receptor like 2) gene encodes members of immunoglobulin receptor family. It has an expression in appendix, lymph node, spleen, and bladder. *FCRL2* gene is also associated with lymphocytic leukemia (Nückel et al., 2009). Even though its function and expression pattern were not correlating with a nervous system defect we decided not to exclude it without testing the familial segregation since even very common housekeeping genes like tRNA synthetases were known to cause neuropathy. As expected, though, the gene variant was excluded by direct sequencing of the family samples. Unlike *FCRL2* gene, *POGK* gene has a ubiquitous expression in brain. Function of the protein coded by *POGK* (Pogo transposable element derived with KRAB domain) is not known. However, it has a transposase domain and KRAB domain for protein-protein interactions. Nevertheless, it was also excluded upon direct sequencing for segregation. *MPZL1* (Myelin protein zero like 1) gene has a general expression including brain. An important paralog of *MPZL1* is MPZ gene that is commonly mutated in CMT patients (Gould et al., 2008). *MPZL1* gene is associated with carcinogenesis by promoting proliferation of tumor cells and migration (Chen et al., 2019). This gene variant was also excluded because the three key members of the family appeared to carry a wild type allele as for other previous candidates of homozygosity mapping. Last candidate gene variant in HOMWES analysis was *FMO2*. A different variant in the same gene was encountered in the autosomal recessive inheritance analyses. This variant residing in the 5'UTR region of the *FMO2* gene was not segregating according to the disease status as others and excluded from the study.

The major problem encountered in this study was the absence of reproducibility of the WES data in segregation studies by Sanger sequencing. Even though WES is a successful technique and main driver for defining novel genes, it does have some limitations. To begin with, it covers about 1-2% of the entire genome and around 96% of the exome (Pipis et al., 2019). Since it does not provide a complete coverage of some protein coding regions, it becomes harder to identify a possibly pathogenic gene variant corresponding to those uncovered regions. Another challenge of WES

technique is the variety of existing methods that differs between companies during the variant calling and alignment (Foo et al., 2012). In our project, WES data of index patient P1403-3 was obtained from a different service provider (Macrogen, Inc.) than the other members of the family (Dnalab, Turkey) and and this difference between datasets caused some inconveniences, such as possibility of a coverage diversity between patients and style differences in final provided files. Another drawback of WES is the presence of false positive and false negative variants (Bertier et al., 2016). In this study, each candidate gene variant was tested with Sanger sequencing for verification to avoid false positive results and most of the results were, indeed, incompatible with WES data, causing the elimination of the variant in question. This drawback increases the possibility of missing the causative variants during strict filtering of WES data. To provide evidence for the fact that the WES data does not reflect the real genotypes for most of the analyzed loci we especially analyzed a specific gene variant of *MYH14* that shows unpredictable familial segregation. Sanger results proved that WES data was inaccurate and suggested that the false positive rate of WES method can be higher than predicted. Any possible false negative result remained unknown due to a huge cluster of gene variants that one encounters among at least 20.000 variants.

The high number of neuropathy families with unknown causative genes directs us to the next important challenge which is the interpretation of variants of unknown significance (VUS) (Bertier et al., 2016). Main complication in WES analysis is determining whether the observed gene variant is pathogenic or not. To give an example, two prediction algorithms, Polyphen and SIFT software that have also been used in this study, can give different predictions for the same amino acid change. This is probably because these algorithms utilize conservation among species while making a prediction and the change of amino acid in question may reside in the poorly conserved position (Foo et al., 2012). This interpretation challenge is one of the reasons that the responsible gene identification ratio with WES analysis is between 19% and 45% in patients with CMT or other neuropathies in literature. More factors contribute to this low diagnosis ratio such as genetic heterogeneity within populations and within families along with phenotypic heterogeneity. A genetic modifier could affect the phenotype of an

affected individual in addition to the effect of pathogenic variant; hence, complicates the process of accurate identification (Pipis et al., 2019).

The major disadvantage of WES method is that it cannot detect large duplications, insertions, and deletions in the genome. Therefore, the possibility of inheritance of a large chromosomal structural abnormality could not be investigated in this family. This is why the repeat expansion responsible for cerebellar ataxia with neuropathy and vestibular areflexia syndrome (CANVAS) could not be detected in the first place. The clinical diagnosis of CANVAS in the index case was only possible prior to three weeks before the presentation of this thesis when he was hospitalized due to worsening of the symptoms. CANVAS is caused by a pathogenic repeat expansion of an AAAAG pentamer changing from 400 to 2000 copies in affected individuals that is found in 11 copies in reference sequence (Rafehi et al., 2019). The molecular confirmation of the CANVAS in the family was than achieved by haplotype analysis using the WES data. A core haplotype was identified by Rafehi et al. (2019) among twenty-two CANVAS-affected families. We have confirmed existence of this core haplotype in our affected family members, P1403-3, P1403-5 and P1403-6. Presence of 81.3% similarity to this haplotype in the index case supported the diagnosis of CANVAS. Considering the high rate of false positives in our WES data, the haplotypes observed in our family members should still be confirmed by Sanger sequencing. A further evidence was planned to be provided for the presence of the pathogenic repeat expansion in the family by using the PCR with flanking primers of the repeat and six internal controls. Observation of the amplification products for the internal controls and absence of amplification for the CANVAS repeat in the affected individuals and its amplification in unaffected individuals from the family would have proven the presence of the causative repeat expansion in the family. This PCR analysis could not be performed due to time limitation but along with the analysis of repeat primed PCR, aimed for determining the number of pathogenic repeats on causative alleles, a complete genetic diagnosis of the patient will be provided. Further analysis is required mainly by repeat primed PCR that is planned for the near future.

7. CONCLUSION

This study has started with the aim of identifying a novel gene that causes HSN disease. WES method was performed for the unaffected mother, affected index case, unaffected brother and two affected sisters. Pedigree of the family suggested autosomal recessive inheritance and two candidate gene variants, *FMO2* and *MYH7B*, were analyzed within this scope, later, eliminated by direct Sanger sequencing. Candidate gene variants for a possible autosomal dominant inheritance were also excluded since they had high frequencies in healthy population. The analyses were then focused on X-linked inheritance and determined a total of ten candidate variants in four genes, *ARSD, TAB3, SLC25A5* and *RBMX*. All of them were excluded since segregation of the allele was not associated with the disease status. Lastly, homozygosity mapping combined with WES data was used to detect the rare disease variants. A locus residing in chromosome 1 was found to be shared among affected individuals and four gene variants, *FCRL2, POGK, MPZL1* and *FMO2* (a different variant from the one analyzed in a possible autosomal recessive inheritance part of the study) were acquired from that homozygous region. Sanger sequencing led the exclusion of these variants, too.

The step of verification with direct sequencing has been a determining factor in all investigated gene variants in this study. Thirty-four results of WES data out of seventy-seven were not in accordance with Sanger sequencing data, constituting a false positive rate of 44%. This percentage of false results interferes with a proper investigation and led us question the yield of WES. It has a number of limitations but the most important one especially for this study was that WES cannot detect large chromosomal insertions, deletions, or duplications. This is the reason why CANVAS, caused by a biallelic pathogenic repeat expansion of a pentamer, diagnosis could not be obtained genetically. The clinical diagnosis of CANVAS was available upon the very end of this thesis. Nevertheless, a haplotype analysis could be performed in three affected individuals and revealed that the index patient's haplotype was 81.3%

identical with the CANVAS core haplotype. Conventional and repeat primed PCR studies should be further performed to prove the existence and the copy number of this causative pentamer expansion that can finally provide a complete genetic diagnosis for this family.

REFERENCES

Adamson B., A. Smogorzewska, F.D. Sigoillot, R.W. King, S.J. Elledge, 2012, "A genome- wide homologous recombination screen identifies the rna-binding protein rbmx as a component of the dna-damage response", *Nature Cell Biology*, Vol. 14, No: 3, pp. 318–328.

Ahmad H., T. Requena, L. Frejo, M. Cobo, A. Gallego-Martinez, F. Martin, J.A. Lopez-Escamez, A.M. Bronstein, 2018, "Clinical and functional characterization of a missense elf2 variant in a canvas family", *Frontiers in Genetics*, Vol. 9, pp. 85.

Alkuraya F.S., 2010, "Homozygosity mapping: one more tool in the clinical geneticist's toolbox", *Genetics in Medicine*, Vol. 12, No:4, pp. 236–239.

Auer-Grumbach M., 2008, "Hereditary sensory neuropathy type I", *Orphanet Journal of Rare Diseases*, Vol. 3, No:1, pp. 1–7.

Bamshad M.J., S.B. Ng, A.W. Bigham, H.K. Tabor, M.J. Emond, D.A. Nickerson, J. Shendure, 2011, "Exome sequencing as a tool for mendelian disease gene discovery", *Nature Reviews Genetics*, Vol. 12, No:11, pp. 745–755.

Bertier G., M. H´etu, Y. Joly, 2016, " Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views", *BMC Medical Genomics*, Vol. 9, No:1, pp. 52.

Boesch S.M., M.A. Nance, 2020, "Intronic pentanucleotide expansion in the replication factor 1 gene (rfc1) is a major cause of adult-onset ataxia", *Neurology Genetics*, Vol.6, No:3, pp. 440.

Burke D., G.M. Halmagyi, 2018, "Normal tendon reflexes despite absent sensory nerve action potentials in canvas: a neurophysiological study", *Journal of the Neurological Sciences*, Vol. 387, pp. 75–79.

Bussmann J., E. Storkebaum, 2017, "Molecular pathogenesis of peripheral neuropathies: Insights from drosophila models", *Current Opinion in Genetics & Development*, Vol. 44, pp. 61– 73.

Chen D., L. Cao, X. Wang, 2019, “Mpzl1 promotes tumor cell proliferation and migration via activation of src kinase in ovarian cancer”, *Oncology Reports*, Vol. 42, No:2, pp. 679–687.

Cortese A., R. Simone, R. Sullivan, J. Vandrovcova, H. Tariq, W.Y. Yau, J. Humphrey, Z. Jaunmuktane, P. Sivakumar, J. Polke, 2019, “Biallelic expansion of an intronic repeat in rfc1 is a common cause of late-onset ataxia”, *Nature Genetics*, Vol. 51, No:4, pp. 649–658

Dawkins J.L., D.J. Hulme, S.B. Brahmbhatt, M. Auer-Grumbach, G.A. Nicholson, 2001, “Mutations in sptlc1, encoding serine palmitoyltransferase, long chain base subunit- 1, cause hereditary sensory neuropathy type I”, *Nature Genetics*, Vol. 27, No:3, pp. 309–312.

Dolphin C.T., D.J. Beckett, A. Janmohamed, T.E. Cullingford, R.L. Smith, E.A. Shephard, I.R. Phillips, 1998, “The flavin-containing monooxygenase 2 gene (fmo2) of humans, but not of other primates, encodes a truncated, nonfunctional protein”, *Journal of Biological Chemistry*, Vol. 273, No:46, pp. 30599–30607.

Endeavour (”n.d”) http://homes.esat.kuleuven.be/bioiuser/endeavour/tool/ endeavourweb.php, last accessed on August 2020.

Ensembl (”n.d”) Ensembl genome browser. http://www.ensembl.org/index.html, last accessed on August 2020.

ExAC (”n.d”) http://exac.broadinstitute.org/, last accessed on August 2020.

Foo J.N., J.J. Liu, E.K. Tan, 2012, “Whole-genome and whole-exome sequencing in neuro- logical diseases”, *Nature Reviews Neurology*, Vol. 8, No:9, pp. 508–517.

GATK (”n.d”) Genome analysis toolkit. http://www.broadinstitute.org/gatk/ index.php, last accessed on February 2018.

Geier M., T. Bachler, S.P. Hanlon, F.K. Eggimann, M. Kittelmann, H. Weber, S. Lütz, B. Wirz, M. Winkler, 2015, “Human FMO2-based microbial whole-cell catalysts for drug metabolite synthesis”, *Microbial Cell Factories*, Vol. 14, No:1, pp. 1–10.

Genome (”n.d”) Uscs genome browser. https://genome.ucsc.edu/cgi-bin/hgPcr, last accessed on August 2020. gnomAD (”n.d”) https://gnomad.broadinstitute.org/, last accessed on March 2020.

Gould R.M., T. Oakley, J.V. Goldstone, J.C. Dugas, S.T. Brady, A. Gow, 2008, “Myelin sheaths are formed with proteins that originated in vertebrate lineages”, *Neuron Glia Biology*, Vol. 4, No:2, pp. 137.

HSF (”n.d”) Human splicing finder. http://www.umd.be/HSF/HSF.shtml, last accessed on January 2020.

Kok C., M. Kennerson, P. Spring, A. Ing, J. Pollard, G. Nicholson, 2003, “A locus for hereditary sensory neuropathy with cough and gastroesophageal reflux on chromosome 3p22-p24”, *The American Journal of Human Genetics*, Vol. 73, No:3, pp. 632–637.

Kruglyak L., M.J. Daly, E.S. Lander, 1995, “Rapid multipoint linkage analysis of reces- sive traits in nuclear families, including homozygosity mapping”, *American Journal of Human Genetics*, Vol. 56, No:2, pp. 519.

Kurth I., T. Pamminger, J.C. Hennings, D. Soehendra, A.K. Huebner, A. Rotthier, J. Baets, J. Senderek, H. Topaloglu, S.A. Farrell, 2009, “Mutations in fam134b, encoding a newly identified golgi protein, cause severe sensory and autonomic neuropathy”, *Nature Genetics*, Vol. 41, No:11, pp. 1179–1181.

Laumonnier F., P.C. Cuthbert, S.G. Grant, 2007, “The role of neuronal complexes in human x-linked brain diseases”, *The American Journal of Human Genetics*, Vol. 80, No:2, pp. 205– 220.

Lelieveld S.H., M. Spielmann, S. Mundlos, J.A. Veltman, C. Gilissen, 2015, “Comparison of exome and genome sequencing technologies for the complete capture of protein- coding regions”, *Human Mutation*, Vol. 36, No:8, pp. 815–822.

Lin Y.F., J. Yang, B.P. Rosen, 2007, “ARSD: an as (iii) metallochaperone for the arsab as (iii)-translocating ATPase”, *Journal of Bioenergetics and Biomembranes*, Vol. 39, No:5, pp. 453– 458.

MutationTaster (”n.d”) http://mutationtaster.org/, last accessed on March 2020.

Nam S.H., B.O. Choi, 2019, “Clinical and genetic aspects of charcot-marie-tooth disease subtypes”, *Precision and Future Medicine*, Vol. 3, No:2, pp. 43–68.

NeuroMuscular (”n.d”) Neuromuscular disease center. https://neuromuscular. wustl.edu/time/hmsn.html, last accessed on August 2020.

Nückel H., C.H. Collins, U.H. Frey, L. Sellmann, J. Dürig, W. Siffert, U. Dührsen, 2009, “FCRL2 mRNA expression is inversely associated with clinical progression in chronic lymphocytic leukemia”, *European Journal of Haematology*, Vol. 83, No:6, pp. 541–549.

Pais´an-Ruiz C., J.C. Jen, 2020, “Canvas with cerebellar/sensory/vestibular dysfunction from RFC1 intronic pentanucleotide expansion”, *Brain*, Vol. 143, No:2, pp. 386–390.

Pareyson D., P. Saveri, C. Pisciotta, 2017, “New developments in charcot–marie–tooth neuropathy and related diseases”, *Current Opinion in Neurology*, Vol. 30, No:5, pp. 471–480.

Pipis M., A.M. Rossor, M. Laura, M.M. Reilly, 2019, “Next-generation sequencing in charcot–marie–tooth disease: opportunities and challenges”, *Nature Reviews Neurology*, Vol. 15, No:11, pp. 644–656.

PolyPhen (”n.d”) http://genetics.bwh.harvard.edu/pph2/, last accessed on March 2020.

Primer3 (”n.d”) http://primer3.ut.ee/, last accessed on August 2020.

Rafehi H., D.J. Szmulewicz, M.F. Bennett, N.L. Sobreira, K. Pope, K.R. Smith, G. Gillies, P. Diakumis, E. Dolzhenko, M.A. Eberle, 2019, “Validation of new bioinformatic tools to identify expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS”, *BioRxiv*, pp. 597781.

Rubio M.D., R. Johnson, C.A. Miller, R.L. Huganir, G. Rumbaugh, 2011, “Regulation of synapse structure and function by distinct myosin ii motors”, *Journal of Neuroscience*, Vol. 31, No:4, pp. 1448–1460.

Saba S., Y. Chen, K.R. Maddipati, M. Hackett, B. Hu, J. Li, 2020, "Demyelination in hereditary sensory neuropathy type-1C", *Annals of Clinical and Translational Neurology*, pp. 42.

SIFT ("n.d") http://sift.jcvi.org, last accessed on November 2019.

SNPCheck ("n.d") https://genetools.org/SNPCheck/snpcheck.htm, last accessed on August.

SUSPECTS ("n.d") http://www.genetics.med.ed.ac.uk/suspects, last accessed on December 2018.

Szmulewicz D.J., C.A. McLean, M.L. Rodriguez, A.M. Chancellor, S. Mossman, D. Lamont, L. Roberts, E. Storey, G.M. Halmagyi, 2014, "Dorsal root ganglionopathy is responsible for the sensory impairment in CANVAS", *Neurology*, Vol. 82, No:16, pp. 1410–1415.

Szmulewicz D.J., L. Seiderer, G.M. Halmagyi, E. Storey, L. Roberts, 2015, "Neurophys- iological evidence for generalized sensory neuronopathy in cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome", *Muscle & Nerve*, Vol. 51, No:4, pp. 600– 603.

ThermoFisher ("n.d") https://www.thermofisher.com/tr/en/home/brands/ thermo-scientific, last accessed on December 2019.

Timmerman V., A.V. Strickland, S. Züchner, 2014, "Genetics of charcot-marie-tooth (CMT) disease within the frame of the human genome project success", *Genes*, Vol. 5, No:1, pp. 13–32.

ToppGene ("n.d") http://toppgene.cchmc.org/prioritization.jsp, last accessed on December 2019.

Tsvetkov A.S., A. Samsonov, A. Akhmanova, N. Galjart, S.V. Popov, 2007, "Microtubule- binding proteins clasp1 and clasp2 interact with actin filaments", *Cell Motility and the Cytoskeleton*, Vol. 64, No:7, pp. 519–530.

Van Bockstaele F., B. Verhasselt, J. Philippe, 2009, “Prognostic markers in chronic lym- phocytic leukemia: a comprehensive review”, *Blood Reviews*, Vol. 23, No:1, pp. 25–47.

Wang C.C., X.R. Ortiz-Gonz´alez, S.W. Yum, S.M. Gill, A. White, E. Kelter, L.H. Seaver, S. Lee, G. Wiley, P.M. Gaffney, 2018, “*β*iv spectrinopathies cause profound intellectual disability, congenital hypotonia, and motor axonal neuropathy”, *The American Journal of Human Genetics*, Vol. 102, No:6, pp. 1158–1168.

Wilson E.R., U. Kugathasan, A.Y. Abramov, A.J. Clark, D.L. Bennett, M.M. Reilly, L. Green- smith, B. Kalmar, 2018, “Hereditary sensory neuropathy type 1-associated de- oxysphingolipids cause neurotoxicity, acute calcium handling abnormalities and mi- tochondrial dysfunction in vitro”, *Neurobiology of Disease*, Vol. 117, pp. 1–14.

Yeung F., E. Chung, M.G. Guess, M.L. Bell, L.A. Leinwand, 2012, “Myh7b/mir-499 gene expression is transcriptionally regulated by mrfs and eos”, *Nucleic Acids Research*, Vol. 40, No:15, pp. 7303–7318.

Zhao J., L. Gai, Y. Gao, W. Xia, D. Shen, Q. Lin, W. Mao, F. Wang, P. Liu, J. Chen, 2018, “TAB3 promotes human esophageal squamous cell carcinoma proliferation and invasion via the nf-*κ*b pathway”, *Oncology Reports*, Vol. 40, No:5, pp. 2876–2885.

APPENDIX A: WHOLE EXOME SEQUENCING RESULTS

WES results of all investigated individuals, except index P1403-3, are given below. WES results of P1403-3 are listed in Section 5.1.

Table A.1. Number of different types of variants identified in the WES data of patient P1403-1.

Intergenic variants:	186,370
Intron variants:	168,409
Exonic variants:	23,211
Synonymous variants:	11,224
3'UTR variants:	5673
Non-coding exon variants:	4617
Upstream variants:	4438
Downstream variants:	2926
5'UTR variants:	2427
Unknown:	721
Nonframeshift insertion & deletion variants:	299
Splice region variants:	140
Frameshift deletion variants:	128
Stop gained variants:	101
Frameshift insertion variants:	82
Stop lost variants:	10
Total number of variants:	421,134

Table A.2. Number of different types of variants identified in the WES data of patient P1403-4.

Intron variants:	162,578
Intergenic variants:	157,234
Exonic variants:	21,632
Synonymous variants:	10,495
3'UTR variants:	5679
Non-coding exon variants:	4082
Upstream variants:	3897
Downstream variants:	2790
5'UTR variants:	2152
Unknown:	660
Nonframeshift insertion & deletion variants:	261
Splice region variants:	131
Frameshift deletion variants:	111
Frameshift insertion variants:	92
Stop gained variants:	88
Stop lost variants:	11
Total variant number:	380,537

Table A.3. Number of different types of variants identified in the WES data of patient P1403-5.

Intergenic variants:	201,812
Intron variants:	195,727
Exonic variants:	22,180
Synonymous variants:	10,749
3'UTR variants:	6640
Upstream variants:	4835
Non-coding exon variants:	4489
Downstream variants:	3513
5'UTR variants:	2364
Unknown:	696
Nonframeshift insertion & deletion variants:	260
Splice region variants:	123
Frameshift deletion variants:	118
Stop gained variants:	86
Frameshift insertion variants:	85
Stop lost variants:	12
Total variant number:	466,734

Table A.4. Number of different types of variants identified in the WES data of patient P1403-6.

Intron variants:	167,412
Intergenic variants:	152,309
Exonic variants:	22,069
Synonymous variants:	10,672
3'UTR variants:	6092
Non-coding exon variants:	4171
Upstream variants:	3890
Downstream variants:	2983
5'UTR variants:	2231
Unknown:	677
Nonframeshift insertion & deletion variants:	272
Splice region variants:	132
Frameshift deletion variants:	111
Stop gained variants:	96
Frameshift insertion variants:	82
Stop lost variants:	14
Total variant number:	381,408

APPENDIX B: CHROMATOGRAMS

All chromatograms for the investigated variants in the thesis study are given below.

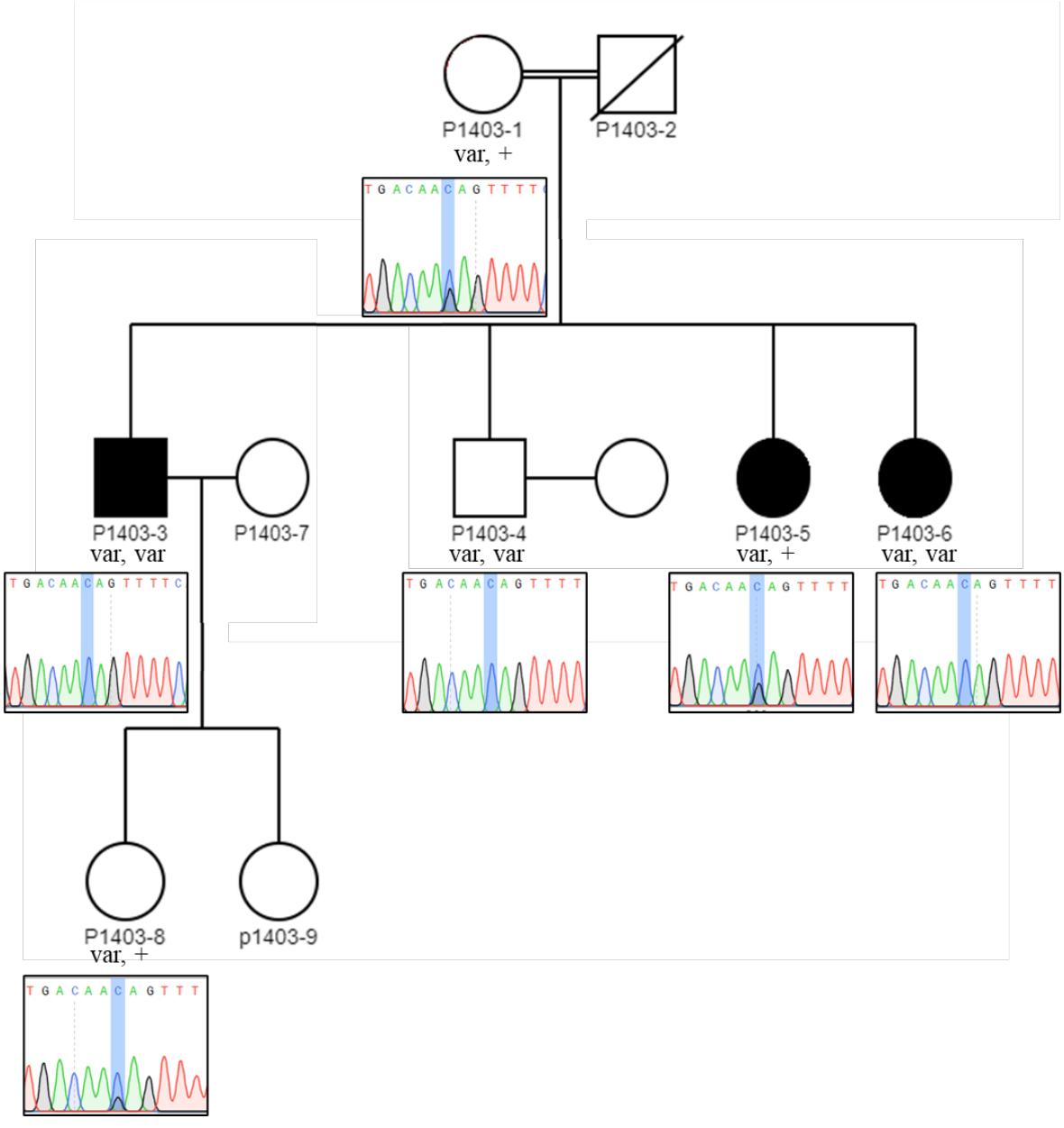

Figure B.1. Pedigree of family P1403 and chromatograms for *FMO2* gene c.1172G>C variant. (var; variant allele, +; native allele)

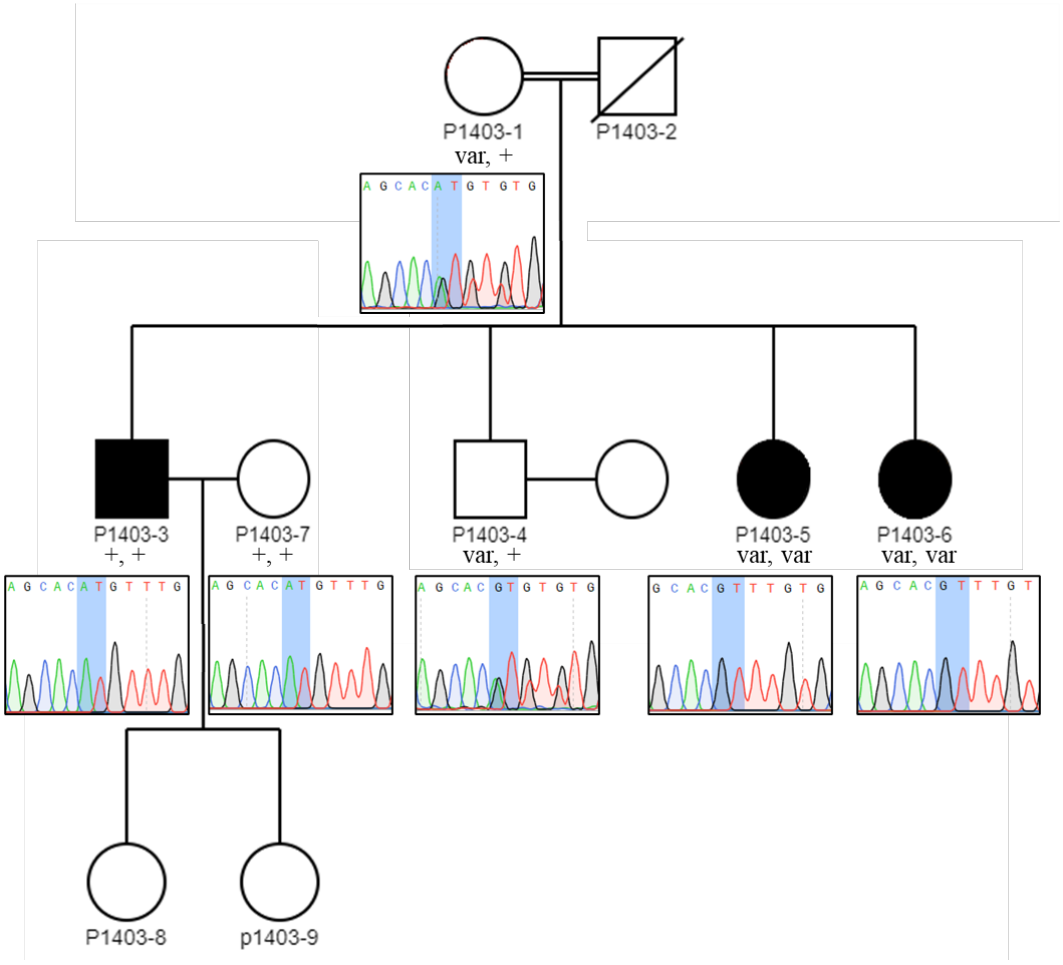

Figure B.2. Pedigree of family P1403 and chromatograms for *MYH7B* gene c.1612_1613del variant. (var; variant allele, +; native allele)

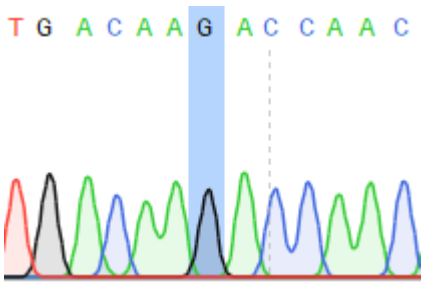

Figure B.3. Chromatogram showing the absence of *CLASP2* gene c.3347C>T variant in patient P1403-6.

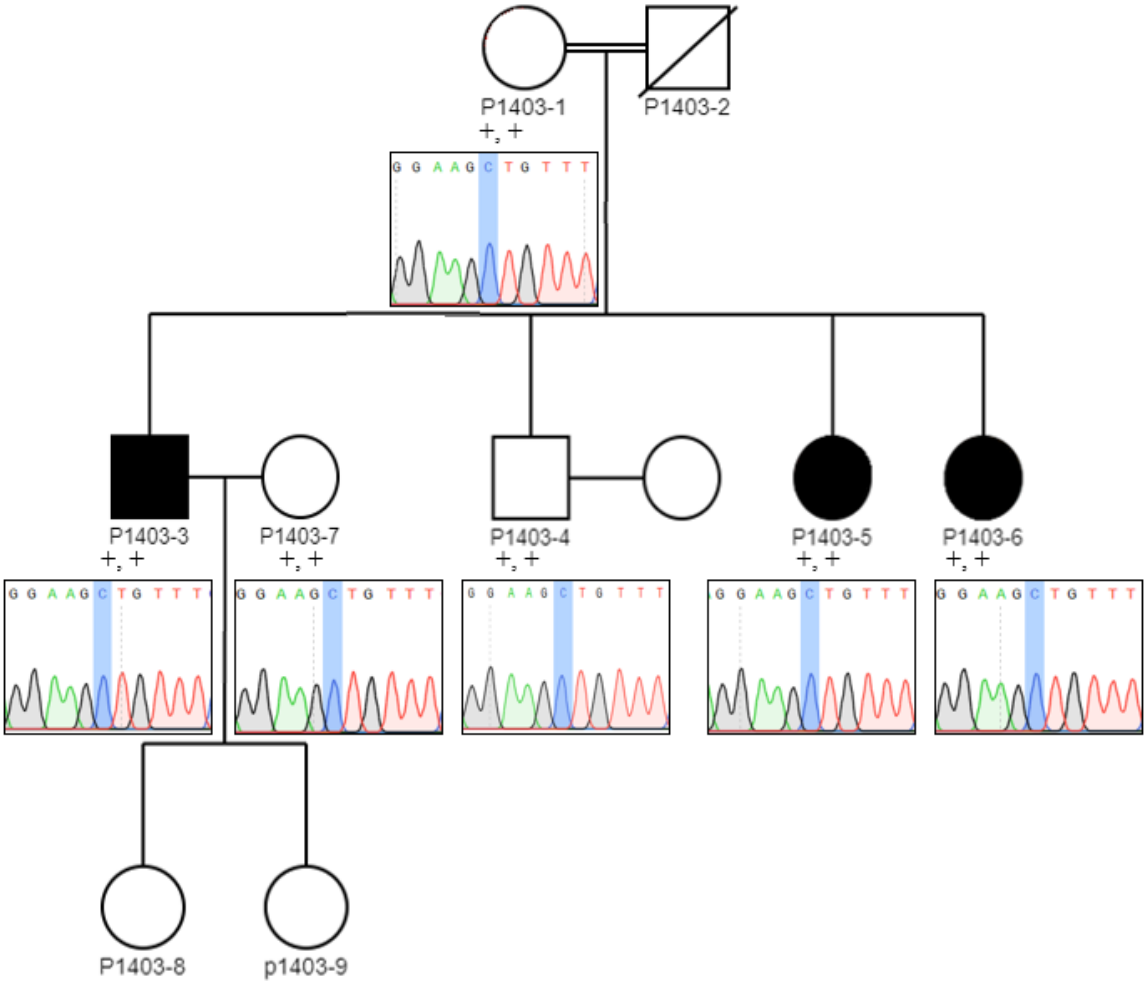

Figure B.4. Pedigree of family P1403 and chromatograms for *ARSD* gene c.845C>A variant. (var; variant allele, +; native allele)

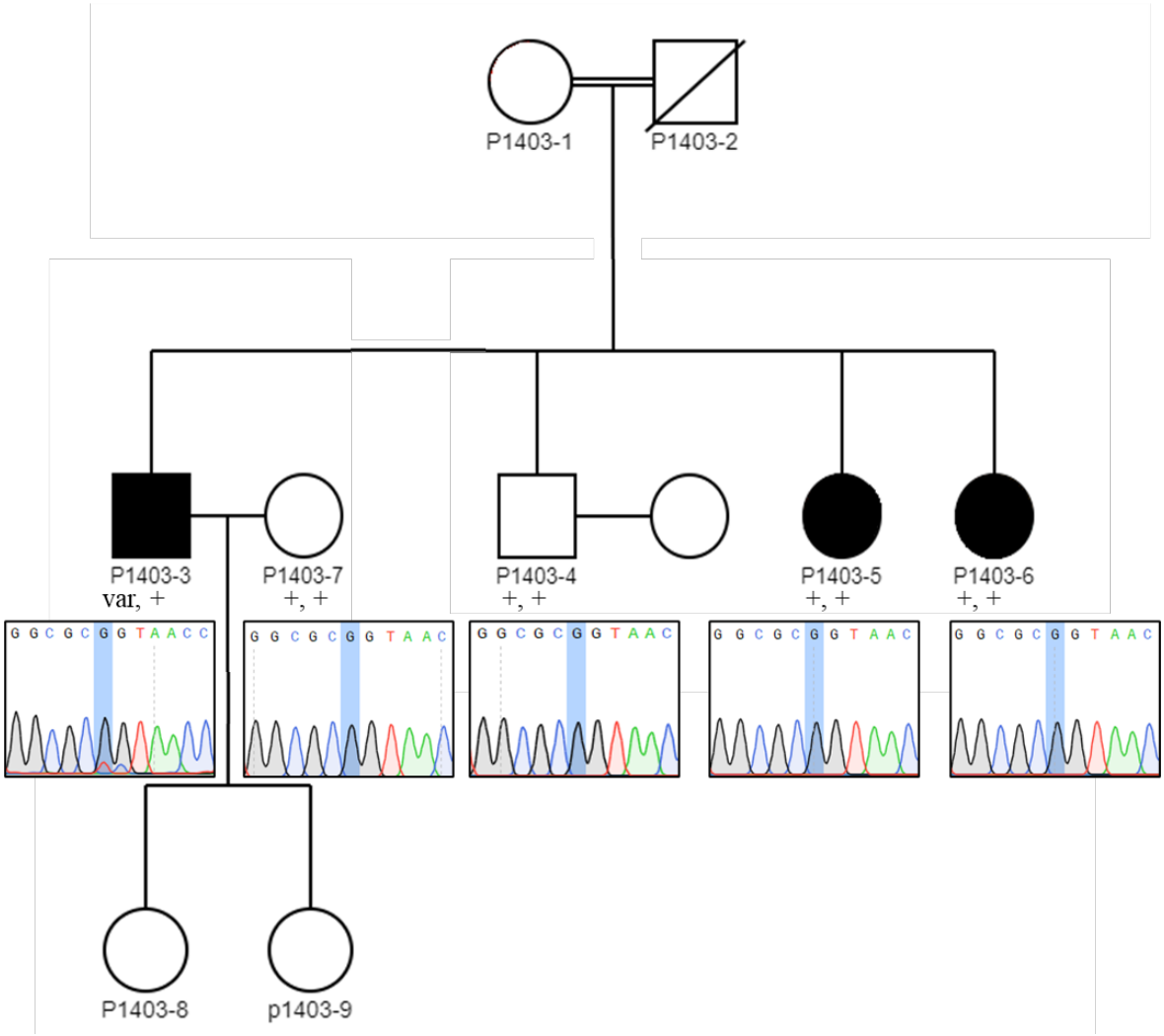

Figure B.5. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.217G>A variant. (var; variant allele, +; native allele)

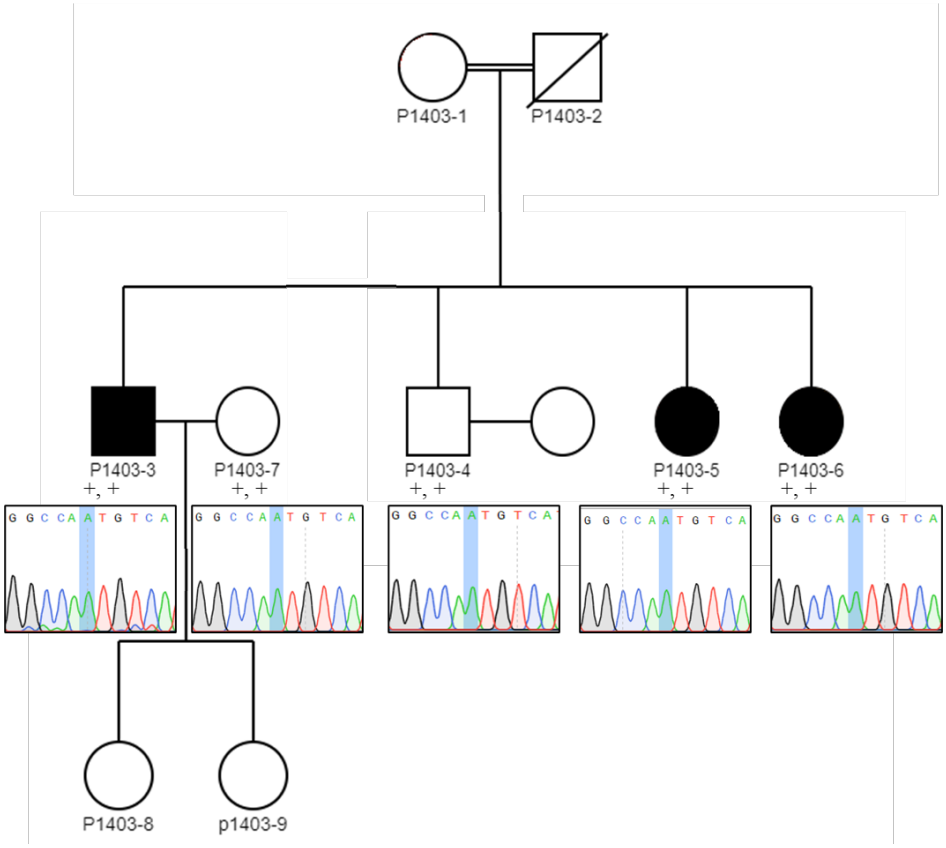

Figure B.6. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.230A>C variant. (var: variant allele, +: native allele)

Table B.1. Sanger Sequencing Results for *SLC25A5* gene c.230A>C variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	var
P1403-5 (Aff. sister)	+, +	var, +
P1403-6 (Aff. sister)	+, +	var, +
Conclusion:	Excluded	

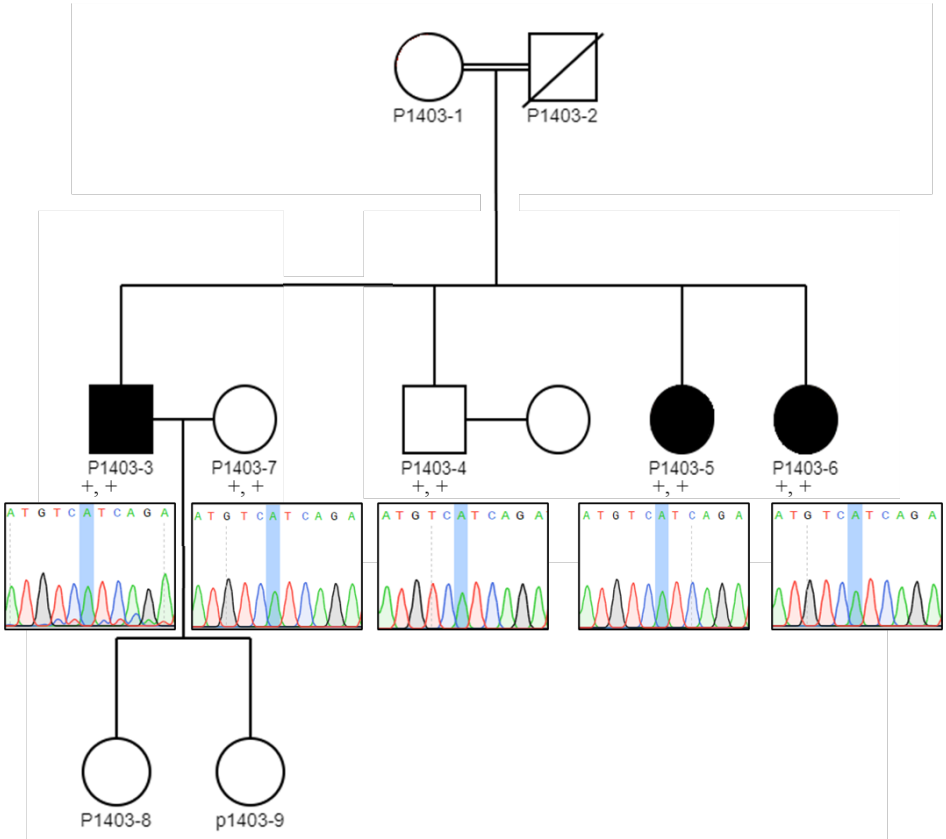

Figure B.7. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.235A>T variant. (var: variant allele, +: native allele)

Table B.2. Sanger Sequencing Results for *SLC25A5* gene c.235A>T variant together with WES data results.(Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	var, +
P1403-6 (Aff. sister)	+, +	var, +
Conclusion:	Excluded	

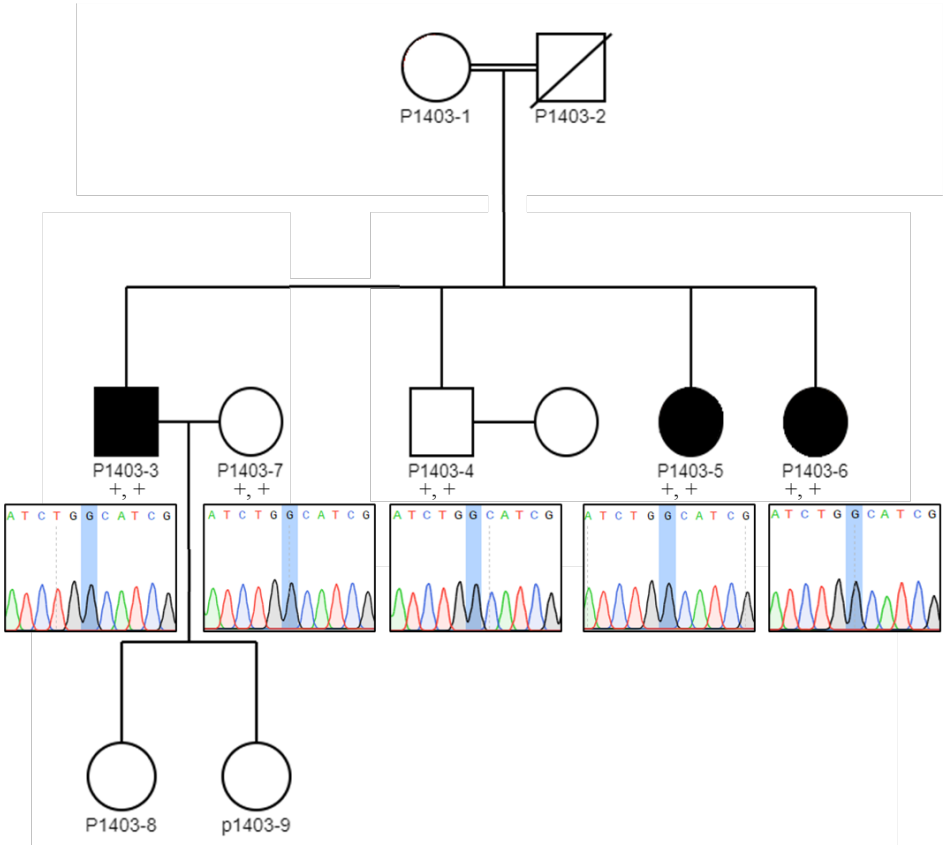

Figure B.8. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.352G>A variant. (var: variant allele, +: native allele)

Table B.3. Sanger Sequencing Results for *SLC25A5* gene c.352G>A variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

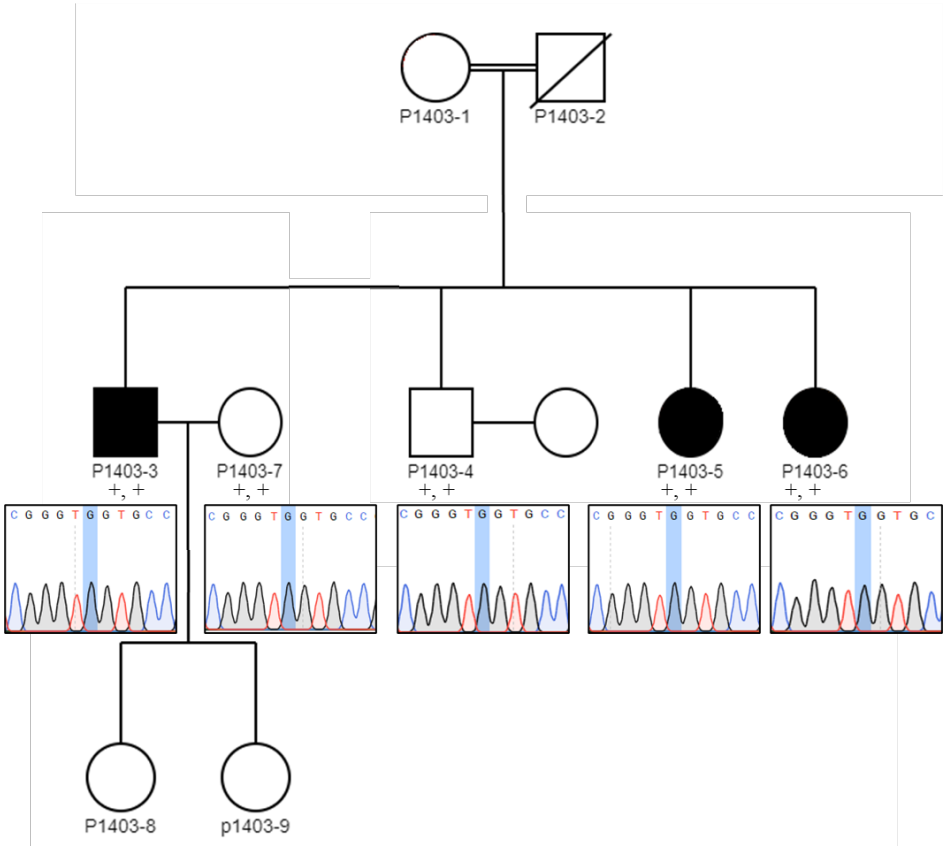

Figure B.9. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.361G>T variant. (var: variant allele, +: native allele)

Table B.4. Sanger Sequencing Results for *SLC25A5* gene c.361G>T variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	+, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

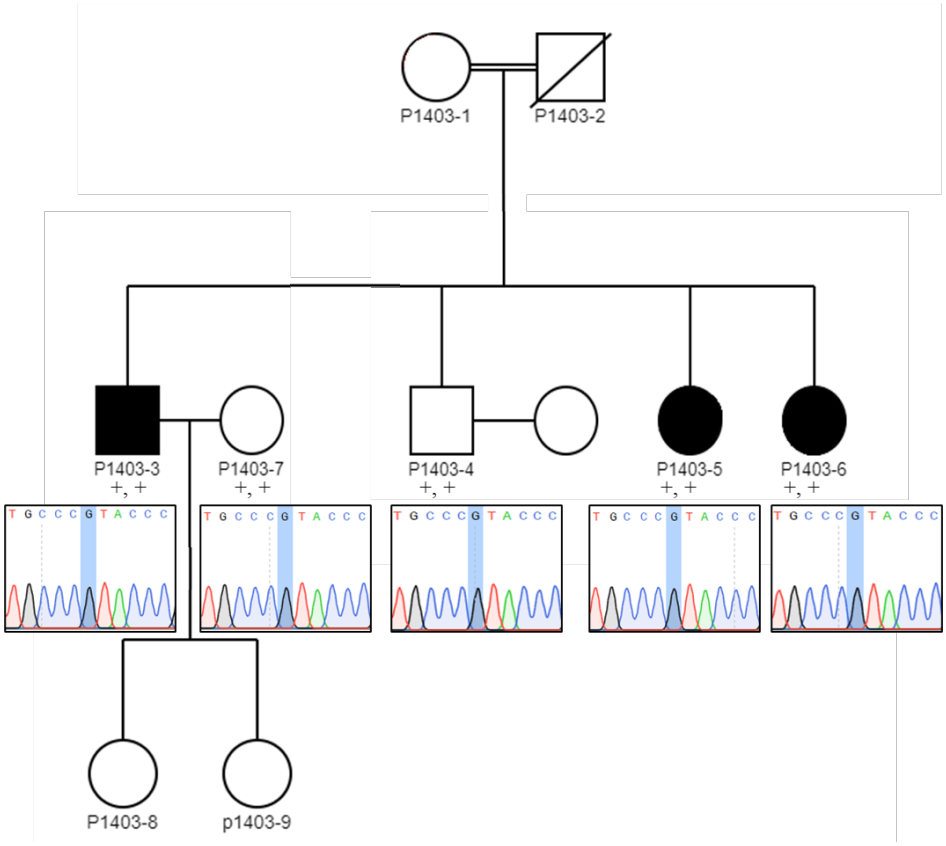

Figure B.10. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.413G>A variant. (var: variant allele, +: native allele)

Table B.5. Sanger Sequencing Results for *SLC25A5* gene c.413G>A variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

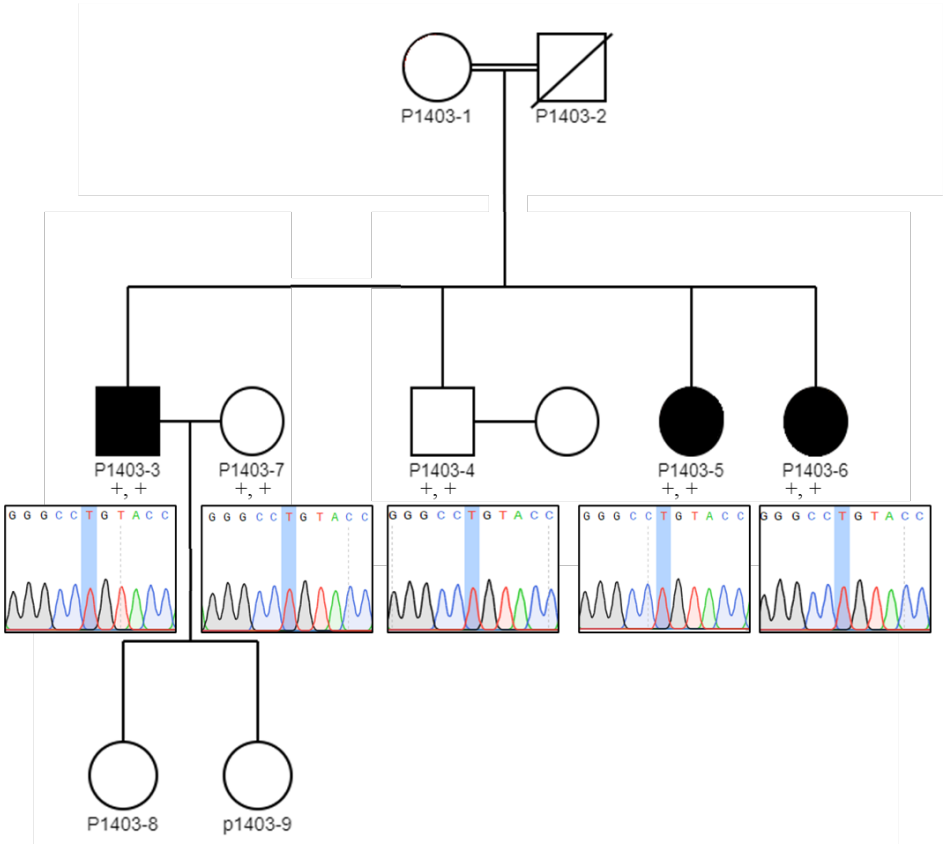

Figure B.11. Pedigree of family P1403 and chromatograms for *SLC25A5* gene c.518T>C variant. (var: variant allele, +: native allele)

Table B.6. Sanger Sequencing Results for *SLC25A5* gene c.518T>C variant together with WES data results. (Var: Variant allele; +: Native allele; Unaff: Unaffected; Aff: Affected.)

***SLC25A5* Gene**	**Sanger Sequencing Results**	**WES Data Results**
P1403-1 (Unaff. mother)	Inconclusive	var, +
P1403-3 (Index case)	+	var
P1403-4 (Unaff. brother)	+	+
P1403-5 (Aff. sister)	+, +	+, +
P1403-6 (Aff. sister)	+, +	+, +
Conclusion:	Excluded	

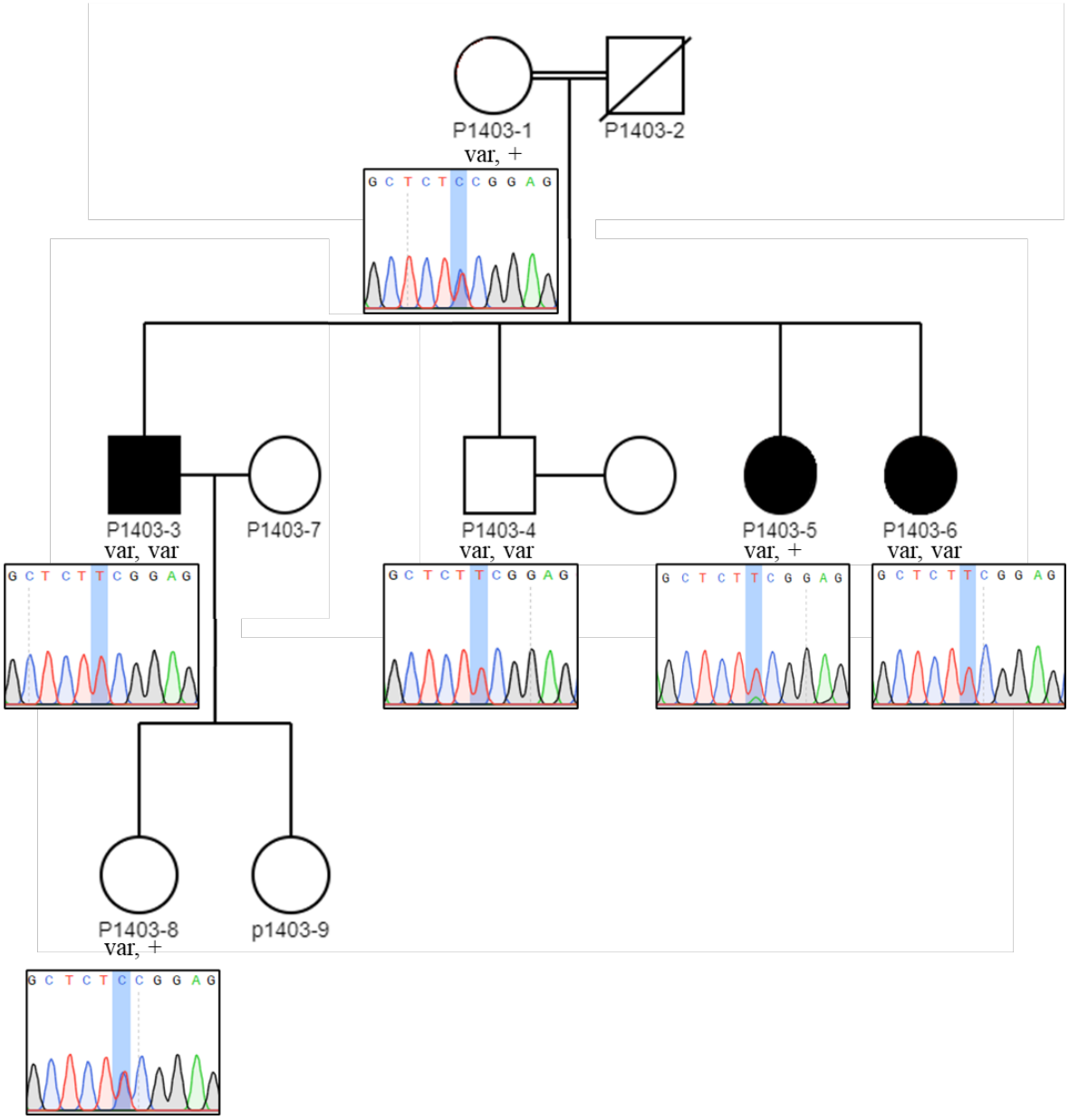

Figure B.12. Pedigree of family P1403 and chromatograms for *FCRL2* gene c.478C>T variant. (var; variant allele, +; native allele)

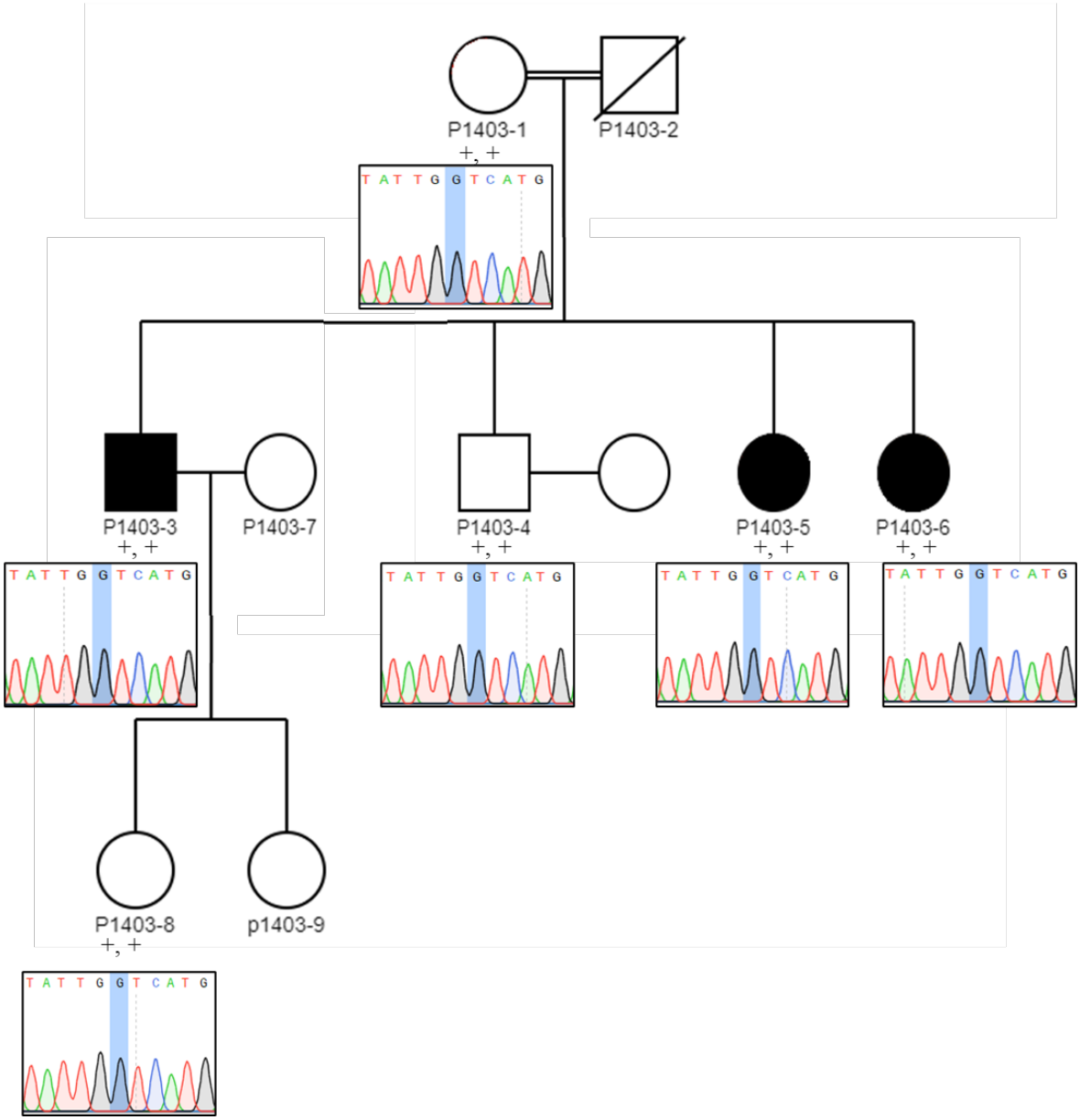

Figure B.13. Pedigree of family P1403 and chromatograms for *POGK* gene c. *1950G>A variant. (var; variant allele, +; native allele)

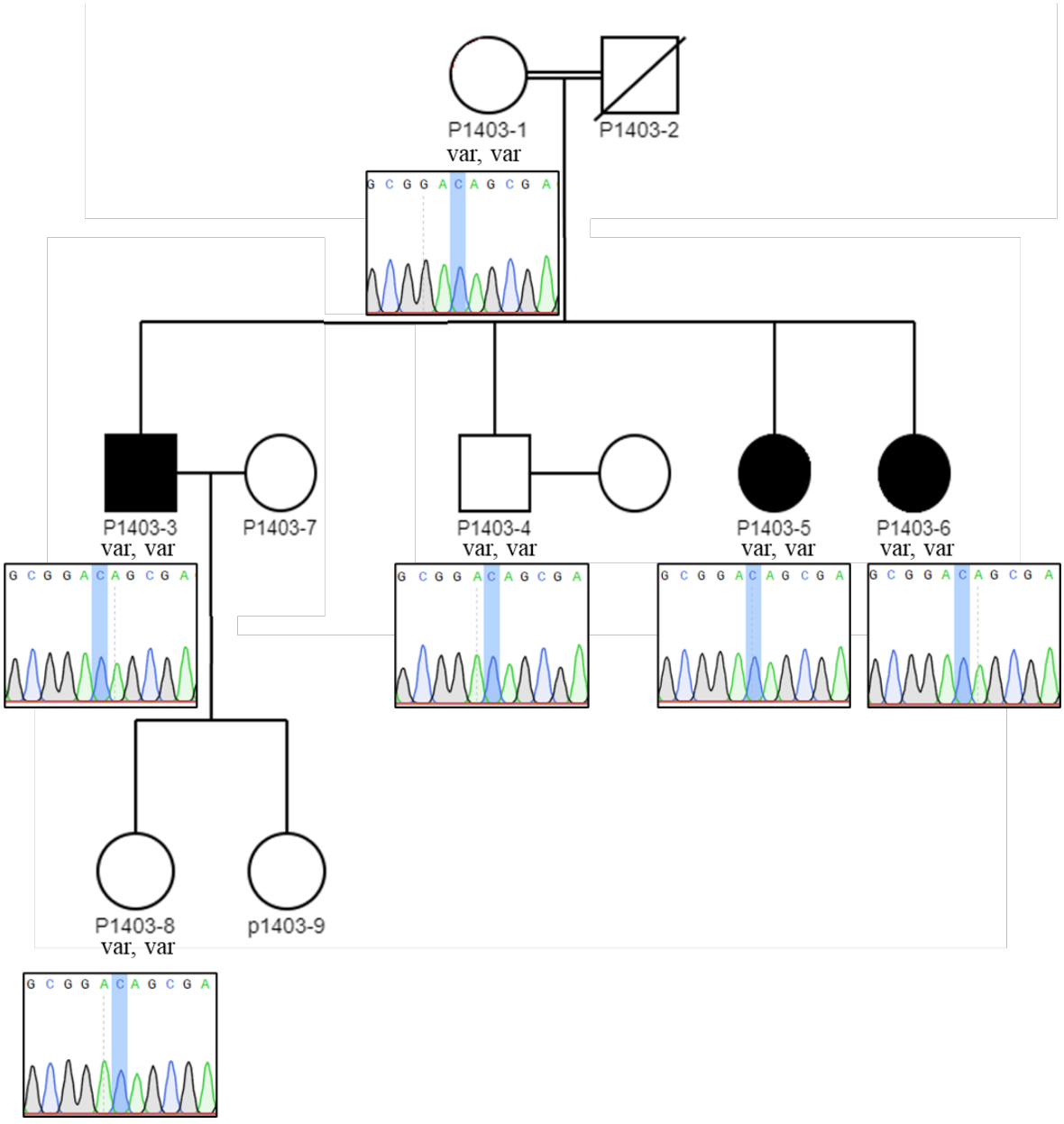

Figure B.14. Pedigree of family P1403 and chromatograms for *MPZL1* gene c.-203T>C variant. (var; variant allele, +; native allele)

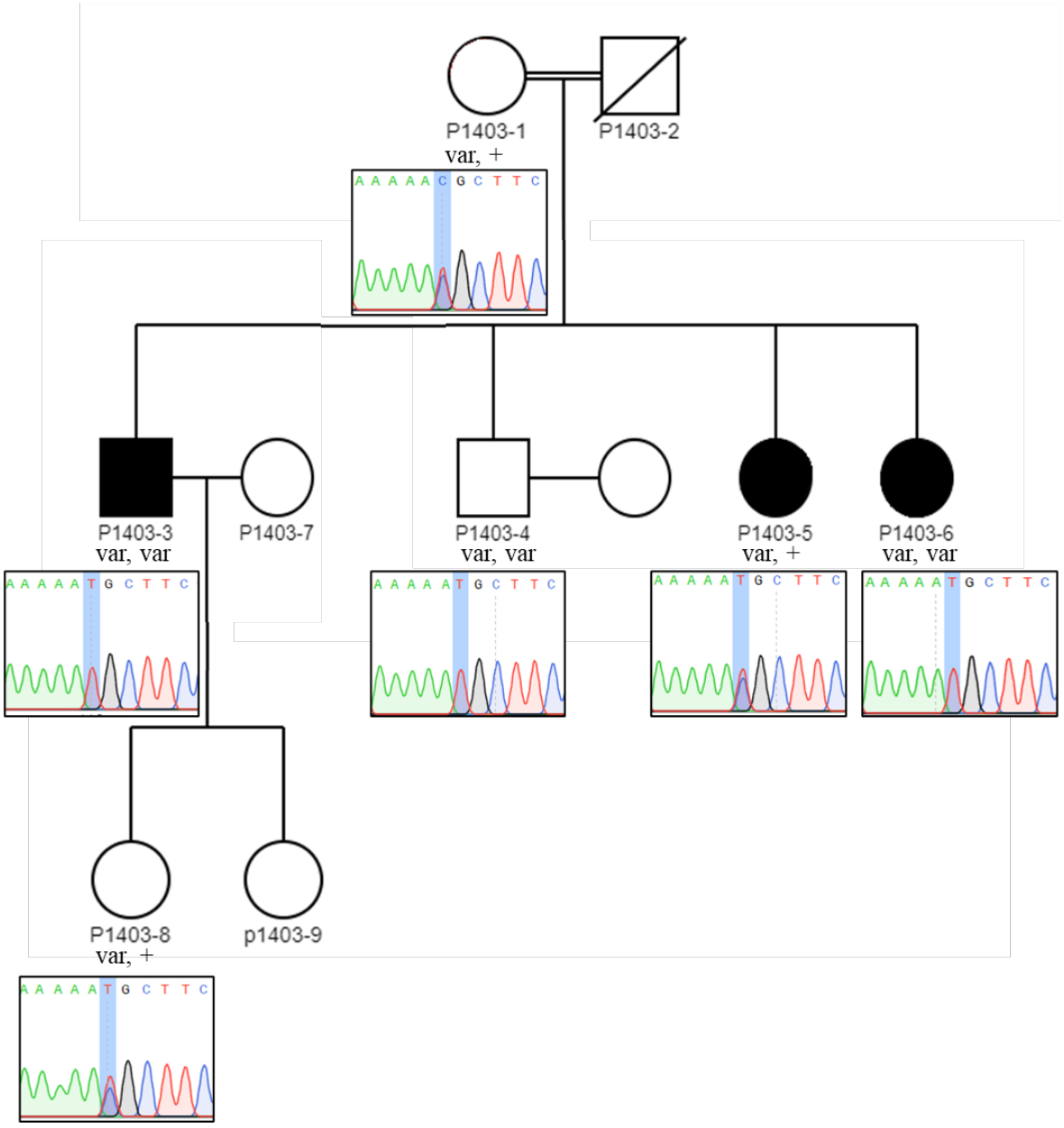

Figure B.15. Pedigree of family P1403 and chromatograms for *FMO2* gene c.-87C>T variant. (var; variant allele, +; native allele)

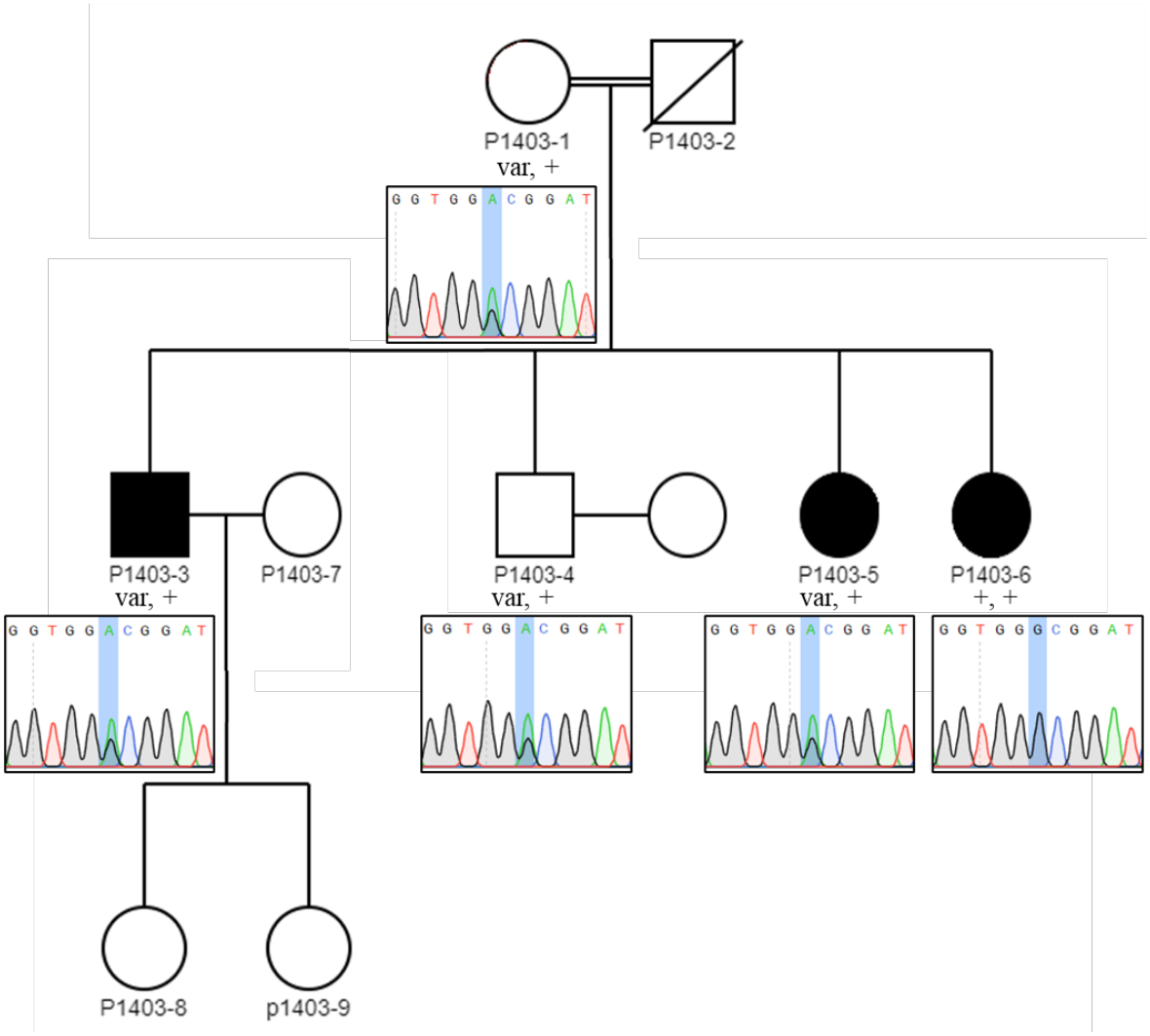

Figure B.16. Pedigree of family P1403 and chromatograms for *MYH14* gene c.4752+197G>A variant. (var; variant allele, +; native allele)

APPENDIX C: HUMAN SPLICING FINDER BIOINFORMATICS TOOL OUTCOMES

The results of the splice site variants analyzed by Human Splicing Finder online prediction program can be found below.

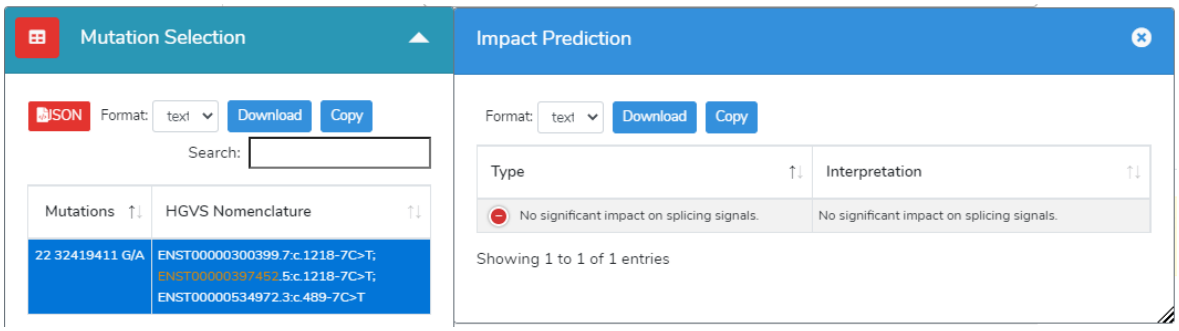

Figure C.1. Human Splicing Finder outcome for *BPIFC* gene c.1218-7C>T variant.

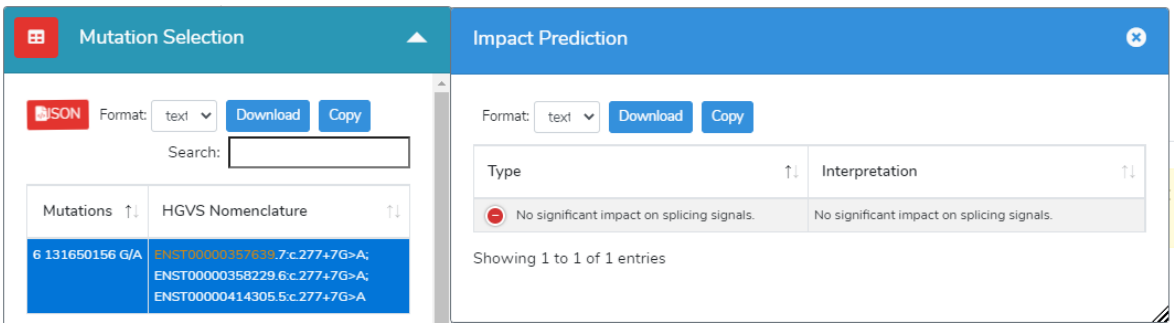

Figure C.2. Human Splicing Finder outcome for *ENPP3* gene c.277+7G>A variant.

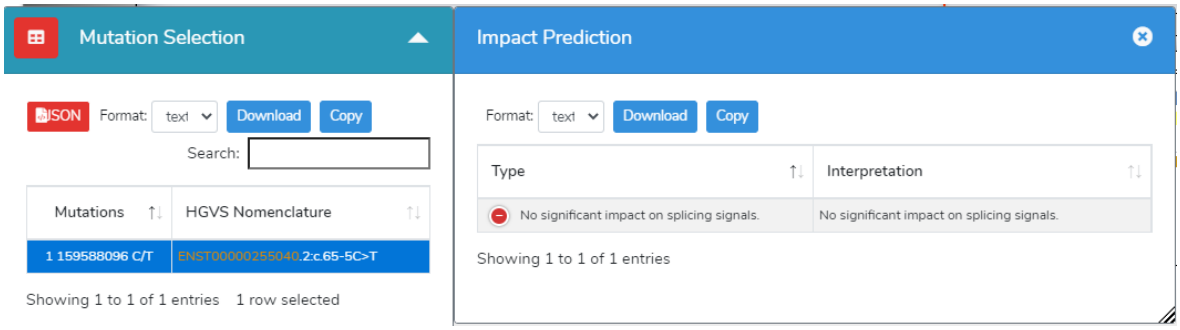

Figure C.3. Human Splicing Finder outcome for *APCS* gene c.65-5C>T variant.

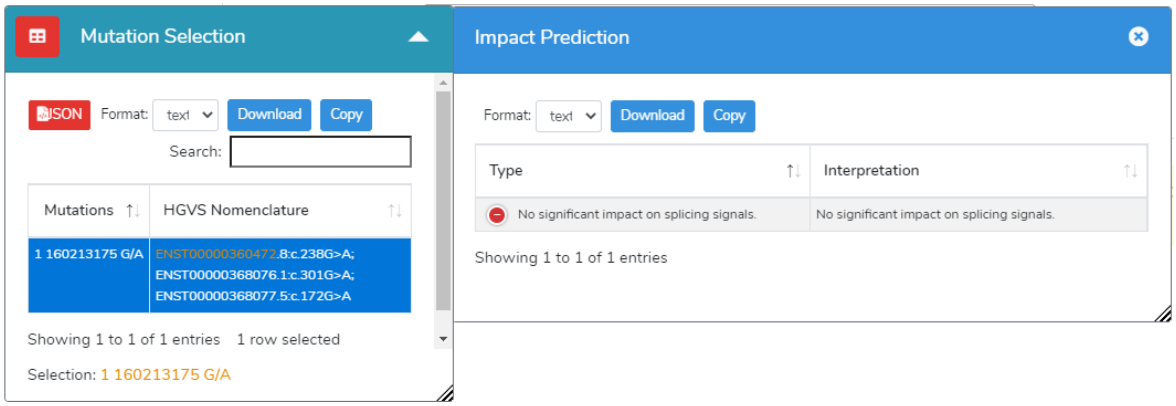

Figure C.4. Human Splicing Finder outcome for *PEA15* gene c.235+3G>A variant.

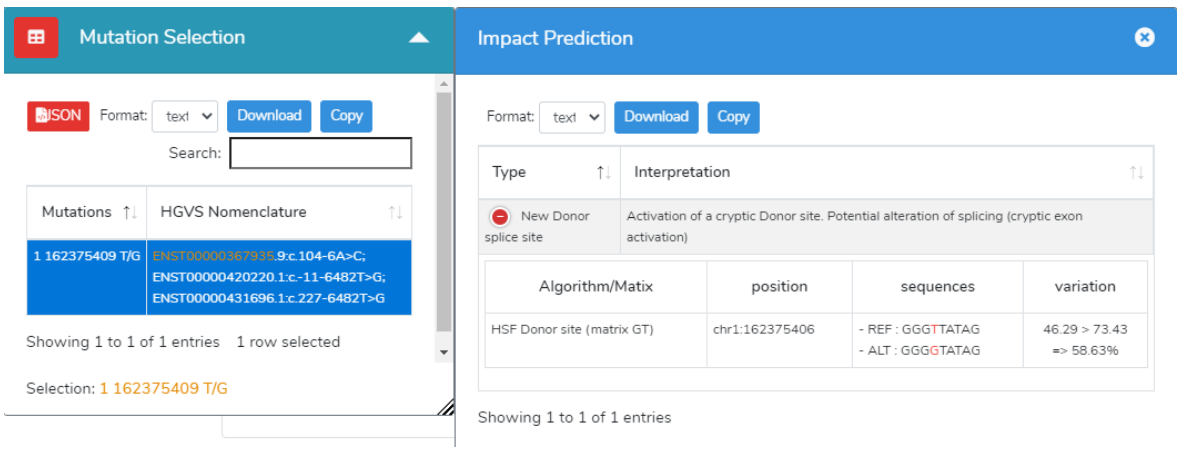

Figure C.5. Human Splicing Finder outcome for *C1orf111* gene c.104-6A>C variant.

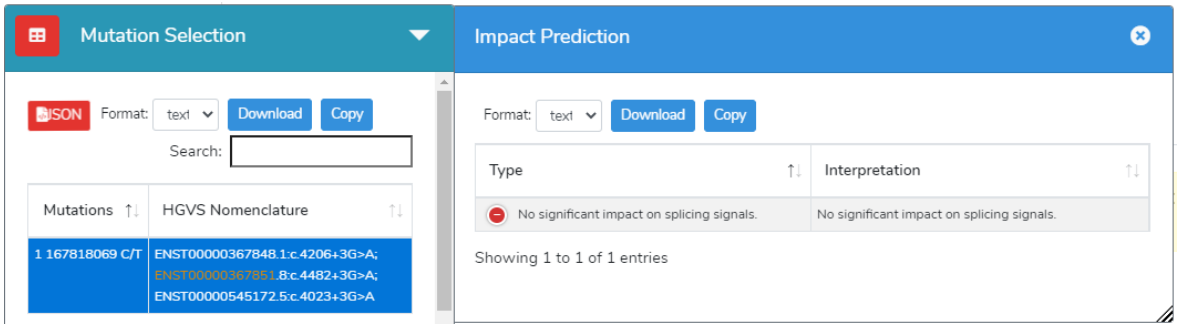

Figure C.6. Human Splicing Finder outcome for *ADCY10* gene c.4482+3G>A variant.

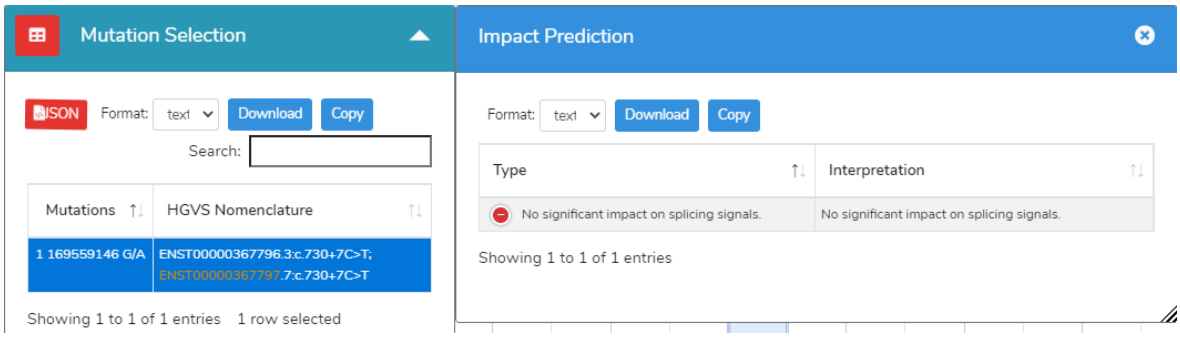

Figure C.7. Human Splicing Finder outcome for *F5* gene c.730+7C>T variant.