

**IDENTIFICATION OF CANDIDATE GENES FOR FAMILIAL
ESSENTIAL TREMOR**

A THESIS SUBMITTED TO
NEUROSCIENCE PROGRAM
OF THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By

MERVE ŞEN

September, 2016

IDENTIFICATION OF CANDIDATE GENES FOR FAMILIAL ESSENTIAL
TREMOR

By Merve Şen,

September, 2016

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Master of Science.

Ayşe Begüm Tekinay (Advisor)

Tayfun Özçelik

Cenk Akbostancı

Approved for the Graduate School of Engineering and Science:

Levent Onural

Director of the Graduate School

ABSTRACT

IDENTIFICATION OF CANDIDATE GENES FOR FAMILIAL ESSENTIAL TREMOR

Merve Şen

M.Sc. in Neuroscience

Supervisor: Ayşe Begüm Tekinay, PhD

September, 2016

Essential tremor (ET) is one of the most common movement disorders in humans and is characterized by action tremors that occur during voluntary motion. However, due to the strong heterogeneity exhibited by ET patients at etiological, clinical and pathological levels, the genetic architecture and pathophysiology of the disease remain largely unknown. In this thesis, whole exome sequencing and pedigree analysis were performed in 3 ET families with histories consistent with an autosomal dominant pattern of inheritance. In two independent families, we observed a rare variant that cosegregated with the disease and was predicted to affect the function of the protein. In one of these families, a homozygous variant was identified in one affected patient and a heterozygous variant was determined in five affected family members. In a second, four-generation Turkish family, the same heterozygous variant was identified in three ET cases while remaining absent in unaffected family members. In addition, whole exome sequencing allowed us to demonstrate that other missense mutation segregate with essential tremor in a different consanguineous Turkish family. Both variants were observed to involve amino acid substitutions of highly conserved domains. Furthermore, both of the affected genes are expressed in the brain and function as

regulatory elements of the central nervous system. Consequently, we propose that these variants are risk factors involved in the etiology of hereditary ET, and suggest that whole exome sequencing can serve as an effective means of identifying other alleles associated with the disease.



Keywords: Essential tremor, human genetics, whole exome sequencing, movement disorders, disease gene identification

ÖZET

AİLESEL ESANSİYEL TREMOR İÇİN ADAY GENLERİN TESPİT EDİLMESİ

Merve Şen

Nörobilim Yüksek Lisans

Danışman: Dr. Ayşe Begüm Tekinay

Eylül, 2016

Esansiyel tremor (ET), insanlarda en sık görülen ve gönüllü hareketlerde aksiyon tremoru ile ortaya çıkan nörolojik bir hareket bozukluğu hastalığıdır. Fakat bu hastalığın genetik nedenleri ve patofizyolojisi bilinmemektedir. Çünkü ET hastaları, etiyolojik, klinik ve patolojik düzeylerde çeşitlilik göstermektedir. Bu çalışmada, ET'nin otozomal dominant kalıtımla aktarıldığı 3 aile, aile ağacı ve ekzon dizilimleri bakımından incelenmiş ve aday mutasyonlar anlatılmıştır. Birbirinden farklı iki ailede nadir bir varyantın proteininin fonksiyonunu etkilediği ve ailede segrege olduğu ortaya çıkarılmıştır. Bu varyant, Türkiye’de yaşayan ve akraba evliliği gözlemlenen 5 kuşaklık bir ailenin hasta bir bireyinde homozigot ve ailenin diğer bireylerinde heterozigot olarak tespit edilmiştir. Başka bir ailenin tüm hasta bireylerinde ise bu aynı varyant heterozigot olarak gözlemlenmiştir. Ayrıca, başka bir akraba evliliği görülen ve Türkiye’de yaşayan diğer bir ailede ise farklı bir gendeki mutasyonun ailede hastalıkla beraber kalıtıldığı keşfedilmiştir. Bu varyantların, evrimsel olarak korunmuş amino asit dizilimlerinde değişikliğe neden olduğu ve proteinlerin önemli bölgelerinde olduğu gösterilmiştir. Hastalıkla ilgili bulunan bu genlerin beyinde ifade edildikleri ve merkezi sinir sisteminde etkilerinin olduğu önceki çalışmalarda bildirilmiştir. Bu çalışma, ET

ile alakalı riskli allellerin keşfedilmesi ile ET'nin genetik ve moleküler mekanizmalarının daha iyi anlaşılmasında önemli bir rol oynayabilir.



Anahtar sözcükler: Esansiyel tremor, insan genetiği, tüm ekzon dizilemesi, hareket bozuklukları, hastalık geni tanımlanması

ACKNOWLEDGEMENT

First and foremost I would like to express my gratitude to my advisor Dr. Ayşe Begüm Tekinay, for the continuous support of my Master's study, for her inspiring guidance, motivation, encouragement and immense knowledge. I consider it as a great opportunity to complete my Master's under her guidance.

I would like to gratefully and sincerely thank Prof. Dr. Tayfun Özçelik for his continuous support, invaluable guidance and his immense knowledge in the field of human genetics.

I am thankful to Prof. Dr. Cenk Akbostancı and Prof. Dr. Kader Karlı Oğuz for the identification and recruitment of patients, for brain imaging studies and for clinical tests.

I am thankful to my best friends from Zeynep Orhan, Fatih Yergöz, İdil Uyan, Nurcan Haştar and Gökhan Günay from Izmir Institute of Technology and also Hatice Kübra Kara, for their emotional support, encouragement and motivation in my research. I will remember those times that we shared together.

I would like to express my sincere thanks to İdil Uyan and Ahmet Emin Topal for their unforgettable and enjoyable friendship together with laugh forever.

I would like to thank Dr. Seher Yaylacı and İslam Oğuz Tuncay for their friendship and persistent support and excellent collaboration during my research. I would also thank new lab member Neva Yaylacı for her moral support during my thesis.

I would like to acknowledge the financial assistance of The Scientific and Technological Research Council of Turkey (TÜBİTAK) under the BİDEB 2210/C Master Scholarship.

I would like to thank Dr. Emre Onat for his help in my experiments and data analysis and his continuous support during my study.

I would like to thank Alper Devrim Özkan for his support during the thesis writing and editing.

I would like to thank all NBT, BML and UNAM family members and especially Mustafa Beter, Çağla Eren, İbrahim Çelik, Canelif Yılmaz, Özge Uysal, Faruk Okur, Mustafa Fadlelmula, Nuray Gündüz, Tuğçe Önür, Side Selin Su Yirmibeşoğlu, Recep Erdem Ahan and Umut Taşdelen for creating such a warm working environment.

I would like to express my most sincere thanks to Yunus Mermerci and Adile Mermerci Foundation for their help, support and valuable suggestions during my study.

I am thankful to all the Essential tremor patients, relatives and other participants for their cooperation in this study.

I would like to thank my aunts Nuray Özer, Şenay Tatlıcı and Songül Koyutürk and my uncles Nejdet Özer, Sabri Özdemir, Remzi Özdemir and Mustafa Tatlıcı for their support and motivation during my research. I would also thank my all 16 cousins for their support and encouragement during my study.

Finally, I would like to express my deep thanks to my parents Gülay Özdemir and Süleyman Şen and my sisters Duygu Arife Tarım, Nazlı Şen and my spiritual brother Evren Tarım, my grandmother Emine Özdemir and my first nephew Emir Tarım whose support, understanding and love have enabled me to reach the present position in life.

Thank you,

Merve ŞEN

CONTENTS

CHAPTER 1	1
Introduction	1
1.1 Introduction to Essential Tremor	1
1.1.1 Tremor in History	1
1.1.2 Clinical Features of Essential Tremor	1
1.1.3 Diagnosis of Essential Tremor	4
1.1.4 Pathology of Essential Tremor	4
1.1.5 Risk and Etiological Factors for Essential Tremor	8
1.1.6 Genetic Heterogeneity in Essential Tremor	10
1.1.7 Treatment of Essential Tremor	15
1.2 Disease Gene Identification Strategies	16
1.2.1 Traditional Methods	16
1.2.2 Next Generation Sequencing	18
1.2.3 Whole Exome Sequencing Approach	19
1.2.4 Identification of Disease-Causing Genes	22
1.3 Consanguinity	25
1.4 Outline of the Thesis	27
CHAPTER 2	28
Material and Methods	28
2.1 Recruitment of Family Members	28
2.2 Clinical Investigations	28
2.3 DNA Isolation from Family Members	29
2.4 Library construction and whole exome sequencing	30
2.4.1 Sample Selection and Preparation	30
2.4.2 Library Construction	30

2.5	Bioinformatics	31
2.5.1	Whole Exome Sequencing Data Analysis	31
2.5.1	Filtration and Prioritization of Candidate Gene Variants	32
2.6	Sanger Sequencing	33
2.7	Model 3D Protein Structure of Candidate Variants	33
CHAPTER 3.....		34
Results.....		34
3.1	Matrix Metalloproteinase 19 (MMP19), p.Arg456Gln Missense Mutation Causes Essential Tremor	34
3.1.1	Clinical Features of the Families	34
3.1.2	Analysis of Whole Exome Sequencing Data.....	41
3.1.3	Variant Annotation and Effect Prediction	45
3.1.4	Identification of Disease Causing Determinants.....	50
3.1.5	Segregation Analysis of Candidate Variants for ET-5 and ET-49 Families 58	
3.2	IMPG1, Interphotoreceptor Matrix Proteoglycan 1, p.Arg580Cys is a Missense Mutation that Causes Essential Tremor in a Consanguineous Family	68
3.2.1	Clinical Features of the Family.....	68
3.2.2	Analysis of Whole Exome Sequencing Data	73
3.2.3	Functional Annotation of Variants for the ET-31 Family	81
3.2.4	Identification of Disease-Causing Determinants	81
3.2.5	Segregation Analysis of Candidate Variants for ET-31 Family	86
3.2.6	Modelling 3D structure of IMPG1 p.Arg580Cys	92
CHAPTER 4.....		94
Discussion.....		94
4.1	MMP19 p.Arg456Gln as the Disease Causing Mutation for Two Different Families with Essential Tremor.....	94

4.2 IMPG1 p.Arg580Cys as the Disease-Causing Mutation for a Consanguineous Family with Essential Tremor	99
CHAPTER 5.....	103
Conclusion and Future Perspectives	103
BIBLIOGRAPHY	105
APPENDIX A.....	123



LIST OF FIGURES

Figure 1.1 Section of the dentate nucleus from an individual with essential tremor.....	7
Figure 1.2 Multiple Lewy bodies are seen in a section of the locus ceruleus from a patient with essential tremor	7
Figure 1.3 Segregation of HTRA2 p.G399S in a six-generation consanguineous Turkish family, illustrating essential tremor.	14
Figure 1.4 Overview of an exome sequencing pipeline.....	21
Figure 1.5 Global prevalence of consanguinity.	26
Figure 3.1 Pedigrees of ET-5 and ET-49 families showing the affected and unaffected individuals with essential tremor	37
Figure 3.2 Archimedes spiral drawings of ET-5 family members.....	39
Figure 3.3 Archimedes spiral drawings of ET-49 family members.....	40
Figure 3.4 Density measurements using agarose gel electrophoresis.....	42
Figure 3.5 Percentage of variants with respect to genomic context for ET-5 and ET-49 families after SnpEff analysis	46
Figure 3.6 Filtration and prioritization pipeline for disease causing determinants.....	51
Figure 3.7 Pedigree of ET-5 family showing the segregation of essential tremor	59
Figure 3.8 Segregation analysis results of the ET-5 family.....	60
Figure 3.9 Pedigree of ET-49 family showing the segregation of ET with the MMP19 p.Arg456Gln	61
Figure 3.10 Segregation analysis results of the ET-49 family.....	62
Figure 3.11 Predicted protein and domain structure of MMP19.....	65
Figure 3.12 Sequence homology of MMP19 protein p.Arg456.....	65
Figure 3.13 3D model protein structure of MMP19 protein.....	67
Figure 3.14 Pedigree of ET-31 family showing the affected and unaffected individuals with ET.....	70
Figure 3.15 Archimedes spiral drawings of ET-31 family members.....	72
Figure 3.16 Density measurements using agarose gel electrophoresis.....	74
Figure 3.17 Percentage of variants with regards to genomic context for the ET-31 family.....	78
Figure 3.18 Pedigree of ET-31 family segregating essential tremor, with genotypes at IMPG1 p.Arg580Cys.	87

Figure 3.19 Segregation analysis result of ET-31 family.	88
Figure 3.20 Predicted protein and predicted domain structure of IMPG1 protein	91
Figure 3.21 Sequence homology of IMPG1 p.Arg580	91
Figure 3.22 3D model protein structure of IMPG1 protein.	93
Figure 4.1. Chromosomal localization of <i>MMP19</i> , protein domain organization of <i>MMP19</i> and the position of the detected mutation.	98
Figure 4.2 Chromosomal localization of <i>IMPG1</i> , protein domain organization of SPACR and position of detected mutation	102



LIST OF TABLES

Table 1.1 Classification of tremors and their characteristics	3
Table 1.2 Heritability estimation for selected diseases.....	9
Table 1.3 Mendelian disease gene identifications by exome or genome sequencing..	24
Table 3.1 Clinical characteristics of tremor in affected and unaffected individuals of families ET-5 and ET-49	38
Table 3.2 Concentrations and qualities of DNA samples isolated from ET-5 and ET-49 families.....	42
Table 3.3 Statistics on the genome sequence produced by the paired-end whole exome sequencing.....	44
Table 3.4 Statistics on all variants based on their regions, as identified through whole exome sequencing and SnpEff analysis.	47
Table 3.5 Statistics on all variants based on their annotation effects, identified through SnpEff analysis.	48
Table 3.6 List of the homozygous damaging variants in proband of the ET-5 family	55
Table 3.7 List of the homozygous variants in proband of the ET-49 family.....	55
Table 3.8 List of the heterozygous variants in the proband of the ET-5 family ... Error!	
Bookmark not defined.	
Table 3.9 Clinical characteristics of tremor in affected and unaffected individuals of the ET-31 family	71
Table 3.10 Concentrations and qualities of DNA samples from the ET-31 family.....	74
Table 3.11 Statistics on the genome sequence produced by the paired-end whole exome sequencing for ET-31 family	76
Table 3.12 Statistics on all variants based on their regions, identified through whole exome sequencing and SnpEff analysis for ET-31 family individuals.....	79
Table 3.13 Statistics of all variants based on their annotation effects, identified through SnpEff analysis for the ET-31 family	80
Table 3.14 List of the homozygous damaging variants in proband of the ET-31 Family	84
Table 3.15 List of the heterozygous variants in proband of the ET-31 Family Error!	
Bookmark not defined.	85

ABBREVIATIONS

BWA	Burrows-Wheeler Aligner
CNS	Central Nervous System
ESP	Exome Sequencing Project
ET	Essential Tremor
ExAC	Exome Aggregation Consortium
GERP	Genomic Evolutionary Rate Profiling
GWAS	Genome-Wide Association Study
HTRA2	High Temperature Requirement protein A2
IMPG1	Interphotoreceptor Matrix Proteoglycan 1
ISH	In Situ Hybridization
MAF	Minor Allele Frequency
MIM	Mendelian Inheritance Of Man
MMP19	Matrix Metalloproteinase 19
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSA	Multiple Sequence Alignment
NGS	Next Generation Sequencing
PD	Parkinson's Disease
RT-PCR	Real Time Polymerase Chain Reaction
SAMtools	Sequence Alignment/Map Tools
SNP	Single Nucleotide Polymorphism

Uniprot	Universal Protein Resource
VCF	Variant Call File
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

URLs

1000 genomes project	1000genomes.org
CADD	cadd.gs.washington.edu/
ExAC	http://exac.broadinstitute.org/
GATK	broadinstitute.org/gatk/
MutationAssesor	mutationassessor.org
MutationTaster	mutationtaster.org
NHLBI exome sequencing project	evs.gs.washington.edu/EVS/
Polyphen-2	genetics.bwh.harvard.edu/pph2/
PROVEAN	provean.jcvi.org
SIFT	sift.jcvi.org/
Swiss-PdbViewer	spdbv.vital-it.ch/

CHAPTER 1

Introduction

1.1 Introduction to Essential Tremor

1.1.1 Tremor in History

Written accounts of pathological tremors are known from ancient India (5000-3000 BC), Egypt (700 BC) and Greece (400 BC)¹. In addition, records left by physicians such as Galen of Pergamon (130 - 200 A.D.), Sylvius de la Boe (1680), Van Swieten (1745) and Sauvages (1768) contain detailed descriptions of patients exhibiting action and resting tremors, suggesting that the distinction was recognized by these authors¹. Action tremors are the characteristic clinical feature of ET and occur during voluntary movement, in contrast to resting tremors that affect stationary extremities². The first modern description of pathological tremor was provided in 1887 by Dr. Charles Dana, a New York neurologist who identified this condition in several large families¹.

1.1.2 Clinical Features of Essential Tremor

Essential tremor (ET [OMIM 190300]) is one of the most common movement disorders, with a frequency of 1% in the population³⁴. The prevalence of ET is also known to be higher among the elderly, with an estimated frequency of 5% in the 65+ age group³. Shaking may occur while patients voluntarily keep a steady posture (postural tremor) or while they move (kinetic tremor)^{5,2}. Kinetic tremor appears during the movement of a body part, such as the rotary movements of the forearm. Postural tremor appears while maintaining a position against gravity, such as holding an outstretched arm in place. Kinetic tremors in the 4-12 Hz range are the characteristic

motor feature of ET (Table 1). A trembling voice and slight movements of the chin and head at rest are other symptoms that might be observed in patients. However, ET is not a monosymptomatic disorder and may include a broad range of symptoms, such as parkinsonism, dystonia, sensory abnormalities, cognitive and psychiatric features and sleep disorders⁶. Recent studies show that certain non-motor disorders, such as cognitive problems, depression, anxiety, social phobia, personal and behavioral changes and hearing and smell disorders are also more frequent in ET patients compared to healthy individuals⁷.

ET causes serious difficulties in daily activities such as eating, drinking and writing by interfering with motor control⁸. ET has an effect on upper limbs in at least 95% of affected individuals. Its effects on other body parts are less frequent: at least 30% of patients have head tremor, at least 20% have voice tremor, at least 10% have chin and face tremor and at least 10% have lower limb tremor^{9,10}. A mild disturbance in gait can also occur in some patients⁹.

In addition, studies in different populations suggest that essential tremor may cause mild to severe levels of attention deficit, memory loss and impairments in executive functions in patients¹¹.

Table 1.1 Classification of tremors and their characteristics

Tremor type		Frequency	Amplitude	Feature	Disease
Resting		Low-moderate	High (increases with movement)	Muscles are not voluntarily activated and the limb is fully supported against gravity	-Parkinson's disease -Drug-induced parkinsonism
Action Tremor					
Kinetic Tremor	Intentional tremor	Low (<5Hz)	Increases with target directed movement	During any type of movement	-Cerebellar lesions (Stroke, Multiple Sclerosis) -Drug-induced (Lithium, alcohol)
	Basic Kinetic Tremor	Variable (3-10 Hz)	Does not change with targeted movement		
	Task specific tremor	Variable (4-10 Hz)	Variable	During specific movement	-Tremor while writing -Tremor in musicians
Postural Tremor		Moderate-High	Low, increases with voluntary movement	When the limb is positioned against gravity	-Physiological tremor -Essential tremor -Metabolic disorders -After alcohol or drug withdrawal

1.1.3 Diagnosis of Essential Tremor

No specific markers or laboratory tests exist for the diagnosis of ET; medical history and neurological examination are the only criteria for diagnosis. Consequently, the disease is misdiagnosed in approximately 30-50% of cases. The Movement Disorder Society (MDS) has developed a set of consensus criteria in an effort to minimize diagnostic errors¹².

According to Deuschl *et al.*, ET can be classified under three groups: sporadic, hereditary and senile. A patient is diagnosed with hereditary ET if they fulfill the consensus criteria of definite (or *classical*) ET and have at least one other family member that is affected by the disease. In addition, the age of onset should be prior to age 65 for both the patient and at least one affected family member¹². In sporadic cases, the patient should fulfill consensus criteria for classical ET at an age of onset lower than 65, and have no other family members with a history of ET. In senile ET, the patient should satisfy the criteria for definite or classical ET with an age of onset after 65 and no family history of the disease¹⁰.

1.1.4 Pathology of Essential Tremor

Pathogenesis describes the totality of the changes that occur in an organism following the onset of disease, including the gradual alterations that accompany its progression. The pathogenesis of ET is not fully understood. According to an initial report with 10 ET cases and 12 controls, pathological findings associated with ET are highly heterogeneous¹³. These patients could nevertheless be grouped into two categories, based on histopathological examinations of the brain: the first group was characterized by cerebellar degeneration and contained four patients, while the second exhibited Lewy bodies on the brainstem and had six patients¹³.

Cerebellar essential tremor was characterized by large numbers of Bergmann astrocytes and axonal torpedoes in Purkinje cell axons. Bergmann astrocytes form as a non-specific response against neuronal injuries, while axonal torpedoes are swollen regions at the proximal regions of Purkinje cell axons that contain disoriented neurofilaments and may cause Purkinje cell loss^{13,14}. Further cerebellar abnormalities were also identified in ET patients, including neuronal loss, microglial clusters, pallor of the white matter and neuronal atrophy in the dentate nucleus¹⁴ (Figure 1.1). These cerebellar changes are unexpected because it is believed that tremor is mediated by cerebellothalamocortical fibers². Nevertheless, magnetic resonance imaging studies clearly demonstrate the presence of neuronal loss in the cerebellum^{15,16}.

Instead of cerebellar degradation, a characteristic pattern of Lewy bodies in the brainstem was observed in the remaining six cases^{13,17}. These individuals had numerous Lewy bodies in the locus ceruleus, while fewer bodies were present in the substantia nigra (Figure 1.2). This arrangement of Lewy bodies has not been detected in PD patients and in normal age-matched controls^{13,17}. However, it has been suggested that ET patients with Lewy bodies in the locus ceruleus are under the risk of developing PD later in life².

The precise relationship between kinetic tremor and Lewy bodies in the locus ceruleus is not known. However, the locus ceruleus is a well-recognized source of norepinephrine in the central nervous system (CNS), and axons from this region extend into the cerebellum to form synapses with γ -aminobutyric acid (GABA)-producing Purkinje cells¹³. The degeneration of the locus ceruleus could compromise this connection, preventing the stimulation of Purkinje cells and decreasing the production of GABA in the cerebellum¹³. As such, medical intervention to restore GABA

production may serve to ameliorate the effects of ET in patients showing Lewy bodies in the locus ceruleus^{18,19}.



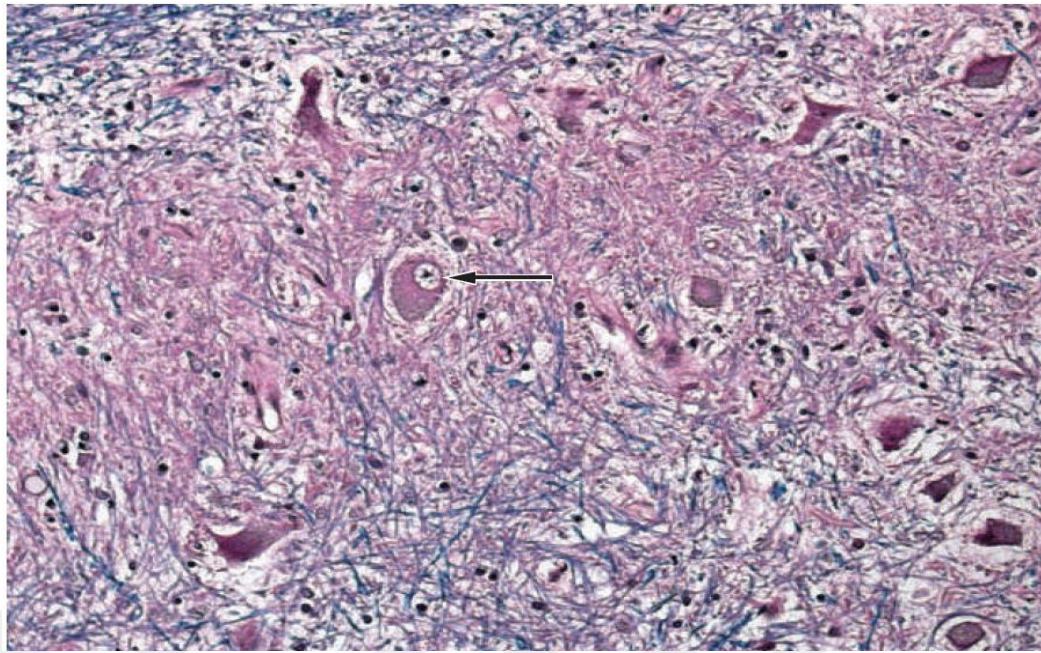


Figure 1.1 Section of the dentate nucleus from an individual with essential tremor, showing neuronal loss. (Copyright 2005 Elsevier. From Louis et al.² with permission)

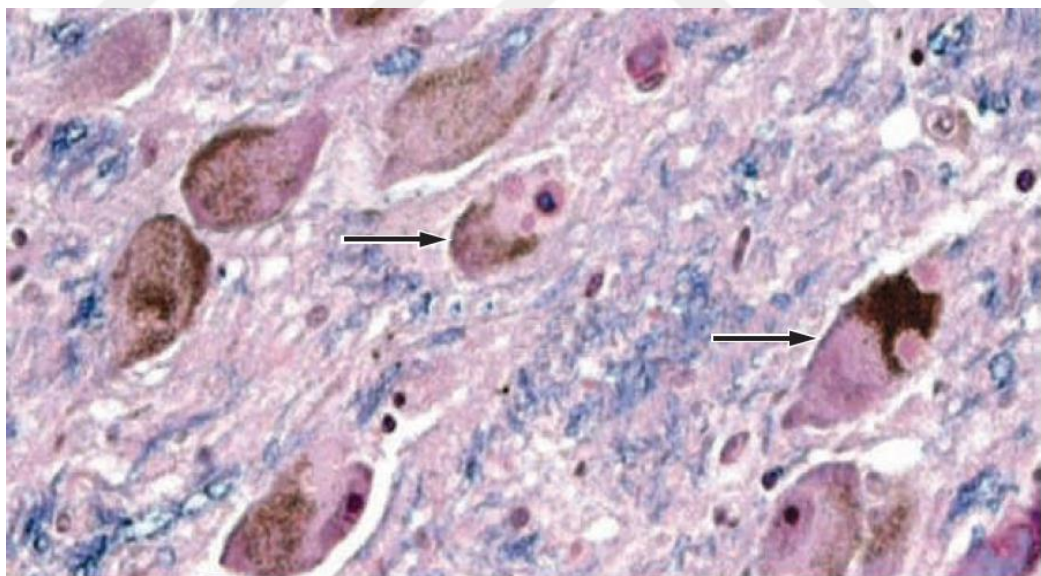


Figure 1.1 Multiple Lewy bodies (arrows) are seen in a section of the locus ceruleus from a patient with essential tremor. (Copyright 2005 Elsevier. From Louis et al.² with permission)

1.1.5 Risk and Etiological Factors for Essential Tremor

Heritability is defined as the degree to which individual genetic variation accounts for phenotypic variation seen in a population. Most heritability estimates have been based on family studies. Epidemiologic evidence shows associations between family history and risk for many diseases (Table). Thus, family history information might aid in assessing risk for a condition, even in the absence of an understanding of the molecular cause of that condition. A family history of ET is a risk factor, and the etiology of ET is often genetic²⁰. The estimates of the proportion of ET patients with a positive family history change between 20% and 90%²¹.

The frequency of ET increases greatly with age and can be as high as 5.5%²². According to community-based studies, the prevalence of ET is higher in whites compared to African-Americans²³. Nevertheless, there is limited information about the genetic factors responsible for ET. Hereditary ET usually exhibits an autosomal dominant pattern of inheritance; however, a non-Mendelian model of inheritance has also been reported²⁴. Twin studies further support the idea that non-genetic determinants strongly influence the incidence and severity of ET. In one study, pairwise concordance in monozygotic twins was 60%, while another study found this concordance to be 77-93%^{25,26}. β -carboline alkaloids and lead are the main environmental factors identified in case-control studies to potentially produce tremor in patients^{27,28}. β -carboline alkaloids in particular are the most common tremorogenic chemicals in the modern human diet, and Louis *et al.* showed that the blood concentrations of these chemicals were higher in ET patients than a healthy control group. These chemicals are abundant in meat products that are cooked at high temperatures under elongated times²⁹. In addition, blood concentration of lead is also found to be higher in ET patients than

controls, and lead was shown to contribute to cerebellar damage and tremor in ET-diagnosed individuals^{28,30}.

Table 1.2 Heritability estimation for selected diseases

Disease	Heritability Estimation
Age-related macular degeneration	49 - 71% ³¹
Alzheimer's disease	58 - 79% ³²
Autism	30 - 90% ³³
Bipolar disorder	70% ³⁴
Breast cancer	25 - 56% ³⁵
Celiac disease	57 - 87% ³⁶
Chronic obstructive pulmonary disease	76% ³⁷
Colon cancer	13% ³⁵
Coronary artery disease	49% ³⁸
Crohn's disease	53% ³⁹
Epilepsy	70 - 88% ⁴⁰
Essential Tremor	50-90% ²¹
Heart disease	34 - 53% ⁴¹
Hypertension	30% ⁴²
Leukemia	1% ³⁵
Lung cancer	8% ³⁵
Obesity	70% ⁴³
Osteoarthritis	30 - 65% ⁴⁴
Ovarian cancer	40% ³⁵
Parkinson's disease	25 - 30% ⁴⁵
Polycystic ovary syndrome	72% ⁴⁶
Prostate cancer	42% ³⁵
Rheumatoid arthritis	55% ⁴⁷
Schizophrenia	81% ⁴⁸
Stroke	32% ⁴⁹
Testicular cancer	25% ³⁵
Thyroid cancer	53% ³⁵
Tourette syndrome	58 - 77% ⁵⁰
Type-1 diabetes	88% ⁵¹
Type-2 diabetes	26% ⁵²

1.1.6 Genetic Heterogeneity in Essential Tremor

Despite the widespread occurrence of sporadic ET, twin and family studies have also demonstrated that genetics is an important factor for the disease. According to twin studies, heritability ranges from 45 to 90% for ET^{25,26}. Genome-wide linkage studies performed in North American and Icelandic ET families have identified genetic loci containing ET-associated genes on chromosomes 3q13 (*ETM1*; OMIM:190300)⁵³, 2p22-p25 (*ETM2*; OMIM: 602134)⁵⁴ and 6p23 (*ETM3*; OMIM: 611456)⁵⁵.

In 1997, a locus (*ETM1*, also called *FET1*) has been identified through the genome-wide scan of 16 Icelandic families, encompassing a total of 75 affected individuals. *FET1* was clearly inherited under a dominant mode of inheritance. The logarithm of odds (LOD) score was found to be 3.71, but the highest single family LOD score was calculated as 1.29, which is below the threshold of significance for mapping a monogenic disorder to a marker⁵³. In 2006, Jeanneteau *et al.* identified a Ser9Gly variant in the DRD3 gene at the *ETM1* locus in 23 of 30 French families⁵⁶. This gene codes for the dopamine D3 receptor, which is expressed in Purkinje cells and mediates ERK activation^{56,57}. However, it is not clear whether the Ser9Gly variant is cosegregated with the 16 Icelandic families linked to *ETM1*⁵³. In addition, the association between ET and the Ser9Gly variant in DRD3 was not observed in other populations^{58,59}.

In 1997, Higgins *et al.* used linkage analysis to identify the *ETM2* locus on chromosome 2p22-p25 in an American family of Czech descent. This gene segregated under autosomal dominant inheritance in this family, with a LOD score of 5.92⁶⁰. In 2003, this group also identified an ancestral haplotype on chromosome 2p24.2, which was found in 29% of 45 American individuals with a family history of ET⁶⁰. The HS1-binding protein 3 gene (*HS1BP3*, OMIM 609359) was also suggested as a gene of

interest for ET, and a relationship has been proposed between the disease and an A265G substitution affecting two genes and seven transcripts on a minimal critical region (MRC)^{61,62}. However, the relationship between ET and *HS1BP3* has not been confirmed by other studies^{55,63}. In addition, various association and linkage studies have failed to establish correlations between ET and mutations in the *ETM2* locus^{64,65}.

The *ETM3* locus is located on chromosome 6p23 and segregates with ET in an autosomal dominant pattern. The locus in question was identified through genome-wide non-parametric and parametric linkage analyses in seven North American families including 325 patients. However, no pathologic variants were found following the sequencing of 15 candidate genes in the *ETM3* locus⁵⁵. In addition, linkage to the *ETM3* locus was not present in two different linkage studies in Italian families^{65,66}.

Genome-wide association studies have led to the identification of two additional genes. The leucine rich repeat and Ig domain containing 1 (*LINGO1*, MIM 609791) gene was identified from families of American and European descent⁶⁷. However, *LINGO1* variants that segregate with the disease are found in non-coding intronic regions⁶⁸. Zhou *et al.* suggested that intronic variants, in conjunction with other environmental components, could disrupt the function of *LINGO1* and facilitate the onset of ET⁶⁹. The solute carrier family 1 (SLC1A2, MIM 600300) gene was also identified as a potential risk factor for ET in European populations through a genome-wide association study⁷⁰. However, meta-analysis demonstrated that these variants are not causative²¹.

FUS/TLS (*fused in sarcoma/translated in liposarcoma*, MIM137070) is the first putative ET gene identified through a combination of whole exome sequencing and linkage analysis. One pathogenic variant of *FUS* is a c.868C>T nonsense mutation that segregated with a large family of French-Canadian origin and was found to cause

autosomal dominant ET⁷¹. In addition, two rare missense variants (M392I, R377W) were also determined by further screening of ET cases, suggesting that FUS mutations account for a subset of ET cases^{72,73}.

Moreover, Rajput *et al.* established a connection between the genetic causes of ET and PD in the context of two VPS35 and DNAJC13 variants⁷⁴. 571 ET patients were genotyped for VPS35 c.1858G>A (p.D620N) and DNAJC13 c.2564A>G (p.N855S) variants and two individuals were observed to harbor the latter mutation, which was previously identified in PD patients⁷⁴.

Our group has also contributed to the identification of ET-associated genes through exome sequencing. Our recent report showed that a p.G399S missense mutation in HTRA2, which encodes a mitochondrial protease, is responsible for both essential tremor and Parkinson's disease. Whole exome sequencing of a six-generation consanguineous Turkish family illustrated that ET is present in both homozygous and heterozygous individuals for this allele, while homozygotes additionally developed Parkinson's disease later in life (Figure 1.3). As such, homozygous family members for HTRA2 p.G399S are more severely affected than heterozygotes, even though the presence of a single variant allele is sufficient to cause ET⁷⁵. HTRA2 produces a serine protease protein which is released from mitochondria to the cytosol and initiates apoptosis by binding to apoptotic inhibitors⁷⁶. According to mouse model studies, HTRA2 was also found to be related with Parkinson's disease. A Mnd2 mouse model study demonstrated that HtrA2 p.S275C causes loss of protease activity and leads to the degeneration of motor neurons, resulting in a phenotype with ataxia, repetitive movements, and akinesia⁷⁷. In another transgenic mice study, HtrA2 p.G399S was identified as a loss-of-function allele through the behavioral comparison of transgenic mice expressing either wild-type HtrA2 or the variant allele. Mice overexpressing the

WT variant showed motor impairments that were not present in HtrA2 p.G399S-overexpressing mice, suggesting that the overexpression of the variant allele can rescue its function⁷⁸.

Another ET-related exome sequencing study has discovered a pathogenic variant of the *TENM4* (MIM 610084) gene in a Spanish population. In particular, two novel missense mutations were shown to exhibit a dominant inheritance pattern in one family⁷⁹.

Clark *et al.* has published a recent study concerning the identification of candidate genes for familial early onset essential tremor. In this study, they identified four genes in five different families following the analysis of 37 early-onset families. In two different families, they determined a heterozygous variant (c.46G4A (p.Gly16Ser)) and a heterozygous variant (c.164C4T (p.Pro55Leu)) in the *NOS3* gene. The p.Asp379Glu variant in *KCNS2*, p.Gly350Arg variant in *HAPLN4* and p.Ala133Val variant in *USP46* genes were also shown to segregate with ET in three different families⁸⁰.

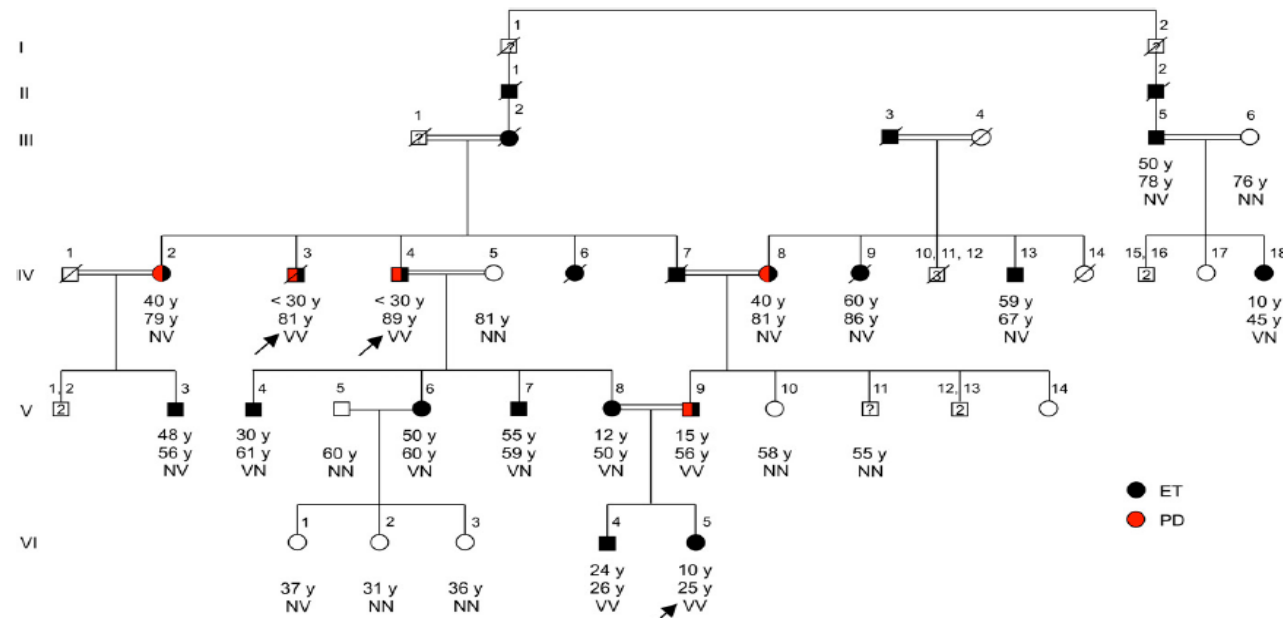


Figure 1.3 Segregation of HTRA2 p.G399S in a six-generation consanguineous Turkish family, illustrating that essential tremor is present in subjects both homozygous and heterozygous for this allele. Homozygous family members for HTRA2 p.G399S are more severely affected than heterozygotes. Individuals that were selected for exome sequencing (IV: 3, IV: 4, VI: 5) are indicated with arrows. Black symbol illustrates ET-diagnosed and red symbol illustrated PD-diagnosed individuals. N: non-variant, wild-type allele, glycine; V: the variant allele serine for HTRA2 p.G399S.

1.1.7 Treatment of Essential Tremor

Several treatments can alleviate the effects of ET for some patients; however, there is no conclusive cure for ET. Drugs can improve the quality of social life by reducing functional disability or discomposure⁸¹. Different β -blockers can be used for treatment, with propranol being the most common drug for this purpose⁸². Atenolol, sotalol and nadolol are other β -blockers that are used for the treatment of ET patients⁸². Primidone is the drug of choice for the treatment of tremor⁸². Primidone is more effective when taken in combination with propranol, and can reduce tremors in 75% of patients^{83,84}. If the patient suffers from both tremors and anxiety, benzodiazepines such as alprazolam and clonazepam are preferred due to their psychoactive effects⁸⁵. These drugs enhance the effect of the neurotransmitter gamma-aminobutyric acid (GABA) on GABA receptors, resulting in sedative, muscle relaxant and anti-anxiety effects⁸⁶. Patients who do not respond to drugs can instead be treated by surgical approaches such as deep brain stimulation⁸⁷.

Nevertheless, it should be emphasized that ET treatments would benefit greatly from the identification of new causative genes for essential tremor, which will enable the development of personalized treatment methods depending on the genetic history of each patient.

1.2 Disease Gene Identification Strategies

1.2.1 Traditional Methods

Genes associated with Mendelian diseases are typically identified through Sanger sequencing of candidate genes, which are selected based on their similarity with genes related to similar diseases, relevance between predicted protein function and the physiology of the disease, and positional mapping analyses⁸⁸. However, non-Mendelian diseases require higher scrutiny due to their variability, and a broad range of analysis techniques have been developed for their identification. The most important genetic mapping methods are karyotyping⁸⁹, linkage analysis⁹⁰, homozygosity mapping⁹¹, copy number variation analysis⁹² and SNP based association analysis⁹³.

Linkage analysis is the primary tool for the positional mapping of Mendelian genes⁹⁴. In 1910, Thomas Morgan used *Drosophila* cultures to determine that traits are co-inherited because of the physical proximity of the factors responsible for their appearance⁹⁵. Accordingly, genes that are clustered in close proximity on a chromosome tend to co-segregate because they are less likely to experience recombination than distant genes. Based on this theory, distances between trait-encoding genes can be determined through recombination studies⁹⁵. Consequently, if two traits are consistently co-segregated in a given family, it is likely that they are linked⁹⁶.

Logarithm-of-odds (LOD) scores can be used to determine the probability that a particular gene is truly associated with the disease of interest rather than co-segregating due to chance⁹⁷. LOD scores greater than 3 are considered to be suitably robust for suggesting linkage in Mendelian disorders⁹⁸.

Linkage mapping has been highly successful in mapping genes and gene variants affecting Mendelian traits. However, it is relatively inefficient in mapping loci underlying common polygenic disorders⁹⁹. Genome wide association studies (GWASs) are based upon the principle of linkage disequilibrium (LD) at the population level and focus on the nonrandom associations between alleles at different loci. These association studies have been utilized for the evaluation of many common genetic variants in different individuals, depending on the associations between single nucleotide polymorphisms (SNPs) and common diseases such as diabetes, heart diseases, autoimmune diseases and psychiatric disorders⁹⁹. Accordingly, if one allele is more frequent in individuals with a specific disease, that variant is evaluated to be *associated* with that disease⁹⁹. The first successful GWAS was published in 2005 and described the relationship between age-related macular degeneration and two SNPs that exhibited significantly changed allele frequencies in affected patients compared to healthy controls¹⁰⁰. However, GWAS is not a suitable technique for the identification of rare genetic diseases, as it usually identifies hundreds of variants that are common in the population (*i.e.* with frequencies over 5%), have to establish biological relevance to the disease of interest and explain only a small portion of the total inherited risk. In addition, only a small number of risk alleles identified by GWAS efforts are nonsynonymous SNPs in exons¹⁰¹.

Homozygosity mapping is another successful approach for the identification of recessive disease genes and is done by determining homozygous regions shared by affected family members¹⁰². This method depends on genome wide single nucleotide polymorphism (SNP) arrays of affected and unaffected family members. However, despite recent advances in sequencing technology, the genetic basis of several common

disorders remains unknown due to the limited number of family members, locus heterogeneity, variable expressivity and reduced penetrance.

1.2.2 Next Generation Sequencing

The first true sequencing method was described in 1977 by the hallmark paper of Sanger *et al.*, whose approach involved the use of dideoxy-nucleotides for controlled chain termination¹⁰³. This discovery contributed to the development of Sanger sequencing in 1987^{104,105} and has been the basis of DNA sequencing for nearly 30 years afterwards. Although technological developments have increased the length of DNA that can be sequenced (to c. 1000 bp) and facilitated the parallel sequencing of DNA strands in capillaries, Sanger sequencing was nevertheless limited by its inability to analyze DNA in a high-throughput manner.

This inability was one of the main impediments to the Human Genome Project (HGP), which was initiated in 1990 with the aim of identifying the 3 billion base pairs that comprise the human genetic code. Despite this rocky start, however, the project eventually finished ahead of schedule through the development of a new approach called *shotgun sequencing*¹⁰⁶. This method involves the production of small DNA fragments through the mechanical or enzymatic breakdown of a DNA isolate, which are subsequently cloned into sequencing vectors¹⁰⁷. Cloned DNA sequences can then be sequenced individually, and the entire genome sequence can be determined by the alignment and assembly of fragmented sequences using overlapping regions¹⁰⁷. The 10-year project culminated in the publication of first drafts by Lander *et al.*¹⁰⁸ and Venter *et al.* in 2001¹⁰⁶, followed by a complete publication by Jasny and Roberts *et al.* after three more years¹⁰⁹. The success of HGP also led to the initiation of certain important projects, such as the International HapMap Project and 1000 Genomes Project. Of these, the International Hapmap project aims to produce a haplotype map of the human

genome and determine the patterns of common genetic diversity in the genetic code¹¹⁰. This approach attracted great attention to the prospect of identifying the genetic causes behind common human disorders, and was completed in 2005 with the genotyping of 4.6 million SNPs¹¹⁰. The 1000 Genomes project aims to uncover more detailed genetic factors involved in human health and disease¹¹¹. This project also provides various resources on human genetic variation, including the genomes of at least one thousand volunteers from a broad range of ethnic groups¹¹¹.

Shotgun sequencing also spurred the development of massive parallel sequencing techniques that would be used in next generation sequencing (NGS) efforts^{112–114}. NGS is performed by breaking the entire genome into smaller chunks and ligating them to pre-designated adapters for rapid, irregular reading during DNA synthesis. Consequently, the technique is also known as sequencing by synthesis¹¹⁵. As NGS produces millions of short, fragmented DNA sequences that only partially overlap, the problem in sequencing moved from the logistics and chemistry of the sequencing reaction to data management and analysis of the resulting fragments. As such, new algorithmic approaches have been developed (and are actively in development) to create a coherent genetic sequence from these disjointed fragments¹¹⁶.

1.2.3 Whole Exome Sequencing Approach

Whole exome sequencing (WES) is a practical application of NGS and limits itself to the variations found in the coding regions (exons) of all currently known genes. Newly developed technologies such as whole genome and whole exome sequencing have replaced traditional methods because of their decreased costs, higher accuracy, capacity for rapid diagnosis and compatibility with a number of user-friendly software for data analysis^{117,118}.

The human genome contains around 3×10^9 bases, but only about 3×10^7 base pairs (*i.e.* 1% of the genome) are responsible for encoding proteins¹¹⁹. The WES approach involves the sequencing of this comparatively small amount of DNA, using hybridization techniques to separate it from the non-coding “chaff”^{120,121}. Two main approaches exist for the “capture” of exomes: solution-based and array-based applications.

In solution-based WES, DNA samples are broken down and hybridized to target regions in the genome by biotinylated oligonucleotide probes. Magnetic streptavidin beads can then be used to bind to biotinylated probes, thereby retaining the coding regions while the rest of the DNA is washed off. The remaining sample is then amplified through polymerase chain reaction (PCR), resulting in the selective enhancement of the coding regions (Figure 1.4)¹²². Array-based WES is similar to its solution-based counterpart, except that the probes are bound to a high-density microarray¹²³. It also needs a lower amount of DNA and is more efficient than solution-mediated approaches^{124,125}.

NGS has greatly advanced our capacity to determine the genetic factors underlying various disorders. The first report about WES was published by Ng *et al.* in 2009¹²¹. In this report, four individuals that were diagnosed with Freeman-Sheldon syndrome were sequenced by whole exome sequencing, and following the exclusion of common variants, the *MYH3* gene was discovered as a previously unknown gene for a rare Mendelian disorder¹²¹. The first application of NGS for molecular diagnosis of a patient came from Choi *et al.*, who reported the clinical diagnosis of a suspected Bartter syndrome individual of Turkish origin¹²⁶. They demonstrated that a well-conserved recessive mutation in *SLC26A3* gene is associated with congenital chloride diarrhea and utilized WES as a clinical tool that is able to facilitate the diagnostic evaluation of patients in addition to disease gene discovery¹²⁶. Their study therefore allowed the

correct diagnosis of individuals who were suffering from congenital chloride diarrhea (CLD) but were misdiagnosed with Bartter syndrome¹²⁶.

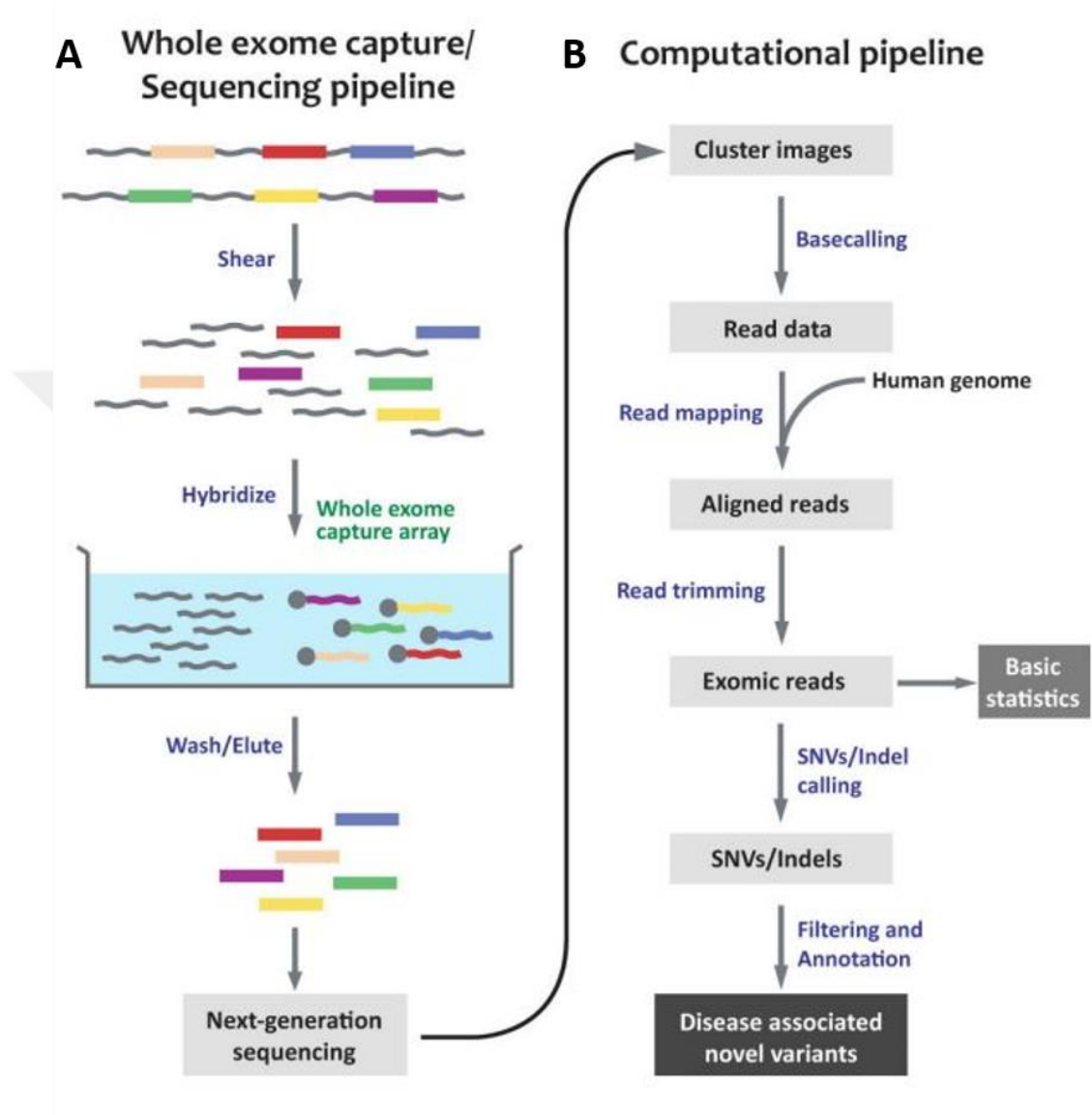


Figure 1.4 Overview of an exome sequencing pipeline. SNV: single nucleotide variant, Indel: Insertion/Deletion. (A) In sequencing pipeline, isolated genomic DNA is sheared into 200-300 bp fragments. Exomes are hybridized with biotinylated probes and non-targeted regions are washed and eluted. At the end, libraries are sequenced. (B) Flowchart of a computational pipeline for WES, including data analysis. (Copyright 2012 by The Korea Genome Organization. From Choi et al ¹²⁷ with permission)

1.2.4 Identification of Disease-Causing Genes

Family based linkage analysis is usually the first step for disease gene identification, although it may be insufficient for the determination of diseases caused by *de novo* mutations¹²⁸. Family based linkage studies have become common place with development of microarrays for GWAS and NGS¹²⁹.

WES has been performed for the determination of disease causing variants in various rare Mendelian diseases, including spastic paraplegia, brain malformations, cerebellar ataxia, retinitis pigmentosa and nonsyndromic mental retardation.^{130–134} (Table 1.2). WES has also been performed for complex and common disorders such as Alzheimer's disease, essential tremor, multiple sclerosis and the familial form of Parkinson's disease to investigate rare genetic variants related with these diseases^{71,135–137}.

Determination of disease causing genes starts with the alignment of WES data to the human reference genome, which is obtained from the HGP. A minimum of 30 reads per base is generally considered to be sufficient for accurate sequencing. Certain databases (dbSNP, ExAC, 1000genomes, EVS) are then used to determine whether the identified variants have previously been reported and exclude common variants. After this step, the remaining variants have their functions annotated, and the neutral-coding variants are eliminated based on functional filtering. This step ensures that all identified variants contain protein-damaging mutations such as nonsense mutations and frameshift insertions or deletions, which generate non-functional or truncated proteins. Several *in silico* analysis methods (such as SIFT, PROVEAN, PolyPhen2, MutationAssessor and MutationTaster) are then used to evaluate the deleteriousness of missense mutations and determine how damaging each variant will be to the complete protein^{138–142}.

The exome sequencing approach has been successfully used for the identification of disease-causing variants. However, this method still suffers from a number of restrictions. First, our knowledge of protein-coding regions is still incomplete; consequently, it is probable that some “non-coding” regions in fact contain genes. Second, some sequences fail to be targeted by the capture probe design, resulting in “leaks” in the sequenced exomes. Third, not all sequences can be mapped to the reference genome. And lastly, it is becoming increasingly evident that other sequences such as microRNAs, promoters and conserved elements also play important roles in the pathogenesis of various disorders, but none of these are considered in exome sequencing⁹⁴.

Overall, however, exome sequencing is a rapid, powerful and cost effective technique for the identification of new genetic variants for Mendelian and non-Mendelian diseases.

Table 1.3 Mendelian disease gene identifications by exome or genome sequencing(Copyright 2011 BioMed Central Ltd. From Gilissen et al¹⁴³ with permission)

Disorder	Inheritance	Gene identified	Scope	References
Congenital chloride diarrhea	Recessive	SLC26A3	Exome	Choi et al. ¹²⁶
Miller syndrome	Recessive	DHODH	Exome	Ng et al. ¹⁴⁴
Charcot-Marie-Tooth neuropathy	Recessive	SH3TC2	Genome	Lupski et al. ¹⁴⁵
Metachondromatosis	Dominant	PTPN11	Genome	Sobreira et al. ¹⁴⁶
Schinzel-Giedion syndrome	Dominant	SETBP1	Exome	Hoischen et al. ¹⁴⁷
Nonsyndromic hearing loss	Recessive	GPSM2	Exome	Walsh et al. ¹⁴⁸
Perrault syndrome	Recessive	HSD17B4	Exome	Pierce et al. ¹⁴⁹
Hyperphosphatasia mental retardation syndrome	Recessive	PIGV	Exome	Krawitz et al. ¹⁵⁰
Sensenbrenner syndrome	Recessive	WDR35	Exome	Gilissen et al. ¹⁵¹
Cerebral cortical malformations	Recessive	WDR62	Exome	Bilguvar et al. ¹³⁰
Kaposi sarcoma	Recessive	STIM1	Exome	Byun et al. ¹⁵²
Spinocerebellar ataxia	Dominant	TGM6	Exome	Wang et al. ¹³¹
Combined hypolipidemia	Recessive	ANGPTL3	Exome	Musunuru et al. ¹⁵³
Complex I deficiency	Recessive	ACAD9	Exome	Haack et al. ¹⁵⁴
Autoimmune lymphoproliferative syndrome	Recessive	FADD	Exome	Bolze et al. ¹⁵⁵
Amyotrophic lateral sclerosis	Dominant	VCP	Exome	Johnson et al. ¹⁵⁶
Nonsyndromic mental retardation	Dominant	Various	Exome	Vissers et al. ¹⁵⁷
Kabuki syndrome	Dominant	MLL2	Exome	Ng et al. ¹⁵⁸
Inflammatory bowel disease	Dominant	XIAP	Exome	Worthey et al. ¹⁵⁹
Nonsyndromic mental retardation	Recessive	TECR	Exome	Caliskan et al. ¹⁶⁰
Retinitis pigmentosa	Recessive	DHDDS	Exome	Züchner et al. ¹³³
Osteogenesis imperfecta	Recessive	SERPINF1	Exome	Becker et al. ¹⁶¹
Dilated cardiomyopathy	Dominant	BAG3	Exome	Norton et al. ¹⁶²
Hajdu-Cheney syndrome	Dominant	NOTCH2	Exome	Simpson et al. ¹⁶³
Hajdu-Cheney syndrome	Dominant	NOTCH2	Exome	Isidor et al. ¹⁶⁴
Skeletal dysplasia	Recessive	POP1	Exome	Glazov et al. ¹⁶⁵
Amelogenesis	Recessive	FAM20A	Exome	O'Sullivan et al. ¹⁶⁶
Chondrodysplasia and abnormal joint development	Recessive	IMPAD1	Exome	Vissers et al. ¹⁶⁷
Progeroid syndrome	Recessive	BANF1	Exome	Puente et al. ¹⁶⁸
Infantile mitochondrial cardiomyopathy	Recessive	AARS2	Exome	Götz et al. ¹⁶⁹
Sensory neuropathy with dementia and hearing loss	Dominant	DNMT1	Exome	Klein et al. ¹⁷⁰
Autism	Dominant	Various	Exome	O'Roak et al. ¹⁷¹

1.3 Consanguinity

Consanguinity has been identified as a risk factor for rare diseases since the turn of the century. Consanguineous individuals are those who descend from strongly related (*i.e.* second cousins or closer) ancestors, and share an extraordinarily high amount of genetic material. The first report detailing the detrimental effects of consanguineous marriages was published in 1902 and concerns alkaptonuria patients¹⁷². This phenomenon can also be seen in Mendelian inheritance. The co-occurrence of two recessive alleles is much more frequent in these families: for example, assuming that each couple shares 1/8 of their genome, their progeny will be homozygous for 1/16 of all loci.

The frequency of recessive diseases strongly correlates with the rate of consanguineous marriages. More than 50% of marriages are consanguineous in certain regions of North Africa, South India, West Asia and eastern and southern rims of the Mediterranean, resulting in strong incidence of genetic disorders in these populations¹⁷³ (Figure 1.5). As such, the likelihood of recessive disorders creates potentially detrimental health effects in consanguineous marriages, and the disease genes in question can be identified by determining the regions that are shared between family members¹⁷⁴. Accordingly, a prospective disease-associated variant should be found in other affected family members, but also be absent from unaffected family members. These regions can be determined by genome-wide SNP genotyping arrays by analyzing shared homozygous regions¹⁷⁵. In addition to autosomal recessive disorders, consanguinity is also important for diseases exhibiting a dominant inheritance pattern because the presence of a second allele often alters the severity of the disease⁷⁵.

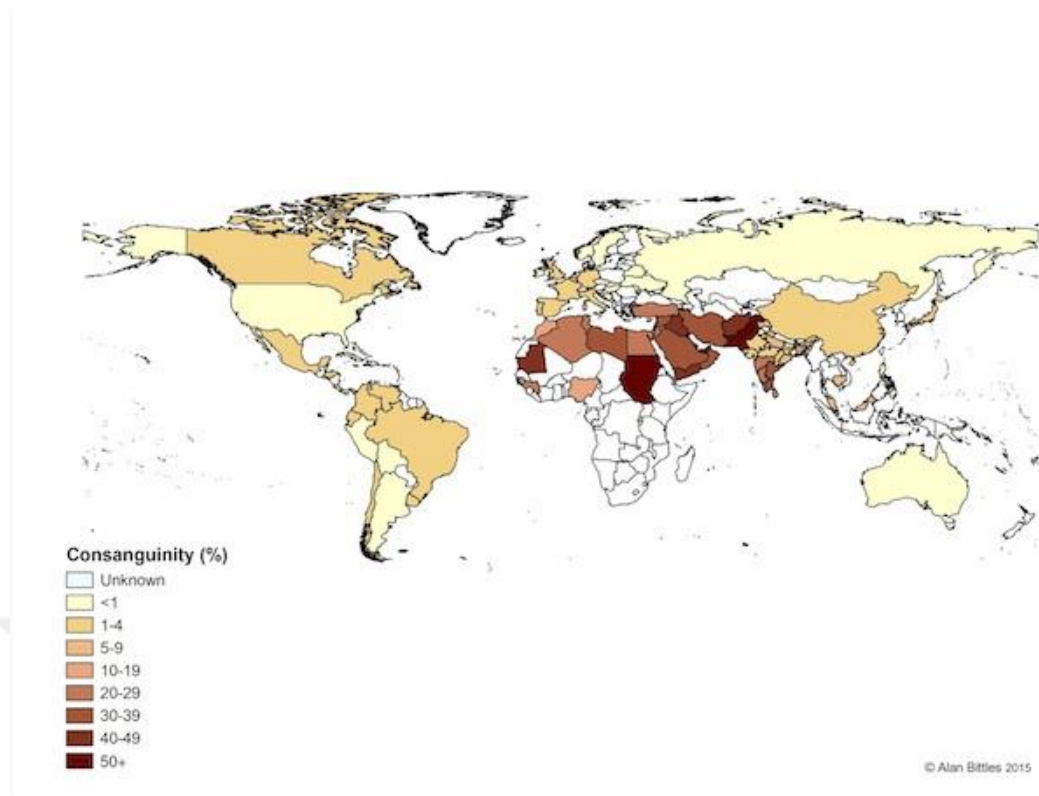


Figure 1.5 Global prevalence of consanguinity. (<http://www.consang.net> with permission from A.H. Bittles)

1.4 Outline of the Thesis

In this dissertation, I detail the identification of the candidate genes related with familial essential tremor under five chapters, including this introductory section.

The first chapter provides a general outlook on the identification of genetic disorders in general and ET in particular. The second chapter is a methods section that explains how clinical assessment of families, whole exome sequencing analysis and disease gene identification were performed.

The third chapter covers the identification of two genes related with ET. ET is one of the most frequent movement disorders in humans; however, its genetic causes and pathophysiology are unknown. We identified candidate genes in three ET families based on WES. In two independent Turkish ET families, we identified a rare and potentially damaging variant in the matrix metalloproteinase gene (*MMP19*) as a causative allele. In a third Turkish family, we determined a rare and damaging missense mutation in *IMPG1*. Our results suggest that *MMP19* and *IMPG1* might be responsible for essential tremor in these families.

The fourth chapter contains a discussion of my findings, and the final chapter summarizes my study and offers future perspectives on determining the genetic causes of ET.

CHAPTER 2

Material and Methods

2.1 Recruitment of Family Members

Three families; ET-5, ET-31 and ET-49, were investigated in the present study. ET-5 is a family from Bayır Village (Muğla, Turkey) and has four individuals (spanning four generations) suffering from essential tremor. ET-49 is a consanguineous family from Ilgın (Konya, Turkey), includes 6 patients and a control group, and has a family history of ET for at least five generations. ET-31 family is a consanguineous family from Corum, Turkey includes 11 patients and 4 control individuals, and has a family history of ET for at least four generations. All individuals are coded by the blood ID. The study was approved by the Institutional Review Board at Bilkent University. Additional cohorts involving patients having similar neurological phenotypes were included in the study to identify other patients with candidate mutation. Before the study, all the participants were asked to sign an enlightened consent form that was prepared according to the guidelines of the Ministry of Health in Turkey.

2.2 Clinical Investigations

Clinical investigations were performed at Ankara University Medical School and Hacettepe University Medical School. All clinical examinations were performed according to the Helsinki Declaration (<http://www.wma.net>). Each participant was examined for essential tremor consistent with criteria of both the Washington Heights-Inwood Genetic Study of Essential Tremor (WHIGET) and the Consensus Statement of the Movement Disorder Society on Tremor (MDS). After the diagnosis of essential tremor, a pedigree was developed and histories about the distribution of the disease was

collected through communication with family members. Each participant's tremor score was graded for resting and postural tremors according to 0 to +3 tremor ratings. As stated in WHIGET criteria, 0 means no visible tremor, +1 is of low amplitude; +2 tremor is of moderate amplitude and +3 tremor is of large amplitude. Participants were asked to carry through four distinct tasks to evaluate kinetic tremor: finger-to-nose movement, pouring water, drinking water from a cup and drawing. Tremor severity was estimated in the course of each task. To separate essential tremor symptoms from Parkinson Disease, bradykinesia, muscular rigidity and postural instability were also evaluated according to the diagnostic criteria of the UK Parkinson Disease Society Brain Bank.

2.3 DNA Isolation from Family Members

Peripheral blood samples from patients and their relatives were collected in K3-EDTA-containing BD Vacutainer® Blood Collection tubes (Becton Drive, NJ, USA) by a specialist utilizing veinpuncture technique. Blood samples were carried to the laboratory under cold chain conditions and stored in 1.5 ml micro centrifuge tubes containing 1 mL blood samples at -80 °C.

Each individual's genomic DNA was isolated from 200 µL of peripheral whole blood samples using Nucleospin® Blood kit (Macherey-Nagel Inc., PA, USA) in accordance with the manufacturer's instructions. To get high DNA concentrations that are required for whole exome sequencing, washing steps were repeated two times and DNA was eluted with double distilled water rather than elution buffer.

Quality of genomic DNA samples was measured by densitometry analysis using horizontal 1% agarose gel electrophoresis with TAE Buffer. To analyze the quantity and purity of DNA samples, spectrophotometric analysis was performed with

NanoDrop™ ND-1000 Spectrophotometer (NanoDrop Technologies, Inc, DE, USA). DNA concentration, A_{260/280} and A_{260/230} values were recorded. A_{260/280} ratio is used for determining the purity of DNA and a ratio of ~1.8 is generally accepted as “pure”. The 260/230 ratio was used for assessing the purity of nucleic acids and ideal values are generally in the range of 2.0-2.2.

2.4 Library construction and whole exome sequencing

2.4.1 Sample Selection and Preparation

Two affected individuals (II-1 and IV-3) and one control (III-2) were selected from the ET-5 family for WES. Likewise, four affected (III-2, IV-2, IV-4, V-2 and V-3) and one unaffected (IV-1) individuals were selected from the ET-49 family, and three affected individuals (III-2, III-3 and III-4) and one control (II-13) were chosen from the ET-31 family for exome analysis. High quality DNA were isolated from selected individuals as described.

2.4.2 Library Construction

Library construction was performed for whole exome sequencing using a commercial kit (Illumina TruSeq DNA Sample Prep Kit), based on the manufacturer’s instructions. Accordingly, DNA samples were diluted in suspension buffer to get a final amount 2.5 µg in 55 µL and sheared into 200-300 bp for 120 s at 6 °C in a sonicator to produce dsDNA fragments with 3’ or 5’ overhangs. For checking the quality of DNA, sheared DNA was run on agarose gel. The final amount of sheared DNA was 1 µg. The overhangs were then converted to blunt ends utilizing End Repair Mix for 30 min at 30 °C. After cleaning DNA fragments with 160 µL of AMPure XP Beads, this mixture was placed on a magnetic stand. The DNA samples with beads were washed with 80% ethanol and dried, and DNA samples were resuspended in Resuspension Buffer. The

DNA fragments were incubated with A-tailing mix for 30 min at 30 °C, allowing single 'A' nucleotides to be added to 3' ends of DNA fragments and preventing them from ligating to one another. The ligation reaction was performed by adding multiple paired-end adapters with 'T' nucleotides on their 3' fragments, which were ligated to complementary dsDNA fragments following incubation for 10 min at 30 °C. This reaction was inactivated by Stop Ligation Buffer and the fragments were cleaned up with AMPure XP Magnetic Beads and amplified by PCR. Exomes were captured by SeqCap EZ Exome Capture Kit (Roche) and hybridized with biotinylated capture probes at 47 °C for 72 h. The exomes were cleaned with Qiagen QIAquick PCR Purification Kit and their concentration and quality were determined with ABI system KAPA Illumina Library Quantification Kit with Real Time PCR. Libraries were sequenced on an Illumina HiSeq2500 and converted to basecall (.bcl) file format.

2.5 Bioinformatics

2.5.1 Whole Exome Sequencing Data Analysis

Qualities of sequences were determined using Illumina Real Time Analysis Software and sequences were converted to .bcl file format. The data included in .bcl files were converted to FASTQ files through Illumina CASAVA software. Burrows-Wheeler Aligner¹⁷⁶ (BWA, v0.6.1-r104) was used to align the fragments to the human reference genome¹⁷⁷ (hg19). PCR duplications were removed by Sequence Alignment/Map Tools (SAMtools) software package¹⁷⁸ and sequence depth of exonic regions were obtained by BED tools¹⁷⁹. SNP and Indel variants and base quality score recalibration were done with Genome Analysis Tool Kit¹⁸⁰ (GATK; v3.0-0-g6bad1c6). Variants were positionally and functionally annotated with SnpEff¹⁸¹ (4.2, 2015-12-05) and ANNOVAR¹⁸² (2016Feb01) software packages. In order to further verify the accuracy

of the determined variants, regions were analyzed using the Integrative Genomics Viewer (IGV) platform, which allows interactive assessment of genomic datasets¹⁸³.

2.5.1 Filtration and Prioritization of Candidate Gene Variants

The following criteria was used to obtain a list of variants for ET-affected families: (1) including only protein-coding region variants, (2) excluding all common variants that have MAF bigger than 0.02 in ExAC (**Exome Aggregation Consortium**), (3) including variants shared between affected individuals within a family. After variants were annotated positionally and functionally with annotation tools, only protein coding region variants were selected based on their annotation impacts, which were determined with SnpEff. ‘HIGH’ and ‘MODERATE’ variants were selected for further analysis. The ‘HIGH’ variants included frameshift, splice acceptor, splice donor, stop loss variants and ‘MODERATE’ variants are composed of inframe insertion/deletion, missense variants and splice region variants. Common SNPs and variants that are frequently seen in the population (minor allele frequency (MAF) is bigger than 0.02) were filtered using ExAC, the 1000 Genomes Project, the NHLBI Exome Sequencing Project and our in-house database and selected for further analysis. Variants were removed without protein altering effect(synonymous, start-retained and stop-retained) and remained with nonsynonymous effect in coding region (i.e. missense, nonsense) or alteration resulting in frameshifts, premature stop codons, loss of stop codons, coding INDELS, splice site and splice region variants. We filtered a variant if it is also found as homozygous or compound heterozygous in the genome of his control sample. For patient-patient pairs, we selected the variants that are shared by affected individuals within the same family. We also considered hemizygous sequence variants on the X chromosome found in brother–brother pair families as candidate mutations. We filtered heterozygous variants only if the control sibling has the same variation and/or any of

the control individual has the same variant in homozygous state. We prioritized as potentially damaging if they generated a premature stop codon or were mutations with the following scores on *in silico* prediction tools: Provean <-2.5 , SIFT ≤ 0.05 , Polyphen2 HDIV; D: Probably damaging ≥ 0.957 , P: possibly damaging $0.453 \leq \text{pp2hdiv} \leq 0.956$, Polyphen2 Hvar D: Probably damaging ≥ 0.909 , P: possibly damaging $0.447 \leq \text{pp2hvar} \leq 0.909$, MutationTaster prediction for A (disease causing automatic) and D (disease causing), MutationAssessor >1.94 (M: Moderate and H: High), CADD ≥ 10 , GERP ≥ 3 and PhastCons_Mammalian ≥ 0.90 .

2.6 Sanger Sequencing

Segregation analysis of the candidate variants for other informative relatives of ET-5, ET-49 and ET-31 families were performed in order to determine if the candidate genes were co-segregated with autosomal ET. The appropriate primers (Table A1) were designed with the Primer3Plus bioinformatics tool. Segregation analysis was performed with Chromas Lite analysis software (Technelysium Pty Ltd) and Multiple Sequence alignment was done with Clustal Omega.

2.7 Model 3D Protein Structure of Candidate Variants

In order to deduce structural alignments and compare the active sites and other relevant regions of the candidate variants, 3D structures of related protein models were derived from Swiss Model, Raptor X or Phyre2 and analyzed with Swiss-PdbViewer (DeepView, v4.1)¹⁸⁴.

CHAPTER 3

Results

3.1 Matrix Metalloproteinase 19 (MMP19), p.Arg456Gln Missense Mutation Causes Essential Tremor

3.1.1 Clinical Features of the Families

The ET-5 family is a four-generation Turkish family with essential tremor and the disease was observed in all generations. All individuals were asked to sign an enlightened consent form that was prepared according to the guidelines of the Ministry of Health in Turkey. Affected and unaffected individuals were each examined by neurologists twice. For WES, IV-3 and II-1 were selected as affected individuals and III-2 was selected as control (Figure 3.1). The proband of ET-5 family, IV-3 (13 year-old, age of onset 5), had ET since she was 5 years old. She exhibits right hand resting tremor with moderate-amplitude, left hand resting tremor with low-amplitude and postural, kinetic and intentional tremors. Her clinical severity is milder than other affected individuals in the ET-5 family. She has no additional clinical findings or diseases. Other affected individuals from the ET-5 family, II-1 (48, 19) and III-3 (71, 20), exhibit moderate clinical severity including resting, postural and intentional tremors with moderate-amplitude. Moreover, III-3 has hyperthyroidism and low-amplitude kinetic tremor and II-1 has asthma and osteoporosis and moderate-amplitude kinetic tremor. The affected individuals IV-3, III-3 and II-1 do not have any hypomimia, head, chin or voice tremor, postural instability or freezing. Also, their Archimedes spiral drawing results show low-amplitude scores. Control individuals from family ET-5 were III-2 (48 y), III-4 (49 y) and IV-4 (17 y), none of whom have

resting, postural, kinetic or intentional tremor. Other ET tests have also resulted in negative scores for these individuals (Table 3.1, Figure 3.2).

The consanguineous ET-49 family is also of Turkish origin and ET is observed in every generation. For WES, IV-2 (64, 8), IV- 4 (51, 20), V-2 (34, 8) and (V-3) were selected for the affected group and IV-1 was chosen as a control (Figure 3.1). All individuals signed a consent form that was prepared according to the guidelines of the Ministry of Health in Turkey. Affected and unaffected individuals were examined twice by neurologists for ET. The affected individual IV-2 exhibited severe clinical severity involving moderate-amplitude resting right hand, postural and kinetic tremors and low amplitude intentional tremor, and his age of onset was 8 years old. This individual also has hypomimia, head and voice tremor. Moreover, he has right wrist bradykinesia and right/left rigidity. He has no symptoms related with freezing and postural instability. His condition is more severe than other affected individuals from the ET-49 family, which is also seen in his Archimedes spiral test (which has the highest score among the individuals tested; Table 3.1, Figure 3.2). V-2 has moderate-amplitude right postural and low-amplitude kinetic, intentional, right hand resting and left hand postural tremors. He exhibits no symptoms related with hypomimia, rigidity, postural instability and freezing. However, he has right-hand bradykinesia. V-3 has moderate postural and kinetic tremors, low amplitude resting right hand tremor and her age of onset is before she was 17-year-old. She also has voice tremor. She exhibits no symptoms related with hypomimia, intentional tremor, head, chin tremors and other Parkinson's disease indications. Another affected individual, IV-4, has moderate right hand postural and right hand kinetic tremor and his age of onset for ET is before the age of 20. He also has head, voice tremor and right hand rigidity. He has no left hand resting tremor, hypomimia, bradykinesia, postural instability and freezing. The proband of ET-49

family, III-2, has moderate clinical severity with moderate resting, postural, kinetic and intentional tremors. He is 86 years old, with an age of onset of 8 years old. The course of his disease is progressive and he has no hypomimia, rigidity, postural instability and freezing. He has moderate-score bradykinesia for both hands. Another affected individual, V-1, only shows left-hand kinetic and left-hand intentional tremor. Other symptoms related with ET and PD are not observed. Archimedes spiral test scores illustrate that IV-2 has higher scores than others. IV-1 is an unaffected individual and has no symptoms related with ET or PD (Table 3.1, Figure 3.3).



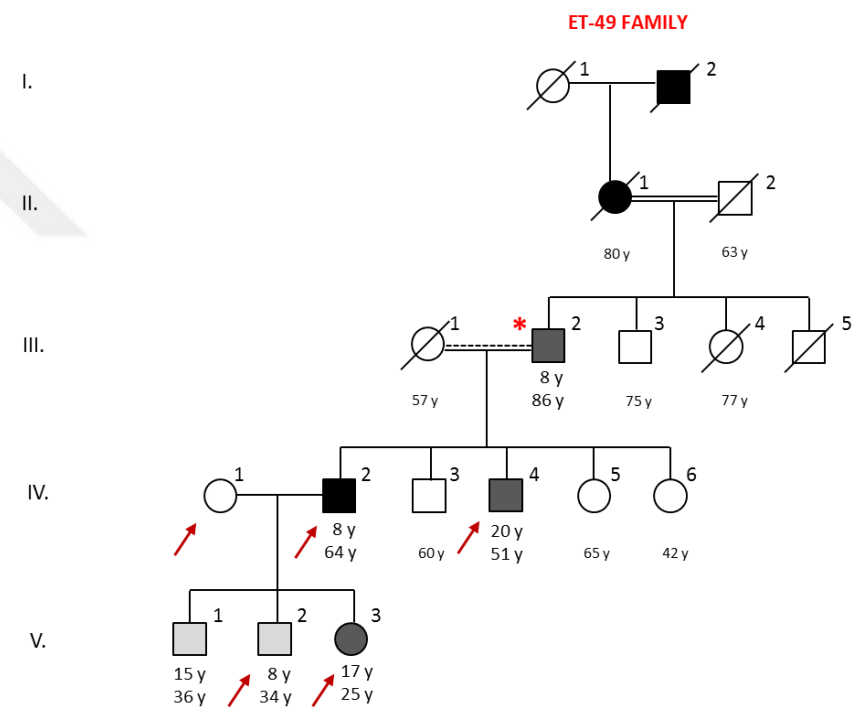
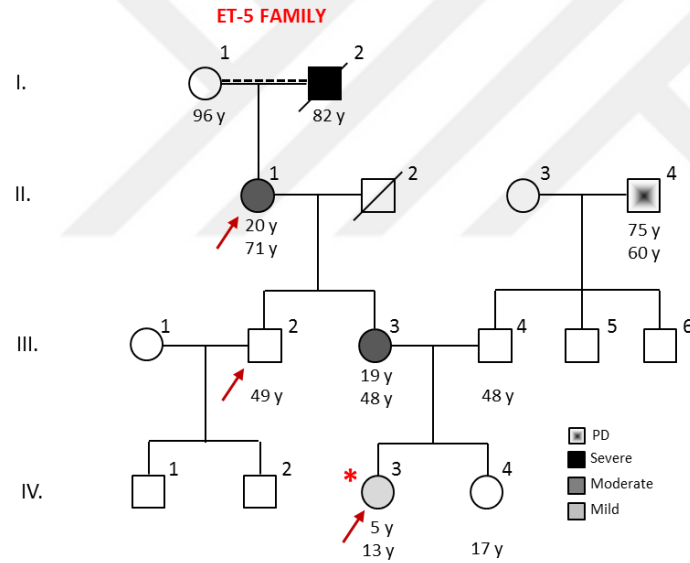


Figure 3.1 Pedigrees of ET-5 and ET-49 families showing the affected and unaffected individuals with essential tremor. Red arrows indicate the individuals which were selected for whole exome sequencing (II-1, III-2 and IV-3 from ET-5 family, IV-1, IV-2, IV-3, V-2 and V-3 from ET-49 family). Probands are illustrated with asterisk (*). Individuals with ET are illustrated with filled dark gray if the clinical severity is moderate, light gray for mild clinical severity and black for severe ET. Individual with PD is illustrated with gray color with black point.

Table 3.1 Clinical characteristics of tremor in affected and unaffected individuals of families ET-5 and ET-49

Family ID	Sample ID	Sex	Clinical Severity	Age	Age of Onset	Additional Clinical Findings	Hypomimia	Tremor										Bradykinesia		Rigidity		Postural Instability	Freezing	Dominant Hand	Archimedes		
								Resting		Postural		Kinetic		Intentional		Head	Chin	Voice	R	L	R				L	R	L
								R	L	R	L	R	L	R	L												
ET-5 FAMILY	IV-3	F	mild	13	5		-	++	+	+	+	+	+	+	+	No	No	No	-	-	-	-	-	No	Left	+	+
	IV-4	F	-	17	-	-	-	-	-	-	-	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	-	-
	III-4	M	-	49	-	-	-	-	-	-	-	-	-	+	No	No	No	-	-	-	-	-	No	Right	-	-	
	III-3	F	moderate	48	19	Hyperthyroidism	-	++	++	++	++	+	+	++	++	No	No	No	-	-	-	-	-	No	Right	+	+
	II-1	F	moderate	71	20	Asthma, osteoporosis	-	++	++	++	++	++	++	++	++	No	No	No	-	-	-	-	-	No	Right	+	+
	III-2	M	-	48	-	-	-	-	-	-	+	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	-	-
ET-49 FAMILY	IV-2	M	severe	64	8	-	+	++	-	++	++	++	++	+	+	No	Yes	Yes	+	-	+	+	-	No	Right	++	+++
	IV-1	F	-	56	-	migraine	-	-	-	-	-	-	-	-	-	No	No	No	-	-	-	-	+	No	Right	-	-
	V-2	M	mild	34	8	-	-	+	-	++	+	+	+	+	+	No	No	No	+	-	-	-	-	No	Right	-	-
	V-3	F	moderate	25	<17	-	-	+	-	++	++	++	++	-	-	No	No	yes	-	-	-	-	-	No	Right	+	+
	IV-4	M	moderate	51	<20	-	-	+	-	++	+	++	+	+	+	Yes	No	Yes	-	-	+	-	-	No	Right	+	+
	III-2	M	moderate	86	8	Lumbar hernia	-	++	++	++	++	++	++	++	++	No	No	No	+	+	-	-	-	No	Right	+	+
	V-1	M	mild	36	15	-	-	-	-	-	-	-	+	-	+	No	No	No	-	-	-	-	-	No	Right	+	+

Abbreviations; For tremor ‘-’ means no visible tremor, +1 is low amplitude; +2, tremor is moderate amplitude; +3, tremor is of high amplitude. For hypomimia, bradykinesia, rigidity, postural instability and Archimedes spiral test +: mild, ++: moderate, +++: severe. R: Right, L: Left.

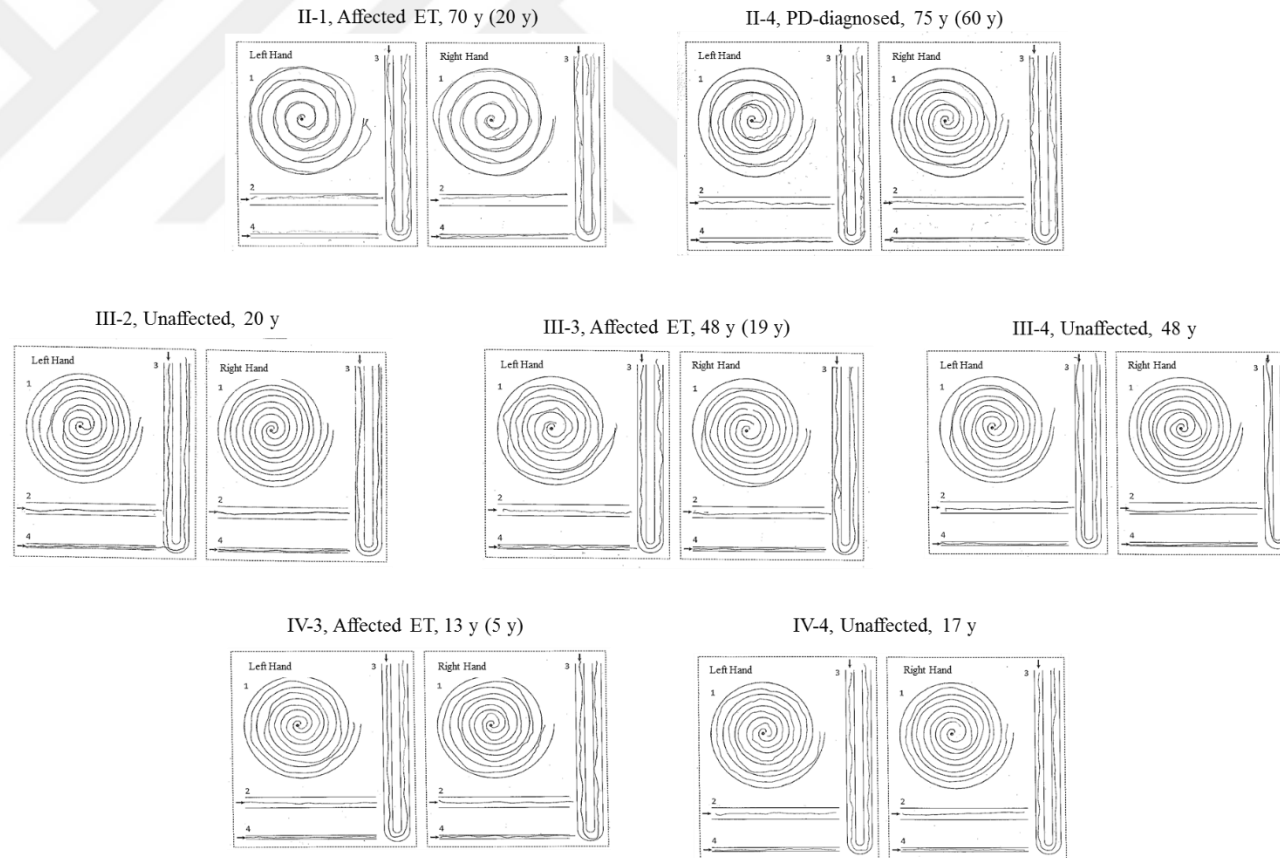


Figure 3.2 Archimedes spiral drawings of ET-5 family members including affected and unaffected individuals. Current ages and ages at onset of the individuals are illustrated.

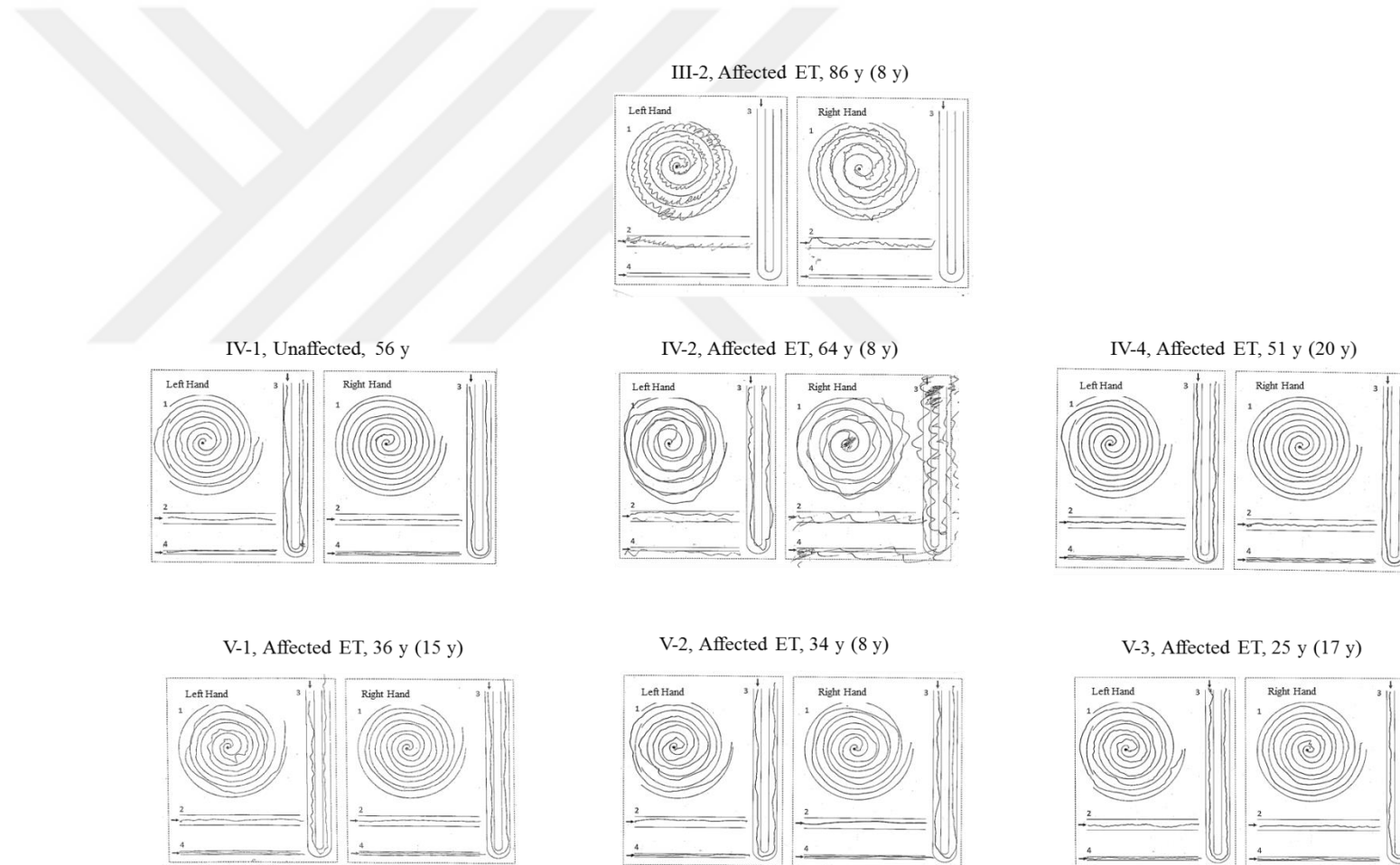


Figure 3.3 Archimedes spiral drawings of ET-49 family members including affected and unaffected individuals. Current ages and ages at onset of the individuals are illustrated.

3.1.2 Analysis of Whole Exome Sequencing Data

3.1.2.1 Preparation of Samples for WES

After clinical examinations by neurologists, the affected individuals were diagnosed with ET and blood samples were taken and stored at -80 °C. To determine the samples to send for WES, at least two patients and one control group were selected. For observing different gene variants, individuals were selected from different generations and the control subject was selected as a patient's sibling. For the whole exome sequencing study II-1 and IV-3 were selected as affected individuals and III-2 was chosen as a control from the ET-5 family, while IV-2, IV- 4, V-2, V-3 were selected as the affected group and IV-1 was the control from the ET-49 family. DNA samples of selected individuals were extracted with MN Nucleospin® Blood kit (Macherey-Nagel Inc., PA, USA) and concentrations of DNA samples were measured with a NanoDrop™ ND-1000 Spectrophotometer (NanoDrop Technologies, Inc, DE, USA) (Table 3.2). In addition, A₂₆₀/A₂₈₀ and A₂₆₀/A₂₃₀ values were used for checking the purity of DNA. To understand whether the DNA samples were degraded, all samples were run on agarose gel and compared with ladder (2-Log DNA Ladder (0.1-10.0 kb), NEB) (Figure 3.4). After concentration measurements and agarose gel electrophoresis, high-quality DNA samples were analyzed by WES.

Table 3.2 Concentrations and qualities of DNA samples isolated from ET-5 and ET-49 families

Family	Sample ID	Nucleic Acid Conc.	Unit	260/280	260/230
ET-5	II-1	197.4	ng/ μ L	1.91	2.23
	III-2	240.3	ng/ μ L	1.91	2.24
	IV-3	231.6	ng/ μ L	1.92	2.29
ET-49	IV-1	163.7	ng/ μ L	1.92	2.29
	IV-2	152.2	ng/ μ L	1.92	2.22
	IV-4	231.8	ng/ μ L	1.96	2.27
	V-2	284.2	ng/ μ L	1.98	2.23
	V-3	187.8	ng/ μ L	1.91	2.29

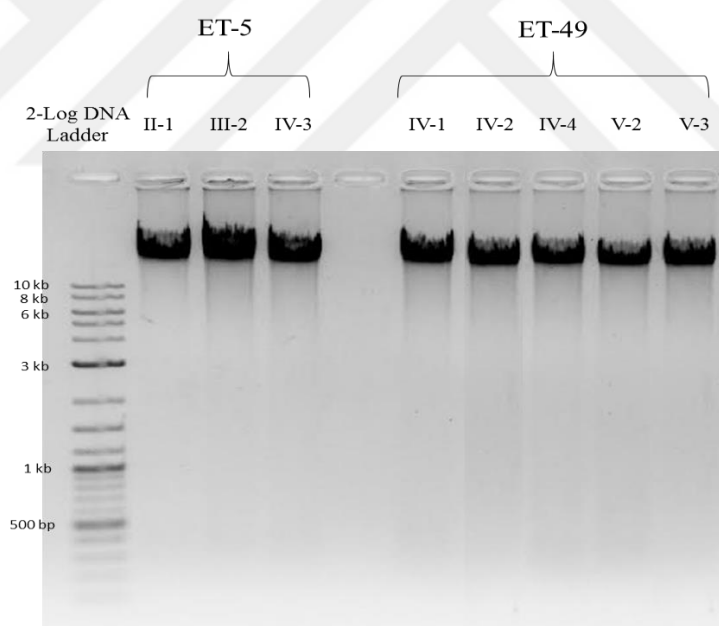


Figure 3.4 Density measurements using agarose gel electrophoresis. DNA samples were run on 1% agarose gel at 70 V for 50 minutes. Lane 1: 2-Log DNA Ladder Mix (NEB) (1 μ l); Lanes 2-5: DNA samples diluted 1:5 in TE (1 μ L). Gel image was captured with BioRad Gel Doc 2000 system. 2-Log DNA Ladder mix (10 kb: 40 ng, 3 kb: 120 ng) used as reference.

3.1.2.2 Whole Exome Sequencing Data Analysis

Sequences were obtained using Illumina HiSeq 2500 sequencer and the quality of sequences were determined with Illumina Real Time Analysis Software. Each fluorescence signal was converted to bases and FASTQ files containing raw data were prepared. Number of reads were calculated as nearly 50 million bases (Table 3.3) for the selected individuals. Using BWA software, fragments were aligned to human reference genome hg19, as illustrated in Table 3.3, and FASTQ files were converted to SAM files containing mapped sequences. Percentages were calculated for mapped reads among total reads. Samtools was used for separating PCR duplications and producing BAM files for clean sequences. In addition, BEDtools was applied to determine the coverage of exonic regions. The proportion of targeted exonic regions covered at least 5-fold was 98.91% for II-1, 99.02% for III-2 and 98.93% for IV-3 from the ET-5 family. For the ET-49 family the proportion of targeted exonic regions was calculated as 98.90% for IV-1, 99.17% for IV-2, 99.07% for IV-4, 99.08% for V-2 and 98.94% for V-3. SNP and Indel variants were determined using GATK (producing VCF (Variant Call Format) files) and characterized by ANNOVAR.

Table 3.3 Statistics on the genome sequence produced by the paired-end whole exome sequencing

Family	Sample ID	Reads ^a	Mapped_Reads ^b	Mapped_% ^c	Removed_dup_% ^d	AVG_Coverage ^e	Covered_at_least_5x_% ^f
ET-5	II-1	50252327	50156963	99.81%	1.39%	39,65756181	98.91%
	III-2	43449742	43336873	99.74%	1.11%	34,1321556	99.02%
	IV-3	52219448	52121947	99.81%	1.34%	40,61587965	98.93%
ET-49	IV-1	50746729	50630830	99.77%	1.39%	40,07058064	98.90%
	IV-2	52096346	51985778	99.79%	1.50%	41,76189517	99.17%
	IV-4	47206790	47120653	99.82%	1.21%	38,3545642	99.07%
	V-2	46100564	46029459	99.85%	1.25%	37,12614443	99.08%
	V-3	53873938	53798633	99.86%	1.41%	43,72322143	98.94%

^aTotal number of reads assembled by WES. ^b Number of sequencing reads mapped to human reference genome (hg19) with BWA. ^c Percentage of mapped reads among total reads. ^d Percentage of removed duplications with Samtools. ^e Effective fold coverage of the genome. ^f Percentage of the bases covered by at least one aligned read.

3.1.3 Variant Annotation and Effect Prediction

VCF (Variant Call Format) files were obtained for selected individuals and combined with GATK software. SnpEff (snpEff_v4_3_core) analysis package was used for variant annotation and effect prediction, with VCF files provided as an input for the genetic coordinates of variants against the human reference genome GRCh37.75 (hg19). SnpEff was used to get more information about these variants, including effect annotations (e.g. missense mutation, frameshift, splice site acceptor etc.), estimation of putative impact (high, moderate, low and modifier), feature type (e.g. transcript, motif, miRNA, etc.), HGVS.c (variant using HGVS notation at DNA level) and HGVS.p (variant using HGVS notation at protein level) values. In addition, cDNA and protein positions were obtained for each variant. For each family, percentage of variants were calculated depending on their region and effect types (Figure 3.5, Table 3.4, Table 3.5). After getting information about annotation impact of variants (qualified as high, moderate, modifier and low), modifier and low variants were filtered. This selection was performed because high and moderate impact variants are estimated to have a stronger deleterious effect compared to the other two levels. High variants include frameshift, splice acceptor, splice donor and stop loss variants, while moderate variants are inframe insertion/deletion, missense variant and splice region variants.

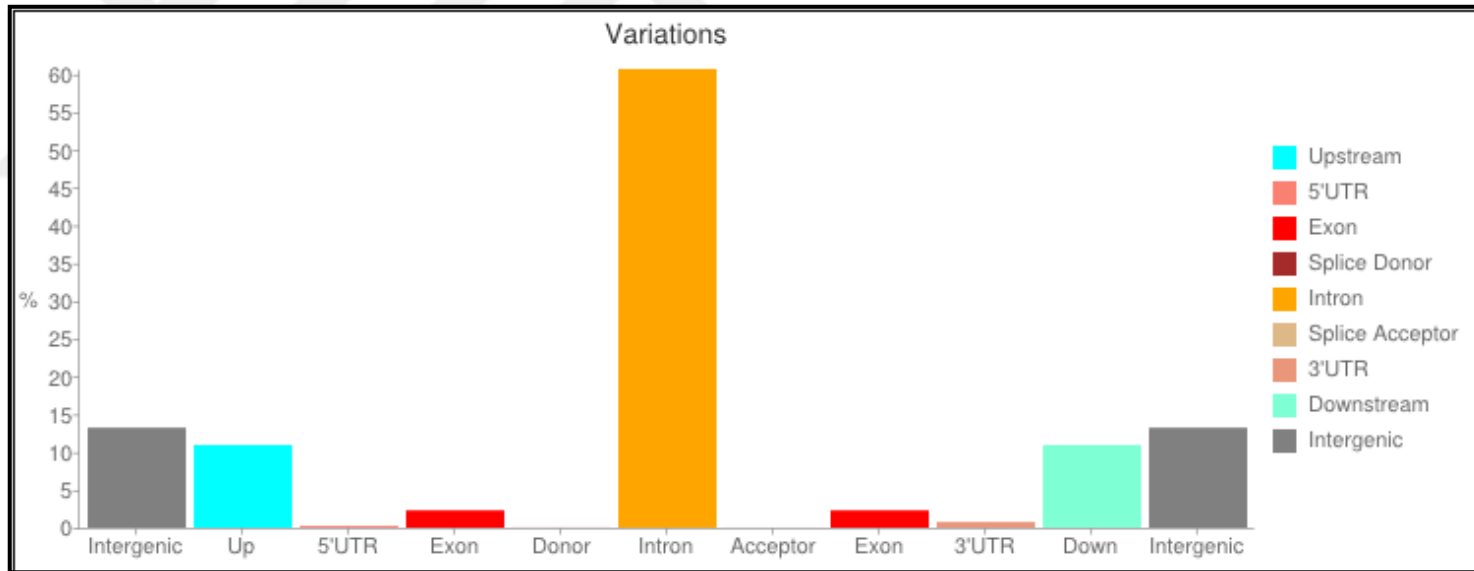


Figure 3.5 Percentage of variants with respect to genomic context for ET-5 and ET-49 families after SnpEff analysis against human database GRCh37.75 (hg19).

Table 3.4 Statistics on all variants based on their regions, as identified through whole exome sequencing and SnpEff analysis.

Family	Sample ID	Genome Total Length	Number of Variants	# of Known Variants	Number of Unknown Variants	Exon	Intergenic	Intron	Splice site Region
ET-5	II-1	37,322,280,797	959,458	875,102 (91.208%)	84,356 (8.8%)	85,806	417,640	2,025,291	6,776
	III-2	37,322,280,796	900,049	819,799 (91.084%)	80,250 (8.92%)	82,266	389,892	1,906,386	6,480
	IV-3	37,322,280,795	982,278	898,233 (91.444%)	84,045 (8.56%)	88,421	427,232	2,076,915	6,696
ET-49	IV-1	37,322,280,795	977,312	892,564 (91.328%)	84,748 (8.68%)	88,027	427,338	2,046,735	6,751
	IV-2	37,322,280,794	993,598	906,636 (91.248%)	86,962 (8.76%)	89,123	432,732	2,104,992	6,823
	IV-4	37,322,280,796	953,304	870,058 (91.268%)	83,246 (8.73%)	84,465	415,850	2,017,402	6,603
	V-2	37,322,280,796	941,493	858,289 (91.163%)	83,204 (8.84%)	84,524	408,696	1,995,097	6,654
	V-3	37,322,280,795	1,045,559	954,581 (91.299%)	90,978 (8.7%)	91,386	458,877	2,186,804	7,281

Table 3.5 Statistics on all variants based on their annotation effects, identified through SnpEff analysis.

Family	Sample ID	Missense	Frameshift	Stop-lost	Nonsense	Splice Site Acceptor	Splice Site Donor
ET-5	II-1	18,400	4,471	53	991	897	764
	III-2	17,216	4,166	52	1,053	883	713
	IV-3	18,557	4,482	43	993	841	768
ET-49	IV-1	18,833	4,455	46	1,094	832	863
	IV-2	18,945	4,630	51	1,074	867	768
	IV-4	17,815	4,366	54	1,020	804	672
	V-2	18,153	4,423	41	905	862	857
	V-3	19,608	4,492	50	1,035	881	826

3.1.3.1 Functional Annotation of Variants

After getting information about genomic context and variant prediction with SnpEff against the human reference genome hg19, the ANNOVAR software package was used to functionally annotate genetic variants. We performed three annotations as 1) Gene-based annotation for identifying SNPs, which cause protein coding changes; 2) Region-based annotation to determine variants in specific genomic regions and to learn the genomic coordination of variants; 3) Filter-based annotation to analyze variants that are reported in specific databases. Filter based annotation includes databases with information about population statistics, dbSNPs, damaging/evolutionary scores and disease-specific variants. ExAC (Exome Aggregation Consortium), 1000 Genome Project (1000G), CG (complete genomics) and ESP (Exome Sequencing Project) databases were utilized to obtain data about allele frequency and population statistics. In addition, prediction of deleteriousness based on protein-sequence and nucleotide-sequence was performed with SIFT, PolyPhen, MutationTaster, MutationAssessor, LRT, FATHMM, MetaSVM, MetalR, GERP++, PhyloP and SiPhy databases, which contain information about evolutionary, structural and biochemical characteristics of variants. Other databases were used to provide further information, including disease-related variants from ClinVar annotation and a combined evaluation of functional, deleteriousness and disease-causal variants from CADD annotation. Other allele frequency scores were also reported from our in-house database containing genomic information from 190 individuals, 52 of which are ET patients and represent 15 different ET families.

3.1.4 Identification of Disease Causing Determinants

3.1.4.1 Prioritization and Filtration of variants for ET-5 and ET-49 families

By using SnpEff analysis, information about variant annotation and effect prediction were completed. In addition, to get more knowledge about frequency and evolutionary deleteriousness of variants ANNOVAR software package was performed. Filtration of variants were carried out in two ways; 1) removing variants which do not have protein altering effect, and 2) removing variants which have minor allele frequency (MAF) bigger than 0.02. First filtration was carried out with filtering low and modifier variants after SnpEff analysis from VCF file. This enabled us to proceed with only protein coding variants including frameshift, missense, non-frameshift indel, stop-gained, stop-loss, start-gain, start-loss, splice site variants which are more potentially damaging than other annotation effects. After this filtration, number of variants was decreased from 5,895,636 variants to 1,011,279 variants. By ANNOVAR analysis, all variants were filtered depending on their allele frequency from ExAC database, and variants which have minor allele frequency (MAF) higher than 0.02 were removed. This filtration decreased the number of variants from 1,011,279 variants to 946,856 variants. We filtered variants that do not have protein altering effect and variant number is decreased from 946,856 to 29,197 variants. Finally, we removed variants that were found homozygous in control individuals and at the end we obtained 357 rare variants that had protein altering effect and allele frequency less than or equal to 0.02 (Figure 3.6)

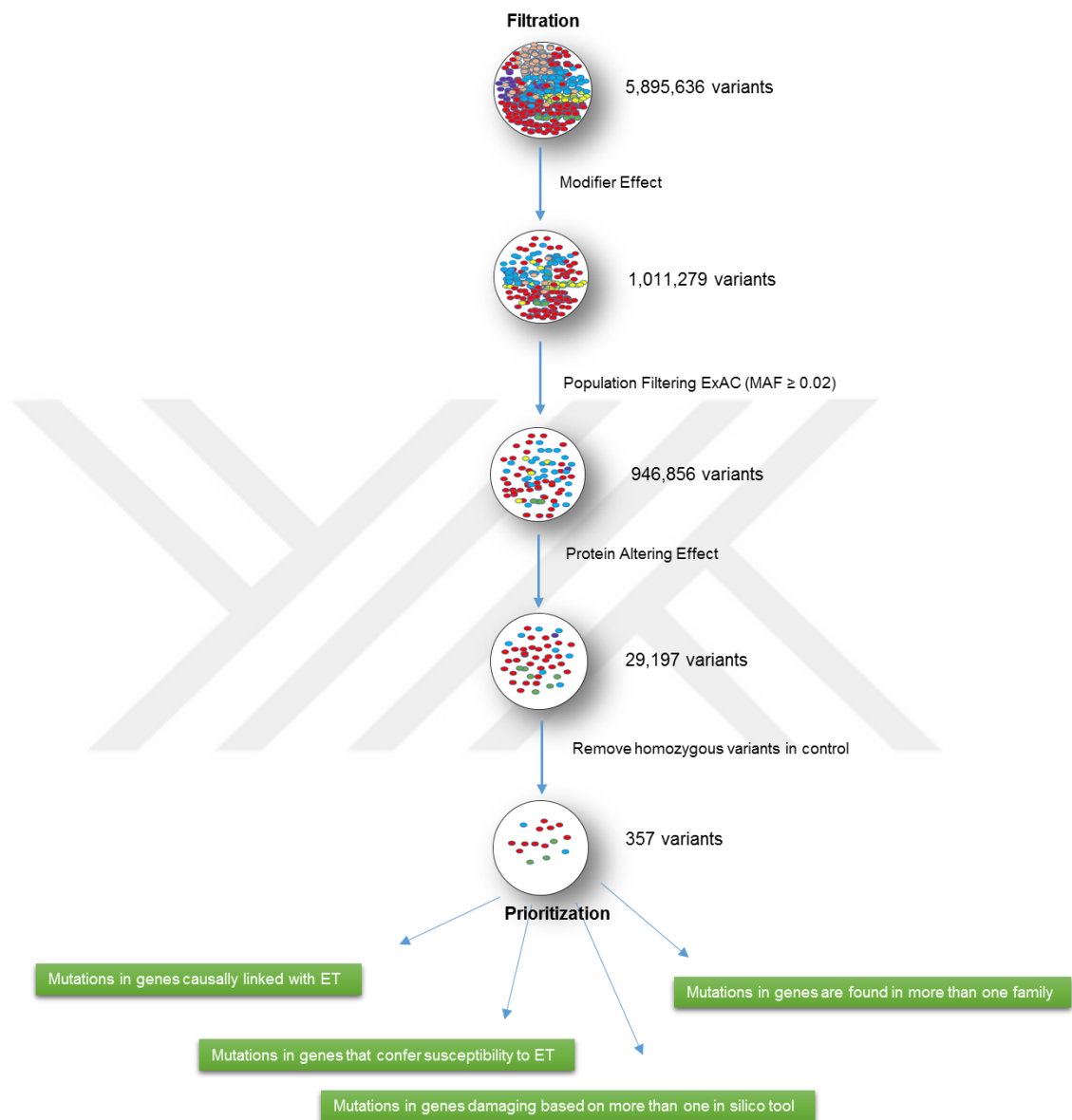


Figure 3.6 Filtration and prioritization pipeline for disease causing determinants.

3.1.4.2 Prioritization of candidate gene variants for ET-5 and ET-49 families

We created an excel table including variants that have protein altering effect and allele frequency is less than or equal to 0.02. First, non-protein coding variants were deprioritized according to their transcript biotype as pseudogene, retained intron, lincRNA, etc. Utilizing our in-house database, variants which were homozygous in more than 10 individuals in 190 individuals were excluded. We also excluded homozygous variants in more than 20 individuals in ExAC database. Variants were eliminated regarding to allele frequency in 1000G and EVS databases which have minor allele frequency larger than 0.02.

Another deprioritization strategy utilized evolutionary conservation databases. Variants which had scores Provean <-2.5 , SIFT ≤ 0.05 , Polyphen2 HDIV; D: Probably damaging ≥ 0.957 , P: possibly damaging $0.453 \leq \text{pp2hdiv} \leq 0.956$, Polyphen2 Hvar D: Probably damaging ≥ 0.909 , P: possibly damaging $0.447 \leq \text{pp2hvar} \leq 0.909$, MutationTaster prediction for A (disease causing automatic) and D (disease causing), MutationAssessor >1.94 (M: Moderate and H: High), CADD ≥ 10 , GERP ≥ 3 , PhastCons_Vertebrate ≥ 0.90 and PhastCons_Mammalian ≥ 0.90 were included and labelled for each specific characteristic. We deprioritized variants if none of the tools predicted them as damaging or if only one of the tools predicted them as damaging, while the score was not too high. We prioritized variants if at least more than one prediction tool determined them as damaging with high damaging scores.

At first, we listed all the damaging variants in probands of the cohort that were reported in at least one *in silico* tool. Simultaneously, we listed all the damaging variants of all controls in the cohort. In the intrafamily prioritization, we filtered a variant if it is also found as homozygous or compound heterozygous in the genome of his/her control

sample. We selected the variants that are shared by affected individuals within the same family.

Under the hypothesis of a recessive inheritance model, we selected all rare variants in autosomal genes that were also present in the presumed heterozygous state in other affected individuals. We filtered heterozygous variants only if the control sibling has the same variation and/or any of the control individuals has the same variant in homozygous state. According to that, Table 3.6 and Table 3.7 were prepared for homozygous variants that were identified in proband of each family.

For ET-5 family, candidates shown with light gray color (U1, LINC00854, etc.) were eliminated for their transcript biotype as lincRNA, antisense, pseudogenes (Table 3.6). Variants which are illustrated by moderate gray were deprioritized because their minor allele frequency is higher than 0.02 according to 1000 genome project database. Candidate variants which appeared as dark gray (MAP3K4, SLC9B1P1) were dismissed because of their frequency in our in-house database. We suspected that CNTNAP3B and SLC9B1P1 could be candidate variants for ET-5 family. After examination of 15 other ET families' sequence data, we recognized these variants were also included in controls of other families. At the end, these two candidate variants CNTNAP3B and SLC9B1P1 were also eliminated.

For ET-49 family, candidates are illustrated with colors showing their elimination reason (Table 3.7). Candidate variants for ET-49 family were selected as MMP19 p.Arg456Gln, TMEM139 p.Arg82Gln and ATXN3 p.Lys304_Gln312dup. These candidates are homozygous for proband and heterozygous for other affected individuals and not included in control individual of ET-49 family. After this decision, we planned for segregation analysis of these determinants.

We searched dominant model candidates for ET-5 family that is shown in Table 3.8. Based on filtering criteria, four important variants were selected for segregation analysis that are heterozygous in all affected individuals of ET-5 family and not included in control group. These are MMP19 p.Arg456Gln, COL24A1 p.Gly657Val, ST7L p.Pro482Arg and OTOF P.Arg49Trp missense variants which had protein altering effect and allele frequency was less than 2%. To determine whether a variant is related with ET or not, segregation analysis was performed for each specific variant.



Table 3.6 List of the homozygous damaging variants in proband of the ET-5 family

Chr	Start	Ref	Alt	Annotation	Impact	Gene_Name	Transcript_BioType	HGVS.c	HGVS.p	ExAC	1000G	ESP	PRVN	SIFT	PPH2	MT	MA	CADD	GERP	PCONS
9	43915500	G	C	missense_variant	M	CNTNAP3B	protein_coding	c.3583G>C	p.Gly1195Arg	0.0032	0.008786	.	D	D	B	D	M	5.507	1.31	0.01
Y	13496371	T	A	missense_variant	M	SLC9B1P1	protein_coding	c.778A>T	p.Met260Leu	.	.	.	N	T	B	.	.	12.86	0.174	0.993
1	17200186	A	C	splice_acceptor_variant	H	U1	lincRNA	n.86-2A>C	.	.	.	.	.	.	.	.	.	5.346	-0.641	0.002
1	15268168	-	AGCTCT(G) _n CTGCTGT	disruptive_inframe_insertion	M	LCE4A	protein_coding	c.143_144insCTGTAGCTCT(G) _n CTG	p.Ser43_Cys48dup	.	0.800519	.	N	.	.	.	.	9.225	4.63	0.01
3	12135131	-	GGCTCAGGCTCA	disruptive_inframe_insertion	M	HCLS1	protein_coding	c.1092_1103dupTGAGCCTGAGCC	p.Glu365_Pro368dup	.	0.241214	.	N	.	.	.	.	10.66	2.32	0.057
3	14937528	G	A	splice_donor_variant	H	WWTR1	protein_coding	c.-4+2C>T	.	.	0.388578	.	.	.	.	.	.	21.7	-0.29	0.997
6	16151935	CTG	-	disruptive_inframe_deletion	M	MAP3K4	protein_coding	c.3593_3598delCTGCTG	p.Ala1198_Ala1199del	.	.	.	N	.	.	.	.	20.9	3.19	0.759
9	35829390	C	T	splice_donor_variant	H	TMEM8B	protein_coding	c.-928+2C>T	.	0.002	0.61901	.	.	.	.	.	.	21.5	3.1	1
12	51740410	A	G	missense_variant	M	CELA1	protein_coding	c.13T>C	p.Tyr5His	0.001	0.367412	.	N	T	B	N	N	8.649	2.28	0.55
12	51740413	-	C	frameshift_variant	H	CELA1	protein_coding	c.9_10insG	p.Leu4fs	.	0.367412	.	.	.	.	.	.	7.595	1.78	0.523
12	51740415	AC	-	frameshift_variant	H	CELA1	protein_coding	c.7_8delGT	p.Val3fs	.	0.367412	.	.	.	.	.	.	8.987	0.701	0.576
17	41382215	A	G	splice_donor_variant	H	LINC00854 RP11-64C12.3	antisense	n.97+2T>C	.	.	.	.	.	.	.	.	.	8.646	0.149	0.075
18	12213597	A	G	splice_donor_variant	H	64C12.3	_pseudogene	n.191+1A>G	.	.	.	.	.	.	.	.	.	1.419	-1.52	0.02
22	20709307	A	G	missense_variant	M	FAM230A	protein_coding	c.1039A>G	p.Asn347Asp	0.0082	0.473842	.	.	.	B	.	.	0.043	-1.94	0
22	20709308	A	C	missense_variant	M	FAM230A	protein_coding	c.1040A>C	p.Asn347Thr	0.006	0.508986	.	.	.	B	.	.	1.669	0.971	0
22	20710985	A	C	stop_lost	H	FAM230A	protein_coding	c.2717A>C	p.Ter906Serext*?	0.0002	0.449081	.	.	.	.	.	.	0.152	-0.846	0
7	14196945	C	T	splice_donor_variant	H	U66059.29	pseudogene	n.498+2G>A	.	.	0.205671	.	.	.	.	.	.	14.08	3.19	0.081
14	92587538	A	G	splice_donor_variant	H	NDUFB1	protein_coding	c.-6+2T>C	.	.	0.539736	.	.	.	.	.	.	6.167	-1.81	0
Y	13496413	T	G	missense_variant	M	SLC9B1P1	protein_coding	c.736A>C	p.Thr246Pro	.	.	.	N	T	B	.	.	0.051	-2.75	0.002

Non-protein coding variants
MAF> 0.02 variants
High score for deleteriousness

Abbreviations used in this table; Chr: Chromosome, Ref: Reference, Alt: Alteration, HGVS.c: Human Genome Variation Society, cDNA, HGVS.p: Human Genome Variation Society, protein, ExAC: Exome Aggregation Consortium, 1000G: 1000 Genomes Project, ESP: Exome Sequencing Project, PRVN: PROVEAN, PPH2: PolyPhen2, MT: MutationTaster, MA: MutationAssesor, CADD: Combined Annotaion Dependent Depletion, GERP: Genomic Evolutionarily Rate Profiling, PCONS: PhastCons

Table 3.7 List of the homozygous variants in proband of ET-49 family

Chr	Start	Ref	Alt	Annotation	Impact	Gene_Name	Transcript_BioType	HGVS.c	HGVS.p	ExAC	1000G	ESP	PRVN	SIFT	PPH2	MT	MA	CADD	GERP	PCONS
12	56230980	C	T	missense_variant	M	MMP19	protein_coding	c.1367G>A	p.Arg456Gln	0.00369	0.0013	0.0035	N	T	P	D	L	19.75	3.88	0.984
7	142983295	G	A	missense_variant	M	TMEM139	protein_coding	c.245G>A	p.Arg82Gln	0.0023	0.0015	0.0019	D	T	P	N	L	20.2	3.13	0.999
14	92537361	-	(GCT) ₉	inframe_insertion	M	ATXN3	protein_coding	c.935_936ins(CAG) ₉	p.Lys304_Gln312dup	.	.	.	N	.	.	.	.	12.83	1.98	0.999
6	170871013	CAACAGCAA	-	disruptive_inframe_deletion	M	TBP	protein_coding	c.216_218delACA	p.Gln73del	.	.	.	N	.	.	.	.	16.29	0.405	0.95
5	156479424	AGT	-	inframe_deletion	M	HAVCR1	protein_coding	c.619_621delACT	p.Thr207del	.	.	.	N	.	.	.	.	14.06	1.05	0.003
X	123195593	-	TT	splice_acceptor_variant	H	STAG2	protein_coding	c.1535-4_1535-3dupTT	.	0.002	.	.	.	.	.	.	.	5.148	0.908	0.009
13	29297506	T	-	splice_acceptor_variant	H	CYP51A1P2	pseudogene	n.518-2delA	.	.	.	.	.	.	.	.	.	9.813	0	0.095
19	36806431	C	T	splice_acceptor_variant	H	LINC00665	lincRNA	n.547-1G>A	.	.	.	.	.	.	.	.	.	1.491	0.149	0.034
4	38120014	-	T	splice_acceptor_variant	H	TBC1D1	retained_intron	n.516-3dupT	.	.	0.6851	.	.	.	.	.	.	6.449	2.46	0.001
6	170871038	-	CAG	disruptive_inframe_insertion	M	TBP	protein_coding	c.213_215dupGCA	p.Gln72dup	.	.	0.73	N	.	.	.	.	17.48	4.6	0.994
19	53703876	G	A	splice_acceptor_variant	H	CTD-2245F17.3	lincRNA	n.153-1G>A	.	.	0.4540	.	.	.	.	.	.	8.36	0.877	0.001
2	171572580	T	C	splice_donor_variant	H	SP5	processed_transcript	n.27+2T>C	.	.	0.2154	.	.	.	.	.	.	13.7	2.61	0.177
4	119517066	G	A	splice_acceptor_variant	H	RP11-384K6.6	lincRNA	n.124-1G>A	.	.	0.2579	.	.	.	.	.	.	11.86	0.14	0.091
6	32489934	C	A	missense_variant	M	HLA-DRB5	protein_coding	c.118G>C	p.Asp40His	.	.	.	N	T	B	N	N	0.001	-9.62	0
12	8386871	C	G	splice_donor_variant	H	FAM86FP	_pseudogene	n.875+1G>C	.	.	0.7340	.	.	.	.	.	.	6.133	0.241	0.071
16	64480	G	C	missense_variant	M	WASH4P	protein_coding	c.1335C>G	p.Asp445Glu	0.0001	.	.	N	T	B	N	.	1.592	-0.758	0.111

Non-protein coding variants
MAF> 0.02 variants
High score for deleteriousness

Abbreviations used in this table; Chr: Chromosome, Ref: Reference, Alt: Alteration, HGVS.c: Human Genome Variation Society, cDNA, HGVS.p: Human Genome Variation Society, protein, ExAC: Exome Aggregation Consortium, 1000G: 1000 Genomes Project, ESP: Exome Sequencing Project, PRVN: PROVEAN, PPH2: PolyPhen2, MT: MutationTaster, MA: MutationAssessor, CADD: Combined Annotation Dependent Depletion, GERP: Genomic Evolutionarily Rate Profiling, PCONS: PhastCons

Table 3.8 List of the heterozygous variants in proband of ET-5 family

Chr	Start	Ref	Alt	Annotation	Impact	Gene_Name	Transcript_BioType	HGVS.c	HGVS.p	ExAC	1000G	ESP	PRVN	SIFT	PPH2	MT	MA	CADD	GERP	PCONS
12	56230980	C	T	missense_variant	M	MMP19	protein_coding	c.1367G>A	p.Arg456Gln	0.0037	0.0014	0.0035	N	T	P	D	L	19.75	3.88	0.995
1	86499587	C	A	missense_variant	M	COL24A1	protein_coding	c.1970G>T	p.Gly657Val	0.0029	0.0004	0.0036	D	D	D	D	M	22.3	5.36	1
1	113093267	G	C	missense_variant	M	ST7L	protein_coding	c.1445C>G	p.Pro482Arg	.	.	.	D	D	D	D	M	28.9	5.69	1
2	26750782	G	A	missense_variant	M	OTOF	protein_coding	c.145C>T	p.Arg49Trp	0.0084	0.0094	0.0043	D	D	D	D	M	34	5.4	1
1	50572019	T	G	missense_variant	M	ELAVL4	protein_coding	c.14T>G	p.Leu5Arg	0.0002	0.0012	.	N	D	B	D	.	21	5.2	1
1	87369131	C	A	missense_variant	M	SEP15	protein_coding	c.76G>T	p.Val26Leu	0.0110	0.0098	0.013	N	T	B	N	N	22	3.71	0.998
1	111998734	A	G	missense_variant	M	ATP5F1	protein_coding	c.250A>G	p.Ile84Val	0.0027	0.0018	0.0029	N	T	B	N	L	15.29	3.91	1
1	14105137	GAT	-	disruptive_inframe_deletion	M	PRDM2	protein_coding	c.861_863delTGA	p.Asp287del	0.0074	.	.	N	.	.	.	.	10.57	2.02	0.972
3	130733047	-	T	frameshift_variant	H	ASTE1	protein_coding	c.1969delA	p.Arg657fs	.	.	.	.	.	.	.	.	12.23	5.81	0.014
3	180320898	-	T	splice_acceptor_varian	H	TTC14	protein_coding	c.287-3dupT	.	.	.	.	.	.	.	.	.	7.68	0.938	0.012
3	124829109	A	G	missense_variant	M	SLC12A8	protein_coding	c.1070T>C	p.Leu357Pro	0.0001	.	.	D	D	D	D	M	17.03	5.26	1
11	1092506	T	C	missense_variant	M	MUC2	protein_coding	c.4325T>C	p.Ile1442Thr	0.0009	0.1639	.	N	T	B	N	.	0.005	-1.53	0.041
11	6977311	G	A	missense_variant	M	ZNF215	protein_coding	c.1103G>A	p.Arg368Gln	0.0027	0.0078	0.0004	N	T	B	N	N	0.001	-9.12	0
11	18729819	C	T	missense_variant	M	IGSF22	protein_coding	c.3022G>A	p.Ala1008Thr	0.0190	0.0192	0.05	N	T	B	N	L	20.1	3.33	0.963
11	56143261	T	A	missense_variant	M	OR8U1	protein_coding	c.162T>A	p.Ser54Arg	.	.	.	N	T	B	N	N	0.001	-3.07	0
11	61026266	C	T	missense_variant	M	VWCE	protein_coding	c.2749G>A	p.Gly917Arg	0.0020	0.0008	0.0023	N	T	B	N	N	0.003	-2.62	0
11	62288649	G	C	missense_variant	M	AHNAK	protein_coding	c.13240C>G	p.Leu4414Val	0.0120	0.0327	0.0014	N	T	P	P	L	11.21	-6.22	0
11	65793121	C	T	missense_variant	M	CATSPER1	protein_coding	c.730G>A	p.Gly244Arg	0.0130	0.0306	0.0007	N	D	D	P	M	13.07	-4.55	0

Non-protein coding variants

MAF> 0.02 variants

High score for deleteriousness

Abbreviations used in this table; Chr: Chromosome, Ref: Reference, Alt: Alteration, HGVS.c: Human Genome Variation Society, cDNA, HGVS.p: Human Genome Variation Society, protein, ExAC: Exome Aggregation Consortium, 1000G: 1000 Genomes Project, ESP: Exome Sequencing Project, PRVN: PROVEAN, PPH2: PolyPhen2, MT: MutationTaster, MA: MutationAssessor, CADD: Combined Annotaion Dependent Depletion, GERP: Genomic Evolutionarily Rate Profiling, PCONS: PhastCons

3.1.4.3 Database Search

Information from a broad range of databases is necessary to best understand the biological functions of candidate variants. For this purpose, the Online Mendelian Inheritance in Man (OMIM) database was used to determine the relationships between genotype and phenotype. The Universal Protein Resource (UniProt) database was accessed for its information on protein function, subcellular location, induction and tissue specificity. Moreover, Mouse Genome Informatics (MGI) was used to evaluate variant function through integrated genetic, genomic, and biological data in mice.

3.1.5 Segregation Analysis of Candidate Variants for ET-5 and ET-49 Families

After primer designing of candidate variants for ET-5 and ET-49 families, PCR products of each individual were prepared for Sanger sequencing. According to Sanger sequencing results, the MMP19 p.Arg456Gln missense variant co-segregated with ET-5 and ET-49 families (Figure 3.7, 3.8). After genotyping MMP19 p.Arg456Gln for all individuals that had DNA samples, we determined that 8 affected individuals from ET-5 and ET-49 families were heterozygous for the variant, while one individual from the ET-49 family was homozygous and all controls were wildtype (Figure 3.9, Figure 3.10). Other candidate variants did not segregate with the disease.

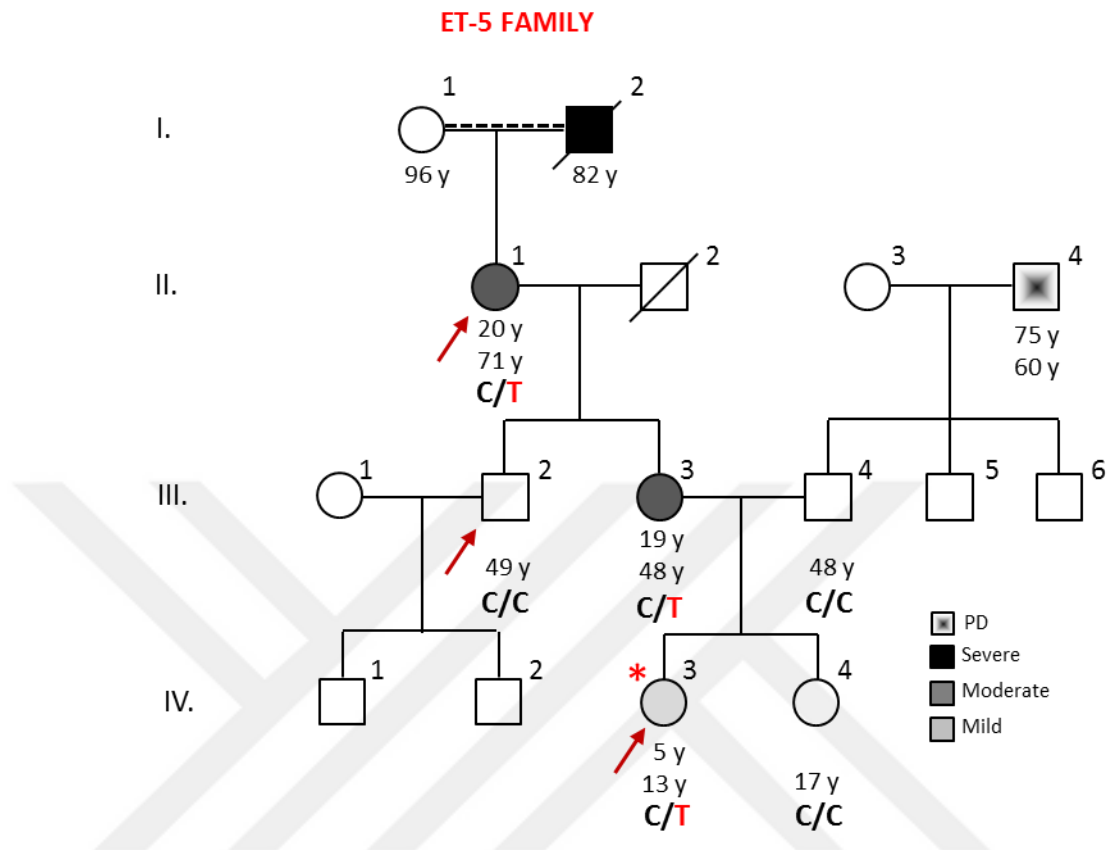


Figure 3.7 Pedigree of ET-5 family showing the segregation of essential tremor with the MMP19 p.Arg456Gln (c.1367 C>T) variant. Red arrows indicate the individuals which were selected for whole exome sequencing (II-1, III-2 and IV-3). Proband is illustrated with asterisk (*). Individuals with ET are illustrated with filled light gray if the clinical severity is moderate, light gray for mild clinical severity and black for severe ET. Individual with PD is illustrated with gray color with black point. Genotypes for MMP19 p.Arg456Gln, age at onset and current ages are shown under the symbols. C indicates the reference base and T indicates the alteration base.

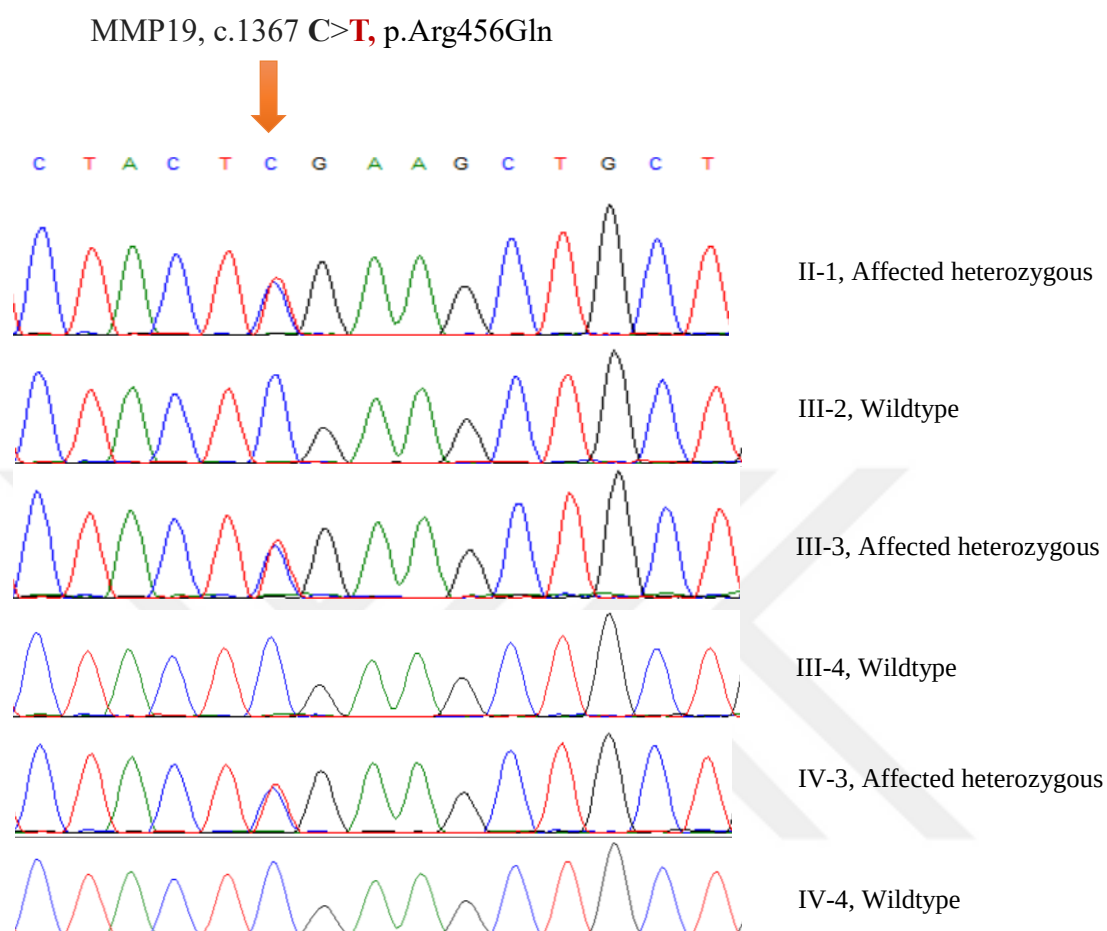


Figure 3.8 Segregation analysis results of the ET-5 family. In affected individuals of the ET-5 family MMP19 p.Arg456Gln (c.1367 C>T) missense mutation segregated in a heterozygous condition that is shown with two signals. **C** base (blue signal) is converted to **T** base (red signal) in ET-diagnosed patients of the ET-5 family.

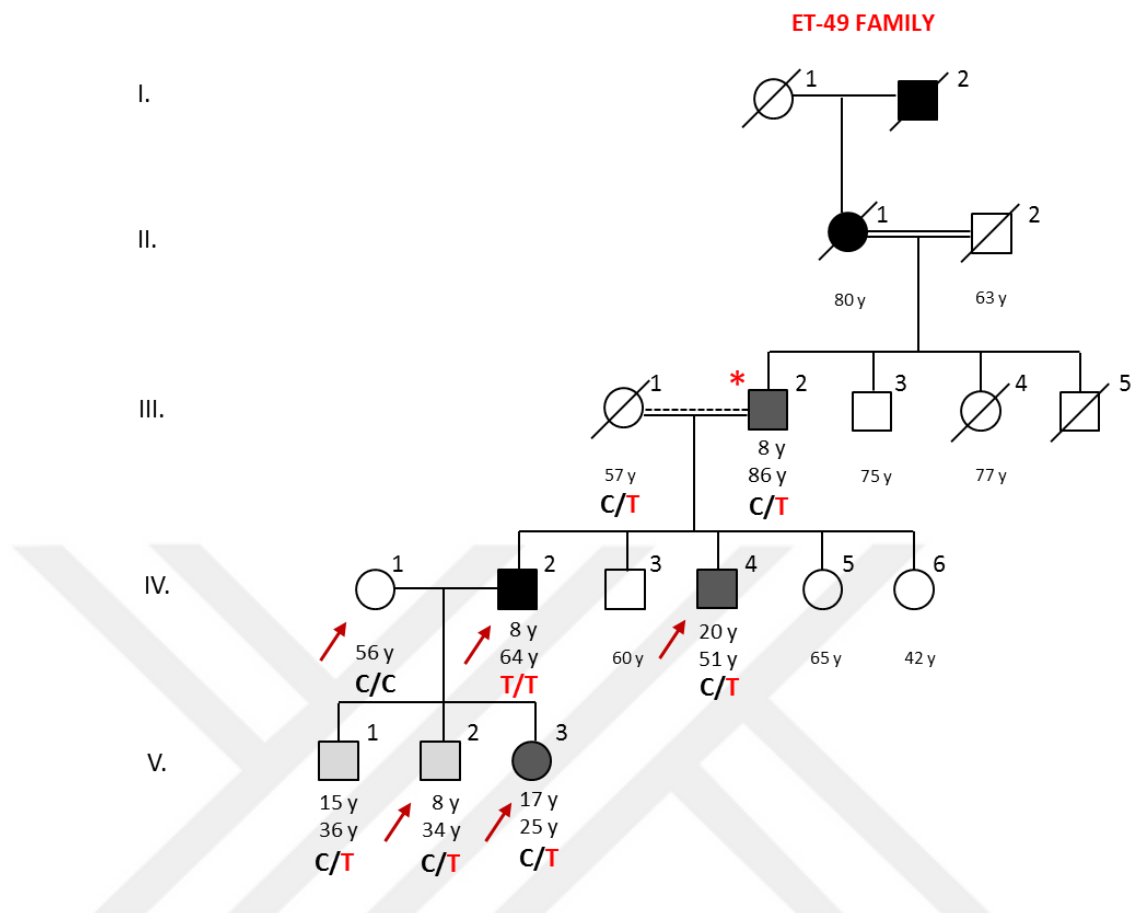


Figure 3.9 Pedigree of ET-49 family showing the segregation of ET with the MMP19 p.Arg456Gln (c.1367 C>T) variant. Red arrows indicate the individuals which were selected for whole exome sequencing (IV-1, IV-2, IV-4, V-2 and V-3). Proband is illustrated with asterisk (*). Individuals with ET are illustrated with filled light gray if the clinical severity is moderate, light gray for mild clinical severity and black for severe ET. Genotypes for MMP19 p.Arg456Gln, age at onset and current ages are shown under the symbols. **C** indicates the reference base and **T** indicates the alteration base.

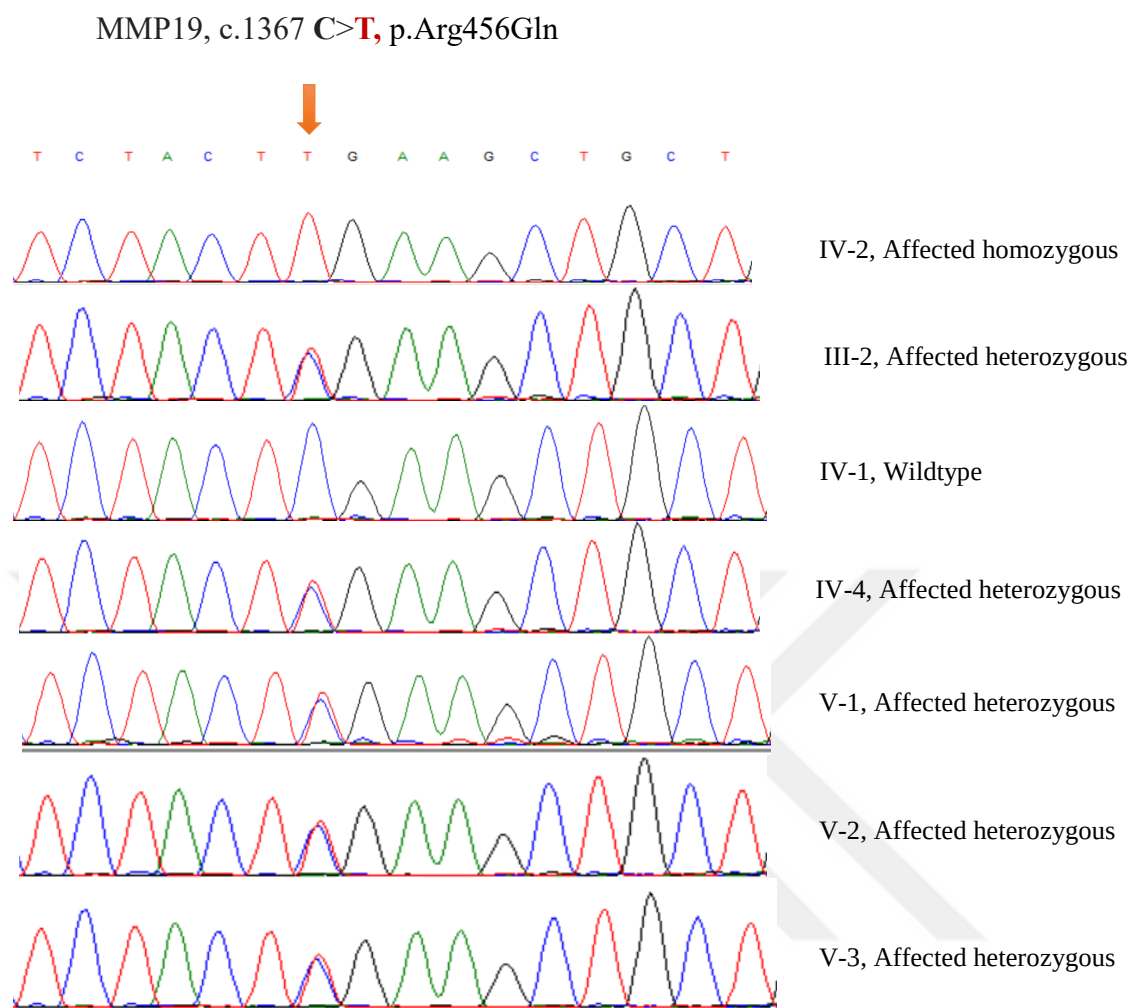


Figure 3.10 Segregation analysis results of the ET-49 family. In affected individuals of the ET-49 family MMP19 p.Arg456Gln (c.1367 C>T) missense mutation segregated in a heterozygous condition that is shown with two signals. One homozygous individual (IV-2) has a single, red signal for the T base. **C** base (blue signal) is converted to **T** base (red signal) in ET-diagnosed patients of the ET-49 family.

3.1.5.1 MMP19 p.Arg456Gln as the Disease-Causing Mutation for ET-5 and ET-49 Families

Following the post-processing of genomic data, a putative ET-causing variant was found at chr12: 56,230,980 (hg19; c.1367G>A), resulting in an arginine (Arg, R) to glutamine (Gln, Q) substitution at the residue 456 of exon 9 of the matrix metalloproteinase-19 (MMP-19, RASI, MMP-18, ENSG00000123342, ENST00000322569) gene. This mutation co-segregates with the disease in the ET-5 and ET-49 families (Figure 3.7).

As a matrix metalloproteinase, MMP-19 plays a role in breaking down various components of the extracellular matrix, such as aggrecan and extracellular matrix protein components, during embryonic development, reproduction and tissue remodeling. Isoform 1 of MMP19 encodes 508 amino acids. According to NCBI conserved domain database search, the protein has four protein family domains. These are ZnMc superfamily (Zinc-dependent metalloprotease), HX superfamily (Hemopexin-like repeats), Peptidase_M10 (Matrixin) domain and PG binding domain (Putative peptidoglycan binding domain) (Figure 3.11).

The mutation we have identified is located in the Hemopexin-like repeats. These repeats are important for creating a region for the binding of tissue inhibitors of metalloproteinases (TIMPs). The catalytic site includes the active region for zinc ion binding and functions in catalytic activities such as substrate degradation¹⁸⁵. Catalytic activities of MMPs are regulated by tissue inhibitor metalloproteinases (TIMPs). MMP19 is strongly inhibited by TIMP-2, TIMP-3 and TIMP-4 and less efficiently by TIMP-1¹⁸⁶.

Multiple sequence alignment analysis using Clustal Omega revealed that the arginine (R) allele is conserved in many species (Figure 3.12). Protein sequences of sequenced species were downloaded from Ensembl database, and the conservation of MMP19 p.456Arginine was also supported by evolutionary scores: A PhastCon score of 0.999 (>0.99 is for strongly conserved sequences), GERP score of 3.88 (>3 is strong evolutionary conservation) and CADD score of 23.4 (>10 is strong evolutionary conservation) were observed for this particular residue.



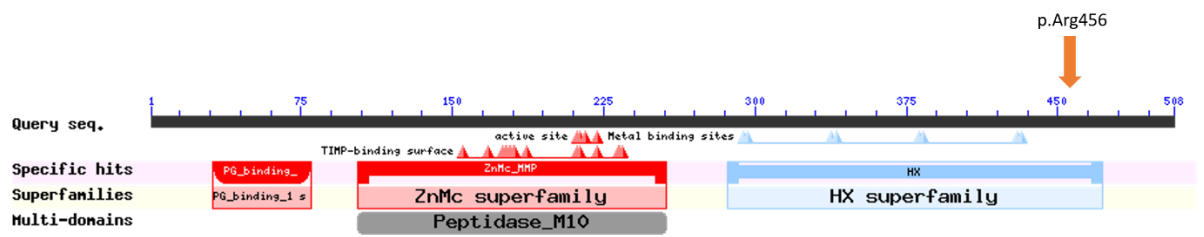


Figure 3.11 Predicted protein and domain structure of MMP19 protein, based on the NCBI Conserved Domain Database.



Figure 3.12 Sequence homology of MMP19 protein p.Arg456 region among various species. The box indicates the mutant amino acid p.Arg456.

3.1.6 Modelling 3D Structure of MMP19 p.Arg456Gln

A model 3D structure of MMP19 was generated using Swiss-Pdb viewer (Deep view version 4.0). 3D structures of wildtype and mutant protein are illustrated in Figure 3.13. 3D structure of wildtype MMP19 was derived from the Phyre2 Protein Fold Recognition Server. We observed that the variant alters the molecular surface and electrostatic indications (which are shown by colored dots), and creates a strong hydrogen bond between Gln456 and Gln454 (Figure 3.13).

In MMP19 p.Arg456Gln missense variation, arginine (which is a positively charged amino acid) was changed to glutamine (which has polar uncharged side chains), suggesting that the residue charge may play a role in the observed alteration in protein function.

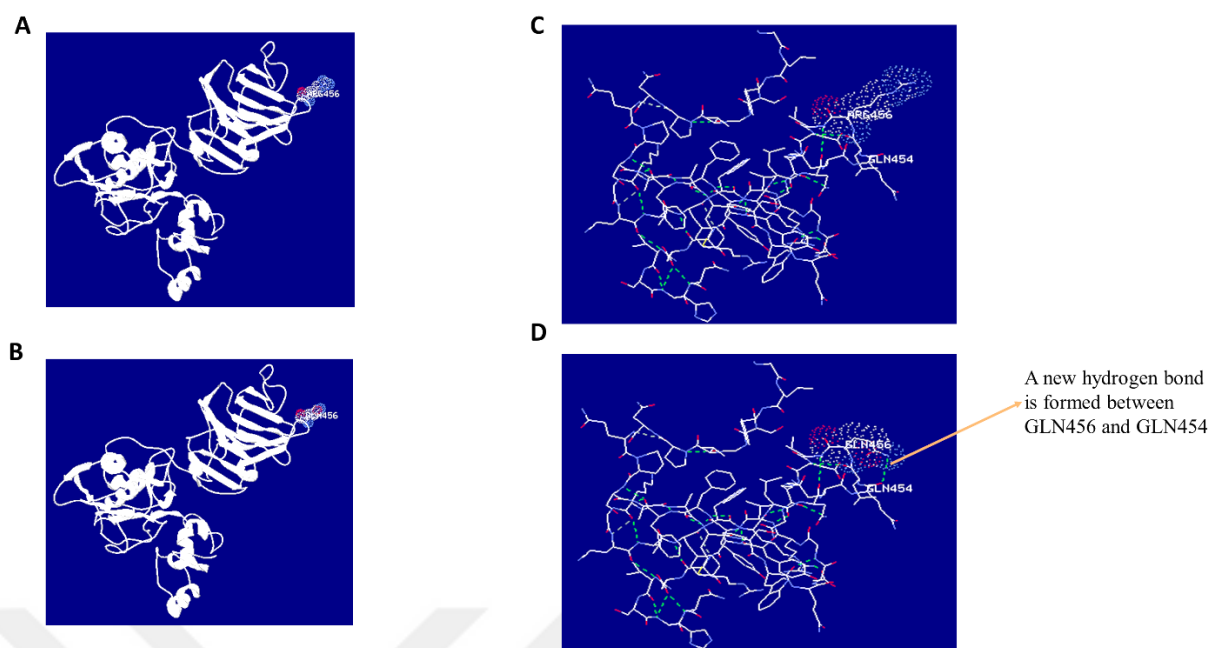


Figure 3.13 Model 3D protein structure of MMP19 protein is represented using Swiss-Pdb viewer. (A) Predicted 3D structure for p.Arg456 wildtype MMP19 protein. The amino acid 456 is indicated. (B) Predicted 3D structure for p.Gln456 mutant MMP19 protein. (C) Wildtype MMP19 protein structure showing specific bonds between amino acids (D) Mutant MMP19 protein structure showing specific bonds between amino acids. Conversion of p.Arg456Gln resulted in formation of hydrogen bond between Gln456 and Gln454 that is illustrated by arrow.

3.2 IMPG1, Interphotoreceptor Matrix Proteoglycan 1, p.Arg580Cys is a Missense Mutation that Causes Essential Tremor in a Consanguineous Family

3.2.1 Clinical Features of the Family

The ET-31 family originates from northern Anatolia. The proband (subject III-2), who is a 57 year-old male, was referred to Hacettepe University Medical School because of his severe hand tremor at age 19. His ET is clinically severe, and he has large amplitude, postural and kinetic tremors for both hands. He also has head and voice tremor. Additionally, he has hypomimia and right hand bradykinesia and rigidity. His Archimedes spiral test also reflects the severity of his ET.

According to his report, his mother and father's siblings had ET. Other siblings of the proband were also diagnosed with ET. III-4 is an affected individual who has severe clinical severity and is 65 years old; her age of onset was 50. She has moderate postural and kinetic tremor and low-amplitude right hand resting tremor. She has also voice and head tremor. She has additional clinical conditions in acathisia and depression. Another sister of the proband, III-8, has severe ET conditions like III-2 and III-4. She is 63-years-old and her age at onset was 40. She has higher scores for left hand resting, left hand postural and left hand kinetic tremors. She also has head and voice tremors and moderate right hand resting, postural and kinetic tremors. Her intentional tremor was reported as low-amplitude. In addition, she has mild right hand bradykinesia and moderate right and left hand rigidity (Figure 3.14, Table 3.9).

Other individuals (III-15, III-21, IV-1, IV-13, IV-16 and IV-17) have been observed to present with moderate-severity ET, including only low score postural tremor. In

addition to this, III-21 has left hand rigidity and IV-16 has low left and right hand rigidity. Control groups of the ET-31 family (II-13, III-1 and V-1) have no resting, postural, kinetic or intentional tremors. III-1 has right hand rigidity and depression problems. IV-6 has right and left hand rigidity. IV-7 has right hand rigidity and mild postural instability. All clinical features for these individuals are illustrated in Table 3.9, with additional information and Archimedes spiral tests for affected and unaffected individuals in Figure 3.15.



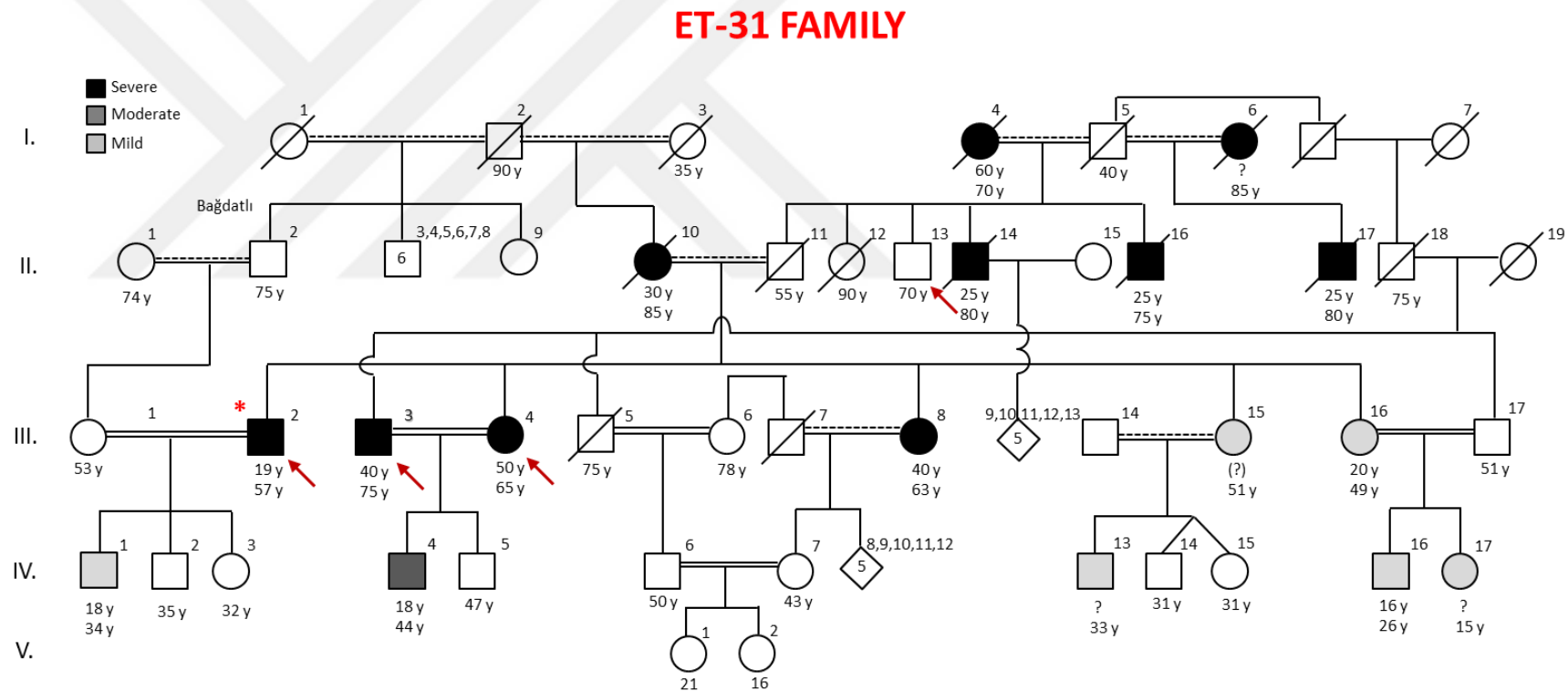


Figure 3.14 Pedigree of ET-31 family showing the affected and unaffected individuals with essential tremor. Red arrows indicate the individuals which were selected for whole exome sequencing (II-13, III-2, III-3 and III-4). Proband is illustrated with asterisk (*). Individuals with ET are illustrated with filled light gray if the clinical severity is moderate, light gray for mild clinical severity and black for severe ET.

Table 3.9 Clinical characteristics of tremor in affected and unaffected individuals of the ET-31 family

		Tremor																										Archimedes		
		Clinical		Age of		Disease	Additional Clinical Findings		Resting		Postural		Kinetic		Intentional					Bradikinesia		Rigidity		Postural			Dominant	Spiral Test		
Sample ID	Sex	Severity	Age	Onset	Course		Hypomimia	Right	Left	Right	Left	Right	Left	Right	Left	Head	Chin	Voice	Right	Left	Right	Left	Instability	Freezing	Hand	Right	Left			
II-13	M	-	70	-	-	-	-	-	-	-	-	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	-	-			
III-1	F	-	53	-	-	Depression	-	-	-	-	-	-	-	-	-	No	No	No	-	-	+	-	-	No	Right	-	-			
III-2	M	severe	57	19	Stable	-	+	-	-	+++	+++	+++	+++	-	-	Yes	No	Yes	+	-	+	-	-	No	Right	+++	+++			
III-3	M	severe	75	40	-	-	-	-	-	++	++	++	++	-	-	No	No	No	-	-	-	-	-	No	Right	+++	+++			
III-4	F	severe	65	50	Progressive	Acathisia, Depression	-	+	-	++	++	++	++	-	-	Yes	No	Yes	-	-	-	-	-	No	Right	+++	+++			
III-8	F	severe	63	40	-	-	-	++	+++	++	+++	++	+++	+	+	Yes	No	Yes	+	-	++	++	-	No	Right	++	+++			
III-15	F	mild	51	?	-	Diabetes	-	-	-	+	+	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	+	+			
III-16	F	mild	49	20	-	-	-	-	-	+	+	-	-	-	-	No	No	No	-	-	-	+	-	No	Right	-	-			
IV-1	M	mild	34	18	-	-	-	-	-	-	+	-	-	-	-	No	No	No	+	-	-	-	-	No	Right	-	+			
IV-4	M	moderate	44	18	Progressive	-	-	-	-	+	++	-	+	-	-	No	No	Yes	-	-	-	-	?	No	Right	-	+			
IV-6	M	-	50	-	-	-	-	-	-	-	-	-	-	-	-	No	No	No	-	-	+	+	-	No	Right	-	-			
IV-7	F	-	43	-	-	-	-	-	-	-	-	-	-	-	-	No	No	No	-	-	+	-	+	No	Right	-	-			
IV-13	M	mild	33	?	-	-	-	-	-	+	+	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	+	+			
IV-16	M	mild	26	16	-	-	-	-	-	+	+	-	-	-	-	No	No	No	-	-	+	+	-	No	Right	-	+			
IV-17	F	mild	15	?	-	-	-	-	-	+	+	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	-	-			
V-1	F	-	21	-	-	-	-	-	-	-	-	-	-	-	-	No	No	No	-	-	-	-	-	No	Right	-	-			

Abbreviations: For tremor ‘-’ means no visible tremor, +1 is low amplitude; +2, tremor is moderate amplitude; +3, tremor is of large amplitude. For hypomimia, bradykinesia, rigidity, postural instability and Archimedes spiral test +: mild, ++: moderate, +++: severe.

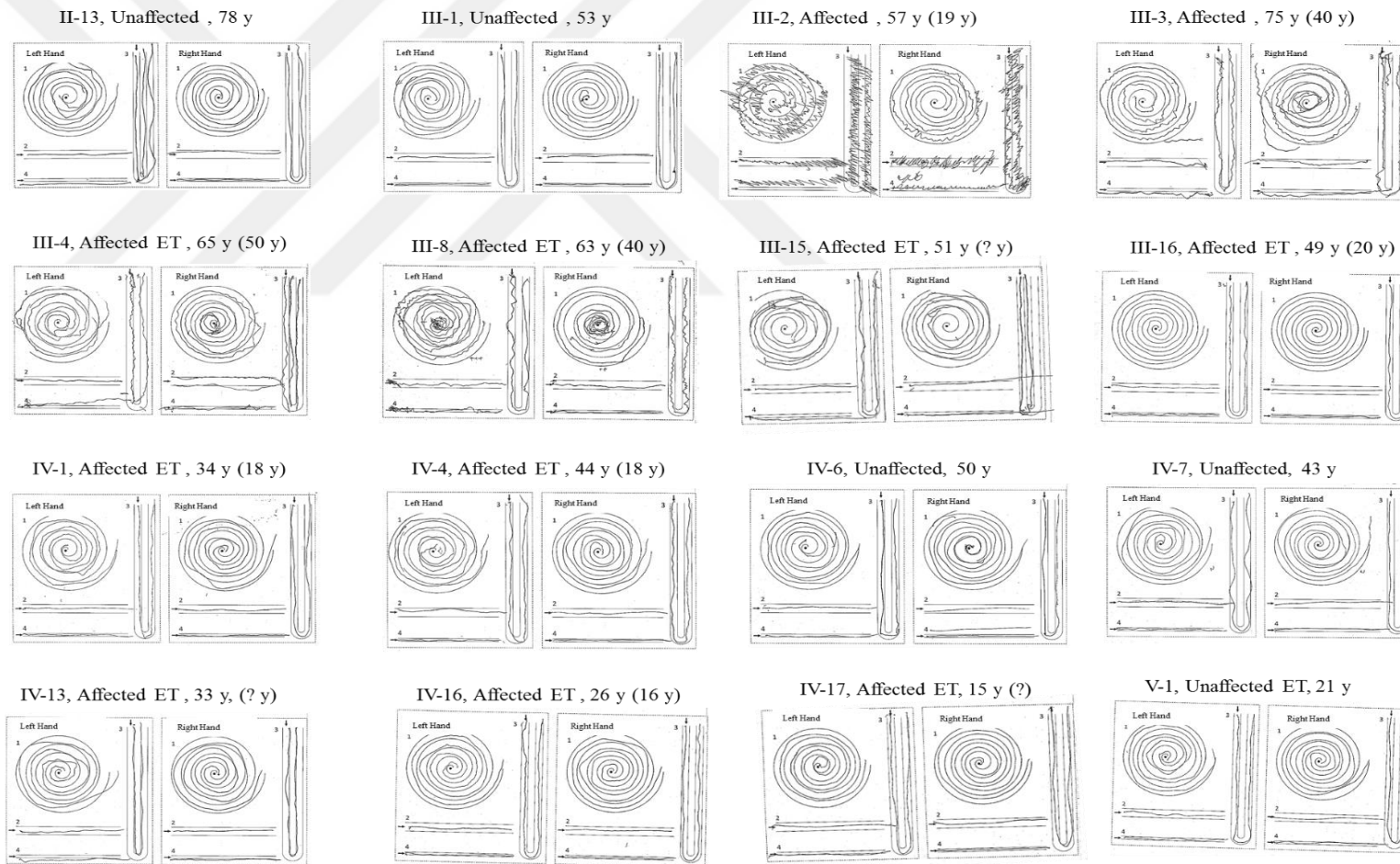


Figure 3.15 Archimedes spiral drawings of ET-31 family members, including affected and unaffected individuals.

3.2.2 Analysis of Whole Exome Sequencing Data

3.2.2.1 Preparation of Samples for WES

After clinical examinations by neurologists and the diagnosis of affected individuals with ET, blood samples were taken and were stored at -80 °C. At least two patients and one control were selected for WES to ensure sufficient representation from healthy and ET individuals. In order to effectively identify disease-causing variants, patients were selected from different generations and the control was selected as a patient's sibling. III-2, III-3 and III-4 were selected as affected individuals and II-13 was chosen as the control for WES. DNA samples of selected individuals were extracted with MN Nucleospin® Blood kit (Macherey-Nagel Inc., PA, USA) and concentrations of DNA samples were measured with a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc, DE, USA). In addition, A260/280 and 260/230 values were used for checking the purity of DNA (Table 3.10). All samples were run on an agarose gel and compared with a ladder (2-Log DNA Ladder (0.1-10.0 kb), NEB) to determine whether the DNA samples were degraded during storage. After concentration analyses and agarose gel electrophoresis, WES was performed on high-quality DNA samples (Figure 3.16).

Table 3.10 Concentrations and qualities of DNA samples from the ET-31 family

Sample ID	Nucleic Acid Conc.	Unit	260/280	260/230
II-13	121	ng/ μ L	1.93	2.16
III-2	110.9	ng/ μ L	1.93	2.06
III-3	125.7	ng/ μ L	1.9	2.21
III-4	118.5	ng/ μ L	1.91	2.14

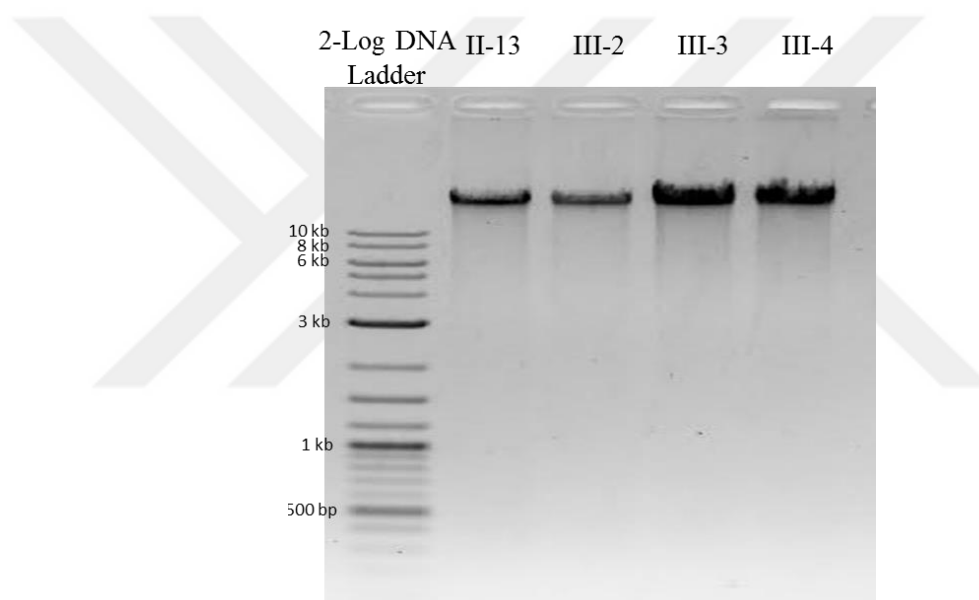


Figure 3.16 Density measurements using agarose gel electrophoresis. DNA samples were run on 1% agarose gel at 70 V for 50 minutes. Lane 1: 2-Log DNA Ladder Mix (NEB) (1 μ l); Lanes 2-5: DNA samples diluted 1:5 in TE (1 μ L). Gel image was captured with BioRad Gel Doc 2000 system. 2-Log DNA Ladder mix (10 kb: 40 ng, 3 kb: 120 ng) used as reference.

3.2.2.2 Whole Exome Sequencing Data Analysis

Sequences were obtained using Illumina HiSeq 2500 sequencer and the quality of sequences were determined with Illumina Real Time Analysis Software. Each fluorescence signal was converted to bases and FASTQ files containing raw data were prepared. Numbers of reads were calculated as nearly 50 million bases (Table 3.11) for the selected individuals. Using BWA software, fragments were aligned to the human reference genome hg19 and FASTQ files were converted to SAM files containing mapped sequences. PCR duplications were removed by Sequence Alignment/Map Tools (SAMtools) software package and sequence depths of exonic regions were obtained by BED tools. Percentages were calculated for mapped reads among total reads. SAMtools was used for separating PCR duplications and producing BAM files for clean sequences. The proportion of targeted exonic regions which were covered at least 5-fold was 98.16% for II-13, 97.76% for III-2, 97.70% for III-3 and 97.74% for III-4 (Table 3.11). SNP and Indel variants and base quality scores were determined with GATK and characterized by ANNOVAR software (ANNOVAR v.2016Feb01). Variants were positionally and functionally annotated with SnpEff and ANNOVAR tools with SnpEff and ANNOVAR.

Table 3.11 Statistics on the genome sequence produced by the paired-end whole exome sequencing for ET-31 family

Sample ID	Reads ^a	Mapped_Reads ^b	Mapped_% ^c	Removed_dup_% ^d	AVG_Coverage ^e	Covered_at_least_5x_% ^f
II-13	59,905,992	59,247,608	98.90%	1.93%	49.17419021	98.16%
III-2	52,883,072	52,279,552	98.86%	2.04%	44.00026393	97.76%
III-3	47,699,160	47,164,450	98.88%	1.63%	37.23520614	97.70%
III-4	55,896,532	55,315,299	98.96%	1.90%	45.87688925	97.74%

^aTotal number of reads assembled by WES. ^bNumber of sequencing reads mapped to human reference genome (hg19) with BWA. ^cPercentage of mapped reads among total reads. ^dPercentage of removed duplications with Samtools. ^eEffective fold coverage of the genome. ^fPercentage of the bases covered by at least one aligned read.

3.2.2.3 Variant Annotation and Effect Prediction for the ET-31 Family

SnpEff (snpEff_v4_3_core) analysis package was used for variant annotation and effect prediction against the human reference genome GRCh37.75 (hg19). This software was used to get more information about the identified variants, including effect annotations, estimation of putative impact, feature type and HGVS.c and HGVS.p values. Percentages of variant types based on their regions are illustrated in Figure 3.16. According to SnpEff analysis, approximately 790,360 (93.983%) – 913,364 (94.256%) of the identified variants were previously reported. Consequently, 50,599 (6.02%) – 55,659 (5.74%) unknown variants were present in the ET-31 family. The percentage of variants that were located on intron regions (2,138,739 variants for II-13) was higher than other variants that are found on exons (99,327 variants for II-13), intergenic regions (408,288 variants for II-13) and splice sites (6,993 variants for II-13) (Figure 3.17). After this step, high and moderate impact variants were selected for further analysis. Statistics of all variants depending on their genotype and effect types are illustrated in Table 3.12 and Table 3.13.

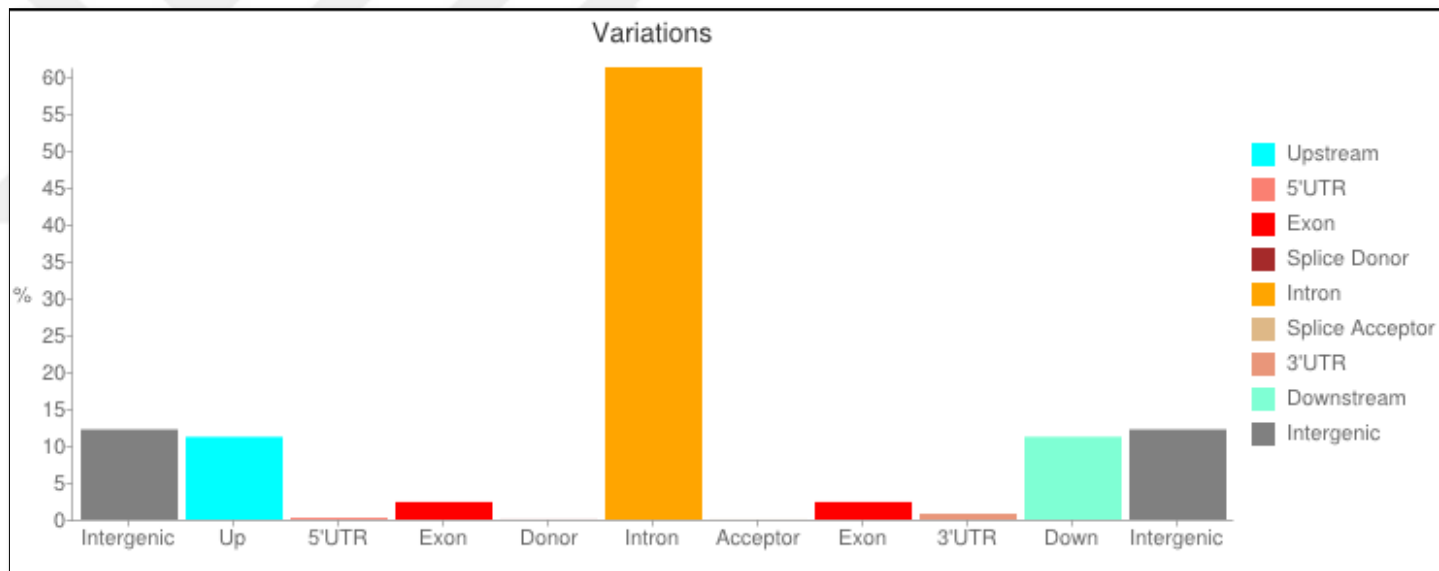


Figure 3.17 Percentage of variants with regards to genomic context for the ET-31 family after SnpEff analysis against the human database GRCh37.75 (hg19).

Table 3.12 Statistics on all variants based on their regions, identified through whole exome sequencing and SnpEff analysis for ET-31 family individuals.

Sample ID	Genome Total Length	Number of Variants	Number of Known Variants	Number of Unknown Variants	Exon	Intergenic	Intron	Splice site Region
II-13	37,322,280,791	969,023	913,364 (94.256%)	55,659 (5.74%)	88,327	408,288	2,138,739	6,993
III-2	37,322,280,794	867,697	815,408 (93.974%)	52,289 (6.03%)	81,783	363,301	1,917,818	6,354
III-3	37,322,280,791	840,959	790,360 (93.983%)	50,599 (6.02%)	81,676	348,954	1,878,583	6,325
III-4	37,322,280,791	917,974	864,132 (94.135%)	53,842 (5.86%)	85,258	386,422	2,020,034	6,793

Table 3.13 Statistics of all variants based on their annotation effects, identified through SnpEff analysis for the ET-31 family

Sample ID	Missense	Frameshift	Stop-lost	Nonsense	Splice Site Acceptor	Splice Site Donor
II-13	19,045	4,302	43	1,228	918	843
III-2	17,745	4,076	36	1,200	758	704
III-3	18,046	3,948	44	1,012	843	841
III-4	18,286	4,228	46	1,157	836	782

3.2.3 Functional Annotation of Variants for the ET-31 Family

SnpEff analysis was employed to predict the potential consequences of the identified variants, and the functional annotation of each variant was performed using a combination of ANNOVAR and several databases. Three different functional annotations were performed with ANNOVAR which are region-based, gene-based and filter-based annotations. Deleteriousnesses, evolutionary scores and population statistics (1000 Genome, ExAC, etc.) were calculated utilizing the databases described in 3.1.2. ClinVar and CADD annotations were also performed for the combined analysis of functional, deleteriousness and disease-causal variants through the integration of multiple annotation methods. Moreover, Online Mendelian Inheritance in Man (OMIM) database was used for determining the relationships between genotypes and phenotypes, and the Universal Protein Resource (UniProt) database was utilized to obtain information on protein function, subcellular location, induction and tissue specificity. The Mouse Genome Informatics (MGI) database was used to assess genetic, genomic, and biological data related with mouse genetics.

3.2.4 Identification of Disease-Causing Determinants

3.2.4.1 Filtration of variants for the ET-31 Family

Following annotation by SnpEff and ANNOVAR, we had the genomic coordinates, genomic regions, population statistics, deleteriousness and evolutionary scores associated with every variant we have identified. We filtered variants that had a significant effect on protein function by excluding modifier and low impact variants according to the results of SnpEff analysis.

We also filtered variants with minor allele frequencies higher than 0.02 based on the ExAC frequency database.

First filtration was carried out with filtering low and modifier variants after SnpEff analysis from VCF file. This enabled us to proceed with only protein coding variants including frameshift, missense, non-frameshift indel, stop-gained, stop-loss, start-gain, start-loss, splice site variants which are more potentially damaging than other annotation effects. After this filtration, number of variants was decreased from 5,905,992 variants to 1,003,179 variants. By ANNOVAR analysis, all variants were filtered depending on their allele frequency from ExAC database, and variants which have minor allele frequency (MAF) higher than 0.02 were removed. This filtration decreased the number of variants from 1,003,179 variants to 853,456 variants. We filtered variants that do not have protein altering effect and variant number is decreased from 853,456 to 24,197 variants. Finally, we removed variants that were found homozygous in control individuals and at the end we obtained 214 rare variants that had protein altering effect and allele frequency less than or equal to 0.02

3.2.4.2 Prioritization of Candidate Gene Variants for the ET-31 Family

Following SnpEff and ANNOVAR annotation and subsequent screening, we were left with only the variants that exhibited protein-altering effects and were very infrequent in the population. Following these steps, we further deprioritized variants with allele frequencies higher than 0.02 in 1000G and EVS databases.

We excluded variants which have been reported to be homozygous in over 20 individuals on the ExAC database. Moreover, we also eliminated homozygous alleles that were found in more than 10 individuals in our in-house database.

Based on damaging scores, we also deprioritized variants predicted to be non-damaging by all available analysis tools. Variants were labelled according to these criteria; Provean < -2.5 , SIFT $= < 0.05$, Polyphen2 HDIV; D: Probably damaging ≥ 0.957 , P:

possibly damaging $0.453 \leq \text{pp2hdiv} \leq 0.956$, Polyphen2 Hvar D: Probably damaging ≥ 0.909 , P: possibly damaging $0.447 \leq \text{pp2hvar} \leq 0.909$, MutationTaster prediction for A (disease causing automatic) and D (disease causing), MutationAssessor > 1.94 (M: Moderate and H: High), CADD ≥ 10 , GERP ≥ 3 , PhastCons_Vertebrate ≥ 0.90 and PhastCons_Mammalian ≥ 0.90 .

At first, we selected all variants that are homozygous for the proband and homozygous or presumed heterozygous in other affected individuals under the hypothesis of a recessive model (Table 3.14). We filtered variants that were found to be heterozygous or homozygous in control individuals. Based on this, we had 17 variants and 16 of them were deprioritized with respect to recessive model. Only one variant was included that was CHST15 c.1365_1366delCC, p.Pro455fs, frameshift variant which was also excluded then when we controlled it for other 16 ET families' sequenced individuals. We recognized that this variant was included in control groups of other ET families.

We reported other variants which were heterozygous for proband and affected individuals and wildtype for control. Our strong candidates after this elimination were IMPG1 p.Arg580Cys missense variant, p.Gln1167* ITGA10 stop gained variant, TPRXL p.Pro207Ser missense variant, IQCE p.Pro586Ser missense variant and SLC34A1 p.Arg215Gly missense variant (Table 3.15). Primers (Table A1) were designed with Primer3Plus according to these genomic coordinates and samples were prepared for segregation analysis.

Table 3.14 List of the homozygous damaging variants in proband of the ET-31 Family

Chr	Start	Ref	Alt	Annotation	Impact	Gene_Name	Transcript_BioType	HGVS.c	HGVS.p	ExAC	1000G	ESP	PRVN	SIFT	PPH2	MT	MA	CADD	GERP	PCONS
10	125780754	G	-	frameshift_variant	H	CHST15	protein_coding	c.1365_1366delCC	p.Pro455fs	.	.	.	.	.	.	.	.	12.96	0	0.983
3	195510613	C	T	missense_variant	M	MUC4	protein_coding	c.7838G>A	p.Ser2613Asn	0.0002	.	.	N	T	P	N	N	10.4	0	0.813
3	195510610	A	G	missense_variant	M	MUC4	protein_coding	c.7841T>C	p.Leu2614Pro	0.0036	.	.	N	T	P	N	N	14.26	0	0.452
3	195510611	G	T	missense_variant	M	MUC4	protein_coding	c.7840C>A	p.Leu2614Ile	0.0130	.	.	N	T	B	N	N	0.03	0	0.564
6	170871053	(CAG) ₂	-	disruptive_inframe_deletion	M	TBP	protein_coding	c.270_281delG(CAG) ₃ CA	p.Gln91_Gln94del	.	.	.	N	.	.	.	.	20.7	4.92	1
7	32795021	-	T	splice_acceptor_variant	H	DPY19L1P1	pseudogene	n.326-2dupA	.	.	0.7460	.	.	.	.	.	.	11.32	1.22	0.996
9	100616707	GCCGCC	-	inframe_insertion	M	FOXE1	protein_coding	c.535_537dupGCC	p.Ala179dup	.	.	.	N	.	.	.	.	14.23	3.48	1
12	7045892	(CAG) ₇	-	disruptive_inframe_insertion	M	ATN1	protein_coding	c.1503_1508dupGCAGCA	p.Gln501_Gln502dup	.	.	.	N	.	.	.	.	15.26	3.2	0.972
14	77493765	TGCTGC	-	inframe_deletion	M	IRF2BPL	protein_coding	c.366_371delGCAGCA	p.Gln123_Gln124del	.	.	.	N	.	.	.	.	9.77	2.12	0.998
15	20875300	A	G	splice_donor_variant	H	NBEAP1	retained_intron	n.282+2T>C	.	.	.	.	.	.	.	.	.	10.98	0	0.004
15	78913068	CAG	-	inframe_deletion	M	CHRNA3	protein_coding	c.67_69delCTG	p.Leu23del	0.0098	.	.	N	.	.	.	.	14.97	0	0.478
17	39432007	A	C	missense_variant	M	KRTAP9-7	protein_coding	c.58A>C	p.Lys20Gln	0.0160	0.4219	.	N	T	.	P	N	0.017	-7.55	0
17	39432017	C	T	missense_variant	M	KRTAP9-7	protein_coding	c.68C>T	p.Thr23Ile	0.0076	0.4399	.	N	T	.	P	N	1.213	0.77	0.019
17	40696233	C	G	missense_variant	M	NAGLU	protein_coding	c.2209C>G	p.Arg737Gly	0.0190	0.8592	0.85	N	D	B	P	N	16.19	4.01	0.926
17	61904728	G	A	splice_donor_variant	H	FTSJ3	protein_coding	c.-165+2C>T	.	.	0.7242	.	.	.	.	.	.	8.491	-7.84	0
21	45466387	-	T	splice_acceptor_variant	H	H2AFZP1	pseudogene	n.236-1dupA	.	.	.	.	.	.	.	.	.	5.893	0.149	0.998

Non-protein coding variants
 MAF> 0.02 variants
 High score for deleteriousness

Abbreviations used in this table; Chr: Chromosome, Ref: Reference, Alt: Alteration, HGVS.c: Human Genome Variation Society, cDNA, HGVS.p: Human Genome Variation Society, protein, ExAC: Exome Aggregation Consortium, 1000G: 1000 Genomes Project, ESP: Exome Sequencing Project, PRVN: PROVEAN, PPH2: PolyPhen2, MT: MutationTaster, MA: MutationAssessor, CADD: Combined Annotation Dependent Depletion, GERP: Genomic Evolutionarily Rate Profiling, PCONS: PhastCons

Table 3.15 List of the hoeterozygous damaging variants in proband of the ET-31 Family

Chr	Start	Ref	Alt	Annotation	Impact	Gene_Name	Transcript_BioType	HGVS.c	HGVS.p	ExAC	1000G	ESP	PRVN	SIFT	PPH2	MT	MA	CADD	GERP	PCONS
6	76660365	G	A	missense_variant	M	IMPG1	protein_coding	c.1738C>T	p.Arg580Cys	0.0002	.	0.0002	D	D	D	D	M	33	3.87	0.979
1	145542273	C	T	stop_gained	H	ITGA10	protein_coding	c.3499C>T	p.Gln1167*	0.0012	0.0012	0.0008	.	.	.	N	.	20.3	4.4	0.998
3	14106295	C	T	missense_variant	M	TPRXL	protein_coding	c.619C>T	p.Pro207Ser	0.0003	.	.	D	.	.	N	.	0.307	-0.261	0.923
7	2645522	C	T	missense_variant	M	IQCE	protein_coding	c.1756C>T	p.Pro586Ser	0.0041	0.0024	0.0038	D	T	P	N	L	15.07	1.42	0.011
5	176814873	C	G	missense_variant	M	SLC34A1	protein_coding	c.643C>G	p.Arg215Gly	.	.	.	D	D	P	D	M	25.7	2.08	0.916
1	152129115	T	C	missense_variant	M	RPTN	protein_coding	c.460A>G	p.Arg154Gly	0.0000	.	0.0022	D	D	P	N	M	9.666	-9.31	0
1	145303917	C	G	missense_variant	M	NBPF10	protein_coding	c.1314C>G	p.Asp438Glu	0.0054	0.0018	0.0064	N	T	B	N	L	1.823	-1.05	0
1	152129100	C	T	missense_variant	M	RPTN	protein_coding	c.475G>A	p.Gly159Ser	.	.	0.0001	N	T	B	N	L	0.196	-0.815	0
1	180910200	A	G	missense_variant	M	KIAA1614	protein_coding	c.2938A>G	p.Thr980Ala	0.0074	0.0062	0.0076	N	T	B	N	N	3.94	0.672	0
2	44445605	T	-	splice_acceptor_variant	H	PPM1B	protein_coding	c.1077-3dupT	.	0.0012	.	.	.	.	.	.	.	1.381	-11.9	0
2	120712719	-	T	splice_acceptor_variant	H	PTPN4	protein_coding	c.1814-3dupT	.	.	.	.	.	.	.	.	.	4.1	-5.82	0
2	160918804	A	C	splice_donor_variant	H	PLA2R1	protein_coding	c.109+2T>G	.	0.0110	.	.	.	.	.	D	.	14.28	3.9	0.999
2	90458434	A	G	missense_variant	M	CH17-132F21.1	protein_coding	c.110A>G	p.Asp37Gly	0.0087	.	.	.	.	B	.	.	7.294	-0.932	0.443
3	14106332	G	C	missense_variant	M	TPRXL	protein_coding	c.656G>C	p.Ser219Thr	0.0032	.	.	N	.	.	P	.	2.36	0.13	0.993
3	195507262	T	G	missense_variant	M	MUC4	protein_coding	c.11189A>C	p.His3730Pro	0.0035	.	.	N	T	B	P	N	0.12	0	0.961
7	147335934	-	T	frameshift_variant	H	CNTNAP2	protein_coding	c.46delT	p.Ser16fs	.	.	.	.	.	.	.	.	2.34	2.23	0.008
8	22459873	A	G	missense_variant	M	C8orf58	protein_coding	c.238A>G	p.Arg80Gly	.	.	.	N	D	.	N	.	6.47	1.03	0.009
9	140115394	T	G	missense_variant	M	RNF208	protein_coding	c.271A>C	p.Thr91Pro	0.0140	.	.	N	T	B	D	N	14.11	3.44	0.938
9	70912543	A	T	missense_variant	M	CBWD3	protein_coding	c.864A>T	p.Glu288Asp	0.0170	.	.	N	T	B	P	L	13.43	2.47	0.926
9	68433464	C	T	splice_donor_variant	H	RP11-764K9.4	Pseudogene	n.695+1G>A	.	.	.	.	.	.	.	.	.	23.4	2.01	1

Non-protein coding variants MAF> 0.02 variants High score for deleteriousness

Abbreviations used in this table; Chr: Chromosome, Ref: Reference, Alt: Alteration, HGVS.c: Human Genome Variation Society, cDNA, HGVS.p: Human Genome Variation Society, protein, ExAC: Exome Aggregation Consortium, 1000G: 1000 Genomes Project, ESP: Exome Sequencing Project, PRVN: PROVEAN, PPH2: PolyPhen2, MT: MutationTaster, MA: MutationAssesor, CADD: Combined Annotaion Dependent Depletion, GERP: Genomic Evolutionarily Rate Profiling, PCONS: PhastCons

3.2.5 Segregation Analysis of Candidate Variants for ET-31 Family

After the design of primers for candidate variants (including IMPG1 p.Arg580Cys, TPRXL p.Pro207Ser, IQCE p.Pro586Ser, SLC34A1 p.Arg215Gly missense variants and ITGA10 p.Gln1167* stop-gained variation) for the ET-31 family, PCR products of each individual were prepared for Sanger sequencing. According to Sanger sequencing results, the IMPG1 p.Arg580Cys missense variant co-segregated with in the ET-31 family (Figure 3.18, Figure 3.19). After genotyping IMPG1 p.Arg580Cys for all individuals from the ET-31 family that had blood samples, we recognized that 10 affected individuals were heterozygous, including 9 affected individuals and one unaffected family member.

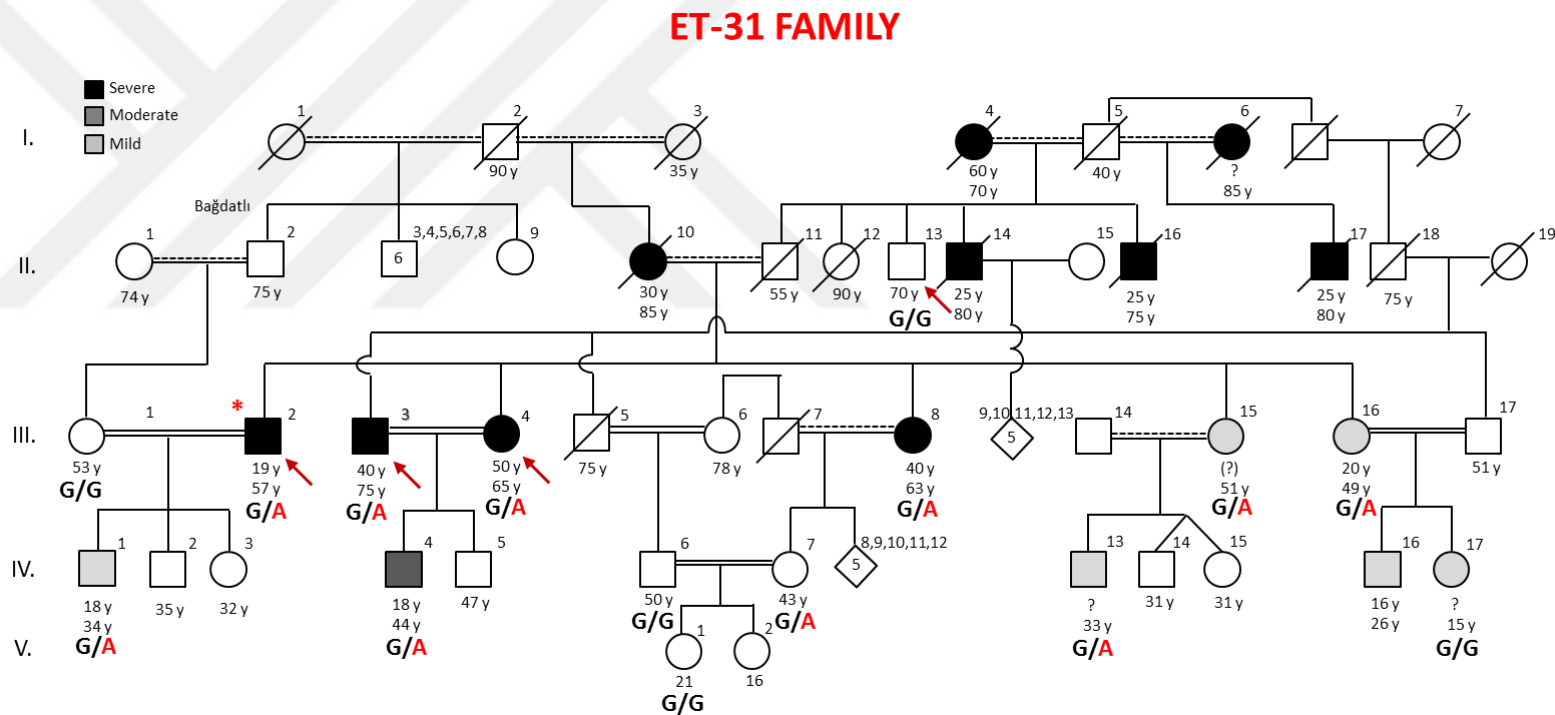


Figure 3.18 Pedigree of ET-31 family segregating essential tremor, with genotypes at IMPG1 p.Arg580Cys (c.1738 G>A). Red arrows indicate the individuals which were selected for whole exome sequencing (II-1, III-2 and IV-3). Proband (III-2) is shown with red asterisk. Individuals with ET are illustrated with filled dark gray if the clinical severity is moderate, light gray for mild clinical severity and black symbol is for severe ET. Genotypes for IMPG1 p.Arg580Cys, age at onset and current ages are shown under the symbols.

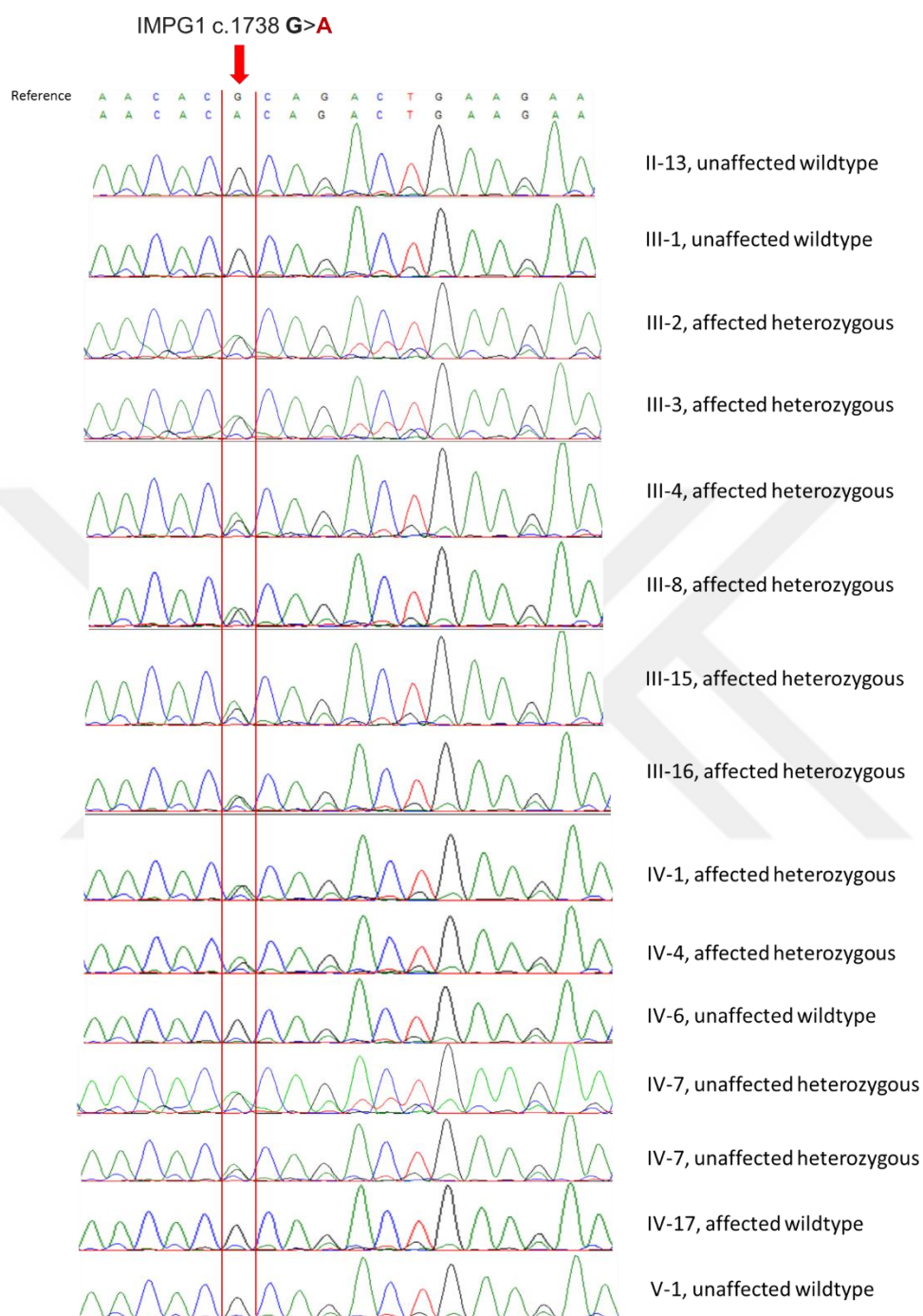


Figure 3.19 Segregation analysis result of ET-31 family. In affected individuals of ET-31 family IMPG1 p.Arg580Cys (c.1378 G>A) missense mutation segregated in heterozygous condition that is shown with two signals. G base (black signal) is converted to A base (green signal) in ET diagnosed patients of ET-31family.

3.2.5.1 IMPG1 p.Arg580Cys as the Disease Causing Mutation for a Consanguineous Family

The remaining variant at chr6, 76,660,365 (hg19; c.1378 G>A), results in an arginine (Arg, R) to cysteine (Cys, C) substitution at residue 580 and is located in exon 13 of IMPG1, the interphotoreceptor matrix proteoglycan 1 (IMPG1, IPM150, SPACR, ENST00000369950) gene. The mutation co-segregated with the disease under autosomal dominant inheritance pattern in the ET-31 family.

IMPG1 encodes a protein interphotoreceptor matrix proteoglycan 1 that is a major component of the retinal interphotoreceptor matrix. The encoded protein is a proteoglycan that is thought to participate in retinal adhesion and in maintaining photoreceptor viability. According to the NCBI conserved domain database, this protein has two SEA domains. SEA domain is found in the Sea urchin sperm protein, Enterokinase and Agrin (SEA), and has been proposed to function in regulating or binding carbohydrate side chains. Recently, proteolytic activity has been recorded in a SEA domain. The IMPG1 protein has another domain, DUF4214, that is of uncertain function and has been reported on a variety of distinct proteins, including transferases and allergen V5/Tpx-1 related proteins (Figure 3.20).

The p.Arg580Cys mutation is located in a SEA domain that is released from the cell surface and has been implicated in cleavage events. Multiple sequence alignment analysis using Clustal Omega illustrated that the arginine (R) allele is conserved in many species (Figure 3.21). In addition, the evolutionary conservation of the p.Arg580 allele was confirmed by *in silico* tools: the wildtype allele's PhastCon score is 0.991 (>0.99 is for high conservation), GERP score is 3.87 (>3 is strong evolutionary conservation) and CADD score is 33 (>10 has strong evolutionary conservation).

The ExAC database was also consulted to determine whether other individuals with the disease allele have been recorded. 22 heterozygous and no homozygous individuals are present for the IMPG1 p.Arg580Cys variation in this database, which contains data from 121,402 individuals. The population frequency this allele is 0.0001812 according to the ExAC database.



3.2.6 Modelling 3D structure of IMPG1 p.Arg580Cys

A model 3D structure of IMPG1 was generated using the Swiss-Pdb viewer (Deep view version 4.0). 3D structures of the wildtype and mutant proteins are illustrated in Figure 3.19. We observed that the variant alters the molecular surface and electrostatic interactions of the protein. In the wildtype, strong hydrogen bonds are present between ARG580 and GLU671, GLN667, LEU670 and HIS669, while hydrogen bonding instead occurs between CYS580 and SER632 and LEU670 in the mutant allele (Figure 3.22).

In IMPG1 p.Arg580Cys missense variation, arginine (which is positively charged) is changed to cysteine (which has polar uncharged side chains, including a thiol side chain), which may account for the differences in hydrogen bonding observed between the wildtype and mutant alleles.

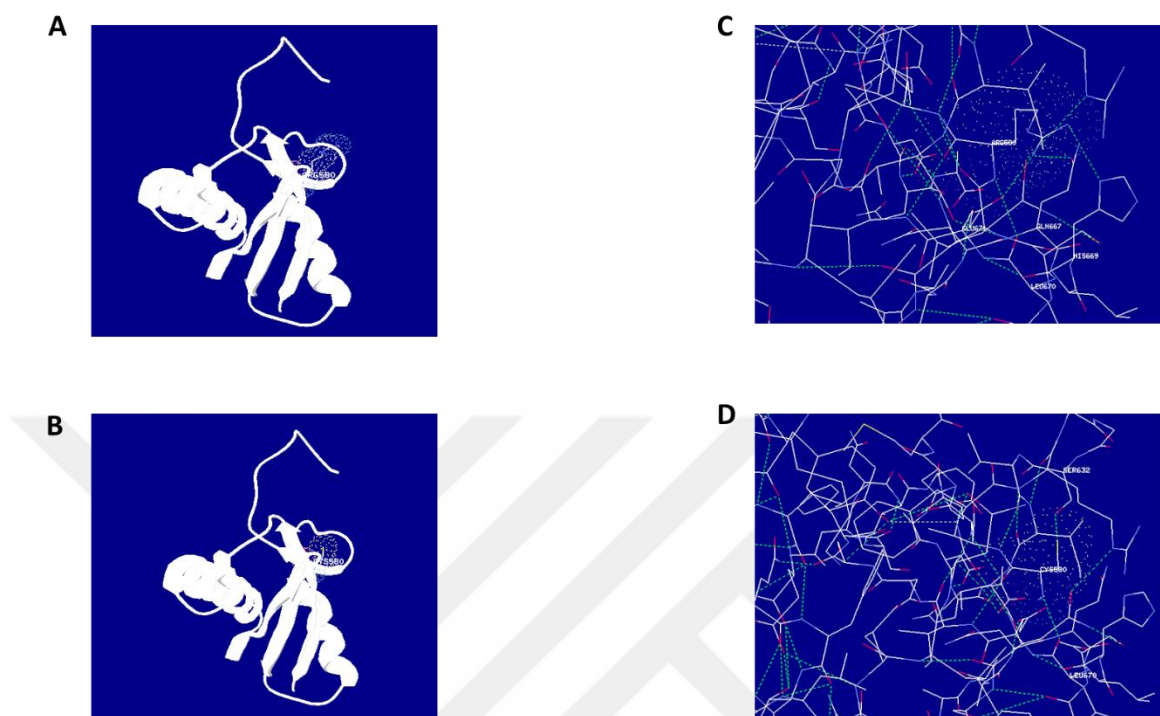


Figure 3.22 3D model protein structure of IMPG1 protein is represented using Swiss-Pdb viewer. (A) Predicted 3D structure for p.Arg580 wildtype IMPG1 protein. The amino acid 580 is indicated. (B) Predicted 3D structure for p.Cys580 mutant IMPG1 protein. (C) Wildtype IMPG1 protein structure showing specific bonds between amino acids (D) Mutant IMPG1 protein structure showing specific bonds between amino acids. Conversion of p.Arg580Cys resulted in formation of hydrogen bonds between CYS580 and SER632, LEU670 that are illustrated by arrows.

CHAPTER 4

Discussion

ET is the most common movement disorder in humans and is genetically and phenotypically a heterogeneous condition. Although significant efforts have been expanded for the characterization of ET, the genetic etiology of the disease remains largely unknown. In the present study, we performed WES to identify the causative mutations in Turkish families in whom ET segregates with autosomal dominant inheritance. We detected a homozygous missense mutation in one member of a family and a heterozygous missense mutation (c.1367G>A[p.Arg456Gln]) in MMP19 in all affected family members from two distinct families. Moreover, we identified a heterozygous missense mutation (c.1738C>T[p.Arg580Cys]) in IMPG1 in affected family members of a consanguineous family.

4.1 MMP19 p.Arg456Gln as the Disease Causing Mutation for Two Different Families with Essential Tremor

In the present study, we performed WES to identify the causative mutation in Turkish families in which ET segregates with autosomal dominant inheritance. We detected homozygous or heterozygous missense mutation (c.1367C>T [p.Arg456Gln]) in MMP19 in all affected members of two distinct families.

WES analyses of three members of the family ET-5, including two affected and one unaffected individuals from the family, generated an average of 50.33 Mb of aligned base reads per sample, of which 99.93% reads reached $\geq 40X$ coverage. A total of 85,856 exon related variants were identified from ET-5 individuals. WES analysis of five members of the family ETM-49, including four affected individuals and one

unaffected individual from the family, produced an average of 50.63 Mb of aligned base reads per sample, of which 99.17% reads reached $\geq 40\times$ coverage.

The pathogenic cause of the disorder was assumed to be the same mutation in a single gene. No deleterious variant in known ET genes was found in these families. All known variants (present in normal family members, the Exome Aggregation Consortium [ExAC] Browser, 1000G, EVS or an in-house Turkish exome database containing 190 individuals at a minor allele frequency [MAF] > 0.02) and non-deleterious SNVs determined by several commonly used *in silico* tools (PROVEAN, SIFT, PolyPhen-2, MutationTaster, MutationAssessor, GERP, CADD, and PhastCons) were removed. Finally, a mutation predicted to be deleterious in MMP19 (Matrix Metalloproteinase 19 [MIM: 601807]), encoding a zinc-binding endopeptidase that degrades various components of the extracellular matrix, was prioritized for further analysis.

Utilizing Sanger sequencing with specific primers (Table A.1), we confirmed that the heterozygous mutation (c.1367C>T) in MMP19 (Transcript ID: ENST00000372330) segregates with the affected individuals, but not the unaffected family members (chromatogram of family members of ET-5 and ET-49 families are illustrated in Figure 3.7 and Figure 3.9). The missense mutation altered the wildtype in codon 456, replacing the wild type arginine with a glutamine residue (p.Arg456Gln) in the hemopexin domain and therefore eliminating the positive charge of the side chain. Multiple-sequence alignment of MMP19 demonstrated that the residue 456Arg is evolutionarily conserved in all of the compared species except *Danio rerio* (Figure 3.10). Moreover, the c.1367C>T (p.Arg456Gln) variation was predicted to be disease-causing by several major algorithms, including Polyphen (score: 0.987), MutationTaster (score: 1), CADD (score: 23.4), GERP (score: 3.88) and PhastCons (score: 0.999). This mutation is also

present in ExAC Browser (3/ 121,358 alleles with MAF = 0.003692), the 1000 genome database (with MAF = 0.0014), and the ESP database (with MAF = 0.0035).

MMP19 is found at chromosome 12q13.2 at location 56229214-56236767. The DNA sequence of this gene includes nine exons and eight introns, spanning 7.55 kb. The MMP19 gene encodes an externally secreted matrix metalloproteinase and contains a signal peptide in order to localize to secretory vesicles. As with most other secreted MMPs, MMP19 can be translated and secreted as a catalytic inactive protein or *zymogens*¹⁸⁵. Zymogens are required to be activated through the proteolytic cleavage of the pro-domain by other extracellular matrix (ECM) proteinases. MMP19 is also a zinc-dependent endopeptidase, as its catalytic center includes an active site for zinc binding and cannot perform its function without this cation. Substrate recognition, meanwhile, is determined by a hemopexin-like domain¹⁸⁵. The catalytic activities of MMP19 are regulated by tissue inhibitors of metalloproteinases (TIMPs). The active form of MMP19 is strongly inhibited by TIMP-2, TIMP-3 and TIMP-4, but less efficiently by TIMP-1¹⁸⁶. MMP19 belongs to a family of proteases known for their ability to degrade the extra cellular matrix. Substrates for MMP19 include gelatin, aggrecan, collagen type IV, fibronectin and large tenascin-C. In addition, MMP19 also has a role in regulating the insulin-like growth factor¹⁸⁷. Recently, Sadowski *et al.* reported that MMP19 is also involved in the processing of laminin, a major vascular brain matrix protein¹⁸⁸. MMP19 was found to be expressed in several tissues, including nasopharyngeal epithelial cells, lung, breast, placenta, skin and intestine^{189–191}.

MMPs are well-recognized to play fundamental roles in the biology and pathology of the nervous system¹⁹². The basic biology and roles of MMPs in the central nervous system range from neurogenesis, axonal guidance, learning and memory, brain development to synaptic plasticity and repairing mechanisms^{193–196}. Recent studies

emphasize their function in distinct brain disorders such as Alzheimer's disease, multiple sclerosis, Parkinson's disease and ischemia/reperfusion^{197–201}.

Accordingly, MMP19 is also important for the nervous system. MMP19 has been reported to be significant for multiple sclerosis lesions, as demonstrated by the immunohistochemistry in different multiple sclerosis lesion stages and the analysis of MMP19 expression in primary microglia and astrocyte cell cultures²⁰². Another study has shown that MMP19 is highly expressed in growing astrocytoma/glioma cells and that its expression may help their invasion through brain extracellular matrix components²⁰³. Fudge *et al.* confirmed the expression of MMP19 protein in the brain extracellular matrix of adult sensory neuron populations cultured on a laminin substrate, and provided new information for manipulating neuronal extracellular matrix components to enhance axonal regeneration²⁰⁴.

The mutation (c.1367C>T, p.Arg456Gln) we have identified is located on the hemopexin-like domain of MMP19, which is important for substrate recognition. This region may be functionally disrupted and thereby inhibit TIMP binding. While functional characterization of the protein is necessary to determine whether this is indeed the mechanism of pathogenesis, it is nevertheless evident that MMP19 is important for the central nervous system and that the disruption of its activity may be a causative factor for ET.

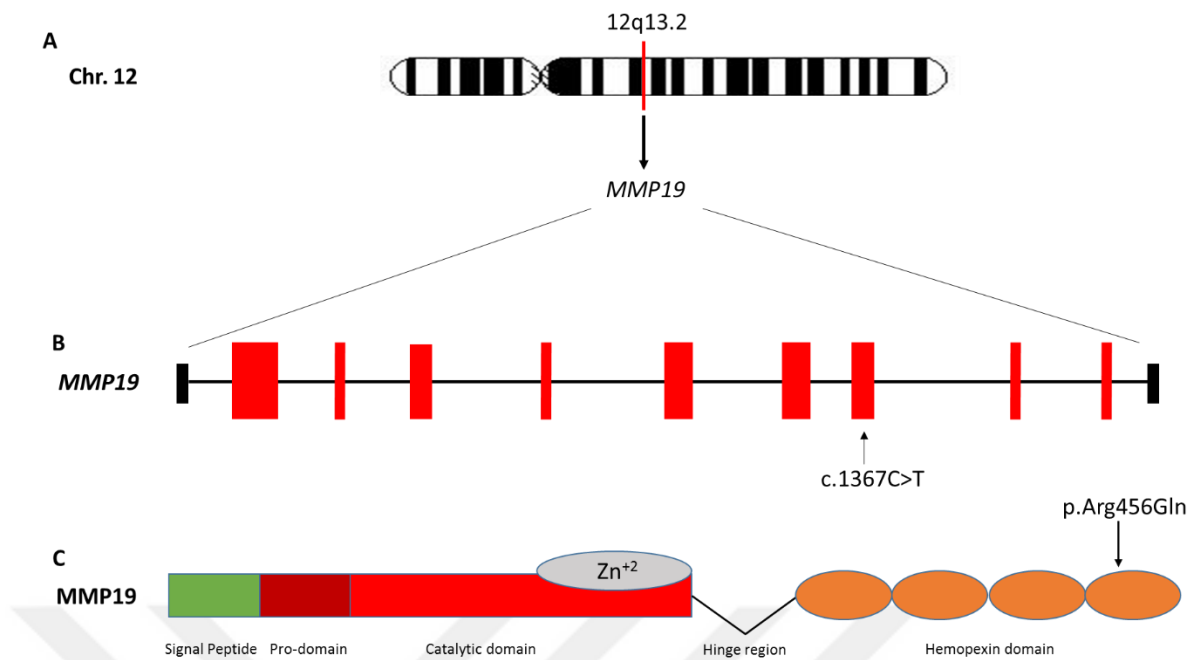


Figure 4.1 Chromosomal localization of *MMP19*, protein domain organization of *MMP19* and the position of the detected mutation. (A) Chromosomal localization of *MMP19* within the region of 12q13.2. (B) Schematic representation of the exon-intron structure of *MMP19*. The location of the c.1367 C > T mutation identified in this study is on exon 7. (C) The domain structure of the *MMP19* protein, including the signal peptide, the pro-domain and catalytic domain containing zinc ion binding site, hinge region and hemopexin domain. The p.Arg456Gln missense alteration is in the hemopexin domain.

4.2 IMPG1 p.Arg580Cys as the Disease-Causing Mutation for a Consanguineous Family with Essential Tremor

WES of three severely affected and one unaffected relatives in family ET-31 revealed IMPG1 p.Arg580Cys as the only potentially damaging allele co-segregating with essential tremor. Heterozygosity for this allele is associated with a variable age of onset for essential tremor, suggesting that the variant's effects could be further modulated through genetic or environmental factors or stochastic modifiers²⁰⁵.

According to segregation analysis, all affected individuals are heterozygous for the p.Arg580Cys variant except IV-7, who was found not to exhibit ET based on clinical assessment (Figure 3.15). This individual is 43 years old age and we observed that the age at onset for the ET-31 family varies between 25 to 50 years. The heterozygous IV-7 individual may develop ET at an older age. Based on clinical assessments, she had right hand rigidity and postural instability, which are potential forerunners for the development of ET.

The IMPG1 p.Arg580Cys missense variant is located in a protein-coding region. The substitution p.Arg580Cys was predicted to be damaging by all *in silico* tools; including PROVEAN (score -7.78), SIFT (P = 0), PolyPhen2 (score 0.996), MutationAssesor (score 2.455) and MutationTaster (score 1) (Table 3.15). The p.Arg580Cys variant was reported at an allele frequency of 0.0001812 in ExAC, with records from 22 heterozygous and 0 homozygous individuals. Moreover, this variant is not reported in the 1000G database and reported in the ESP database allele at a frequency of 0.0002. This mutation lies in the C-terminal region of the encoded protein, corresponding to a functional SEA domain that plays a role in regulating and binding carbohydrate side chains. Multiple sequence alignment between species illustrated that the p.Arg580

region is highly conserved. 3D model structure analysis of p.Arg580Cys also showed that the arginine-to-cysteine substitution on the 580th amino acid resulted in the disruption of some hydrogen bonds between amino acids and production of new strong bonds between other amino acid residues. These results support to the idea that IMPG1 p.Arg580Cys is a highly damaging, conserved and rare variant that co-segregates with essential tremor in the ET-31 family.

Genomic organization of *IMPG1* maps the gene locus to chromosome 6q14.1, which includes 17 coding exons (Figure 4.2). *IMPG1* (Interphotoreceptor matrix proteoglycan 1) encodes SPACR, a 150 kDa sialoprotein that is closely related to photoreceptor cone and rod proteins, where photoreceptors and the retinal pigment epithelium (RPE) are in close collocation²⁰⁶. *IMPG1* corresponds to a 797 amino acid secreted glycoprotein and contains a central mucin-like domain and an EGF-like domain in the C-terminal region (Figure 4.2). SPACR also includes two SEA domains (after sea urchin sperm protein, enterokinase and agrin), a highly conserved structure that contains four anti-parallel β sheets surrounded by α helices in an O-linked glycosylation-rich region²⁰⁷. The C-terminal SEA2 domain of SPACR is known to play a role in the cleavage of peptides, including the autoproteolytic GSVIV sequence that is the cleavage site in mucin 1 and dystroglycan and produces two polypeptides²⁰⁸. However, the N-terminal SEA1 of the SPACR protein does not include a cleavage site. Besides the protein-cleaving function of SEA domains, they are also involved in adaptation of extracellular matrices to specific physical constraints²⁰⁷.

As we mentioned above, the interphotoreceptor matrix (IPM) encoded by the *IMPG1* gene is rich in insoluble glycoconjugates, which are mainly chondroitin sulfate proteoglycans and known to facilitate connections between various cell types, including RPE, photoreceptors, glial Müller cells and choriocapillaris²⁰⁹.

The BioGPS gene portal contains tissue-specific patterns of mRNA expression for 79 human and 61 mouse genes, including IMPG1. This database suggests that the mRNA of the Arg580Cys variant is highly expressed in the retina, cerebellum, spinal cord and brain²¹⁰.

Although the expressional, functional and genetic characterization of IMPG1 is relatively limited, certain variants of *IMPG1* have been found to be related with retinal disorders. Functional characterization and expression patterning of the *IMPG1* gene can be performed through immunocytochemistry and real time PCR experiments in the future. Further functional studies can also be conducted in order to reveal the effects of the p.Arg580Cys missense mutation in IMPG1 in the context of ET.

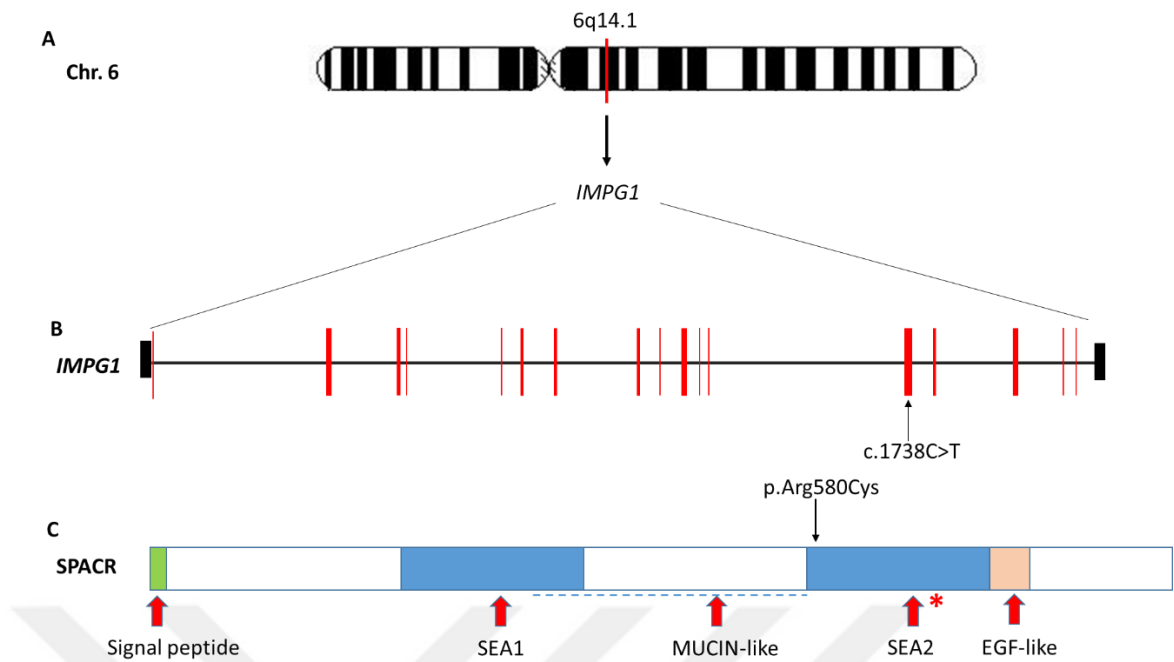


Figure 4.2 Chromosomal localization of *IMPG1*, protein domain organization of SPACR and position of detected mutation. (A) Chromosomal localization of *IMPG1* within the region of 6q14.1. (B) Schematic representation of the exon-intron structure of *IMPG1*. The location of the c.1738 C > T mutation on exon 7 found in this study. (C) The domain structure of its encoded SPACR protein, involving the signal peptide, the N terminal SEA1 and C-terminal SEA2 domains, the mucin-like domain with dotted line, autoproteolytic site that is shown with asterisk (*), and the EGF-like domain. The missense change found is indicated as p.Arg580cys in the C-terminal SEA2 domain.

CHAPTER 5

Conclusion and Future Perspectives

ET is the most common movement disorder in humans. Despite its high heritability and frequency, the genetic basis and pathophysiology of ET is not well understood. A recent report from our group showed that the mitochondrial serine protease HTRA p.G399S is related with ET and Parkinson's disease. To identify new genes related with ET, we performed WES in three different families and identified two new mutations related with ET.

We examined the same mutation in two distinct families found to segregate with autosomal dominant ET across four to five generations. We carried out WES of genomic DNA from a total of seven affected and two unaffected family members and performed pedigree analysis to determine the genetic basis of ET. WES of affected and unaffected relatives of family ET-5 and family ET-49 resulted in the identification of MMP19 p.Arg456Gln as the only damaging allele cosegregating with ET. One individual in the consanguineous family ET-49 (IV-2) was found to be homozygous for this allele and expressed a more severe phenotype of ET. All other affected individuals were heterozygous and control family members were wildtype for this allele. The age at onset was variable between individuals, possibly because ET is highly heterogeneous and is influenced strongly by both genetic and environmental factors.

MMP19 encodes a matrix metallopeptidase that is essential for the brain extracellular matrix. These metallopeptidases play a role in regulating the brain extracellular matrix through the cleavage of several substrates, including gelatin, collagen V and laminin. Our results suggest that the mutation of interest is located on the hemopexin domain of

MMP19, which is important for TIMP binding. TIMPs inhibit the peptide-cleaving activity of MMP19, suggesting that decreased proteolytic activity may be responsible for the onset of ET. We intend to assess the effect of this mutation on the enzymatic activity of MMP19 by comparing the gelatinase activity of wildtype and mutant variants in a zymography assay. We will also examine the expression pattern of MMP19 using *in situ* hybridization, and analyze the mRNA expression patterns of this protein in different tissues through RT-PCR.

We identified another important mutation in a different consanguineous family (ET-31) using whole exome sequencing. In this five-generation family, WES was carried out with three affected and one unaffected family members. WES of affected and unaffected relatives of family ET-31 showed IMPG1 p.Arg580Cys as the only damaging allele cosegregating with autosomal dominant essential tremor. All affected individuals were heterozygous and all unaffected individuals were wildtype for this allele. IMPG1 encodes an interphotoreceptor matrix protein, which is important for regulating carbohydrates. However, the precise function of IMPG1 is not well known. As such, we aim to evaluate the function of this protein in the central nervous system through *in situ* hybridization in mouse brain tissues, and will support our results through the RT-PCR analysis of IMPG1 expression in different tissues.

Our results have led to the discovery of two different mutations that are related with familial ET. ET is a genetically heterogeneous disease, and more than one gene may affect the pathophysiology of the condition. The MMP19 and IMPG1 mutations identified in the present study are potentially damaging mutations and functional assays may further elucidate their role in the central nervous system in the future.

BIBLIOGRAPHY

1. Louis, E. D. Essential Tremor. *Arch. Neurol.* 57, (2000).
2. Benito-León, J. & Louis, E. D. Essential tremor: emerging views of a common disorder. doi:10.1038/ncpneuro0347
3. Louis, E. D., Ottman, R. & Allen Hauser, W. How common is the most common adult movement disorder? Estimates of the prevalence of essential tremor throughout the world. *Mov. Disord.* 13, 5–10 (1998).
4. Lan, E. & Ouis, D. L. Clinical Practice Essential Tremor. *N Engl J Med* 345, (2001).
5. Deng, H., Le, W. & Jankovic, J. Genetics of essential tremor. doi:10.1093/brain/awm018
6. Jankovic, J. Essential tremor: A heterogenous disorder. *Mov. Disord.* 17, 638–644 (2002).
7. Louis, E. D. Non-motor symptoms in essential tremor: A review of the current data and state of the field. *Parkinsonism Relat. Disord.* 22 Suppl 1, S115–8 (2016).
8. Louis, E. D. Essential tremor. *Lancet Neurol.* 4, 100–110 (2005).
9. Elble, R. J. Diagnostic criteria for essential tremor and differential diagnosis. *Neurology* 54, S2–6 (2000).
10. Whaley, N. R., Putzke, J. D., Baba, Y., Wszolek, Z. K. & Uitti, R. J. Essential tremor: Phenotypic expression in a clinical cohort. *Parkinsonism Relat. Disord.* 13, 333–339 (2007).
11. Benito-León, J. & Louis, E. D. in *Hyperkinetic Movement Disorders* 71–94 (Wiley-Blackwell, 2012). doi:10.1002/9781444346183.ch6
12. Deuschl, G., Bain, P. & Brin, M. Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. *Mov. Disord.* 13 Suppl 3, 2–23 (1998).
13. Louis, E. D. *et al.* Neuropathologic findings in essential tremor. *Neurology* 66,

- 1756–9 (2006).
14. Louis, E. D. *et al.* Essential tremor associated with pathologic changes in the cerebellum. *Arch. Neurol.* 63, 1189–93 (2006).
 15. Louis, E. D. *et al.* *Metabolic abnormality in the cerebellum in patients with essential tremor: a proton magnetic resonance spectroscopic imaging study.* *Neuroscience Letters* 333, (2002).
 16. Sharifi, S., Nederveen, A. J., Booij, J. & van Rootselaar, A.-F. Neuroimaging essentials in essential tremor: a systematic review. *NeuroImage. Clin.* 5, 217–31 (2014).
 17. Louis, E. D. *et al.* Essential tremor associated with focal nonnigral Lewy bodies: a clinicopathologic study. *Arch. Neurol.* 62, 1004–7 (2005).
 18. Huber, S. J. & Paulson, G. W. Efficacy of alprazolam for essential tremor. *Neurology* 38, 241–3 (1988).
 19. Louis, E. D. A new twist for stopping the shakes? Revisiting GABAergic therapy for essential tremor. *Arch. Neurol.* 56, 807–8 (1999).
 20. Busenbark, K. *et al.* Accuracy of reported family histories of essential tremor. *Neurology* 47, 264–265 (1996).
 21. Kühlenbäumer, G., Hopfner, F. & Deuschl, G. Genetics of essential tremor: meta-analysis and review. *Neurology* 82, 1000–7 (2014).
 22. Benito-León, J., Bermejo-Pareja, F., Morales, J.-M., Vega, S. & Molina, J.-A. Prevalence of essential tremor in three elderly populations of central Spain. *Mov. Disord.* 18, 389–94 (2003).
 23. Louis, E. D., Thawani, S. P. & Andrews, H. F. Prevalence of essential tremor in a multiethnic, community-based study in northern Manhattan, New York, N.Y. *Neuroepidemiology* 32, 208–14 (2009).
 24. Ma, S. *et al.* Familial essential tremor with apparent autosomal dominant inheritance: should we also consider other inheritance modes? *Mov. Disord.* 21, 1368–74 (2006).
 25. Tanner, C. M. *et al.* Essential tremor in twins: an assessment of genetic vs

- environmental determinants of etiology. *Neurology* 57, 1389–91 (2001).
26. Lorenz, D. *et al.* High concordance for essential tremor in monozygotic twins of old age. *Neurology* 62, 208–211 (2004).
 27. Louis, E. D. & Zheng, W. Beta-carboline alkaloids and essential tremor: exploring the environmental determinants of one of the most prevalent neurological diseases. *ScientificWorldJournal*. 10, 1783–94 (2010).
 28. Louis, E. D. *et al.* Interaction between blood lead concentration and delta-amino-levulinic acid dehydratase gene polymorphisms increases the odds of essential tremor. *Mov. Disord.* 20, 1170–7 (2005).
 29. Louis, E. D., Zheng, W., Applegate, L., Shi, L. & Factor-Litvak, P. Blood harmane concentrations and dietary protein consumption in essential tremor. *Neurology* 65, 391–6 (2005).
 30. Louis, E. D. *et al.* Association between essential tremor and blood lead concentration. *Environ. Health Perspect.* 111, 1707–11 (2003).
 31. Klaver, C. C. *et al.* Genetic risk of age-related maculopathy. Population-based familial aggregation study. *Arch. Ophthalmol. (Chicago, Ill. 1960)* 116, 1646–51 (1998).
 32. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* 63, 168–74 (2006).
 33. Freitag, C. M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* 12, 2–22 (2007).
 34. Smoller, J. W. & Finn, C. T. Family, twin, and adoption studies of bipolar disorder. *Am. J. Med. Genet. C. Semin. Med. Genet.* 123C, 48–58 (2003).
 35. Czene, K., Lichtenstein, P. & Hemminki, K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int. J. cancer* 99, 260–6 (2002).
 36. Nisticò, L. *et al.* Concordance, disease progression, and heritability of coeliac disease in Italian twins. *Gut* 55, 803–8 (2006).
 37. Ingebrigtsen, T. *et al.* Genetic influences on Chronic Obstructive Pulmonary

- Disease - a twin study. *Respir. Med.* 104, 1890–5 (2010).
38. Fischer, M. *et al.* Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. *Circulation* 111, 855–62 (2005).
 39. Tysk, C., Lindberg, E., Järnerot, G. & Flodérus-Myrhed, B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 29, 990–6 (1988).
 40. Kjeldsen, M. J., Kyvik, K. O., Christensen, K. & Friis, M. L. Genetic and environmental factors in epilepsy: a population-based study of 11900 Danish twin pairs. *Epilepsy Res.* 44, 167–78 (2001).
 41. Katzmarzyk, P. T. *et al.* Familial resemblance for coronary heart disease risk: the HERITAGE Family Study. *Ethn. Dis.* 10, 138–47
 42. Agarwal, A., Williams, G. H. & Fisher, N. D. L. Genetics of human hypertension. *Trends Endocrinol. Metab.* 16, 127–33 (2005).
 43. Walley, A. J., Blakemore, A. I. F. & Froguel, P. Genetics of obesity and the prediction of risk for health. *Hum. Mol. Genet.* R124–30 (2006). doi:10.1093/hmg/ddl215
 44. Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J. & Hart, D. Genetic influences on osteoarthritis in women: a twin study. *BMJ* 312, 940–3 (1996).
 45. Do, C. B. *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* 7, e1002141 (2011).
 46. Vink, J. M., Sadrzadeh, S., Lambalk, C. B. & Boomsma, D. I. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J. Clin. Endocrinol. Metab.* 91, 2100–4 (2006).
 47. Harney, S. M. J. *et al.* Fine mapping of the MHC Class III region demonstrates association of AIF1 and rheumatoid arthritis. *Rheumatology (Oxford)*. 47, 1761–7 (2008).
 48. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait:

- evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–92 (2003).
49. Bak, S., Gaist, D., Sindrup, S. H., Skytthe, A. & Christensen, K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke*. 33, 769–74 (2002).
 50. Gamazon, E. R. *et al.* The Convergence of eQTL Mapping, Heritability Estimation and Polygenic Modeling: Emerging Spectrum of Risk Variation in Bipolar Disorder. (2013). at <<http://arxiv.org/abs/1303.6227>>
 51. Ghazarian, L. *et al.* Protection against type 1 diabetes upon Cocksackievirus B4 infection and iNKT-cell stimulation: role of suppressive macrophages. *Diabetes* 62, 3785–96 (2013).
 52. Poulsen, P., Kyvik, K. O., Vaag, A. & Beck-Nielsen, H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia* 42, 139–45 (1999).
 53. Gulcher, J. R. *et al.* Mapping of a familial essential tremor gene, FET1, to chromosome 3q13. *Nat. Genet.* 17, 84–7 (1997).
 54. Higgins, J. J., Pho, L. T. & Nee, L. E. A gene (ETM) for essential tremor maps to chromosome 2p22-p25. *Mov. Disord.* 12, 859–64 (1997).
 55. Shatunov, A. *et al.* Genomewide scans in North American families reveal genetic linkage of essential tremor to a region on chromosome 6p23. *Brain* 129, 2318–31 (2006).
 56. Jeanneteau, F. *et al.* A functional variant of the dopamine D3 receptor is associated with risk and age-at-onset of essential tremor. *Proc. Natl. Acad. Sci. U. S. A.* 103, 10753–8 (2006).
 57. Beom, S., Cheong, D., Torres, G., Caron, M. G. & Kim, K.-M. Comparative studies of molecular mechanisms of dopamine D2 and D3 receptors for the activation of extracellular signal-regulated kinase. *J. Biol. Chem.* 279, 28304–14 (2004).
 58. Tan, E.-K. *et al.* DRD3 variant and risk of essential tremor. *Neurology* 68, 790–791 (2007).

59. Vitale, C. *et al.* DRD3 Ser9Gly variant is not associated with essential tremor in a series of Italian patients. *Eur. J. Neurol.* 15, 985–7 (2008).
60. Higgins, J. J., Loveless, J. M., Jankovic, J. & Patel, P. I. Evidence that a gene for essential tremor maps to chromosome 2p in four families. *Mov. Disord.* 13, 972–7 (1998).
61. Higgins, J. J. *et al.* HS1-BP3 gene variant is common in familial essential tremor. *Mov. Disord.* 21, 306–9 (2006).
62. Higgins, J. J. *et al.* A variant in the HS1-BP3 gene is associated with familial essential tremor. *Neurology* 64, 417–21 (2005).
63. Deng, H. *et al.* Extended study of A265G variant of HS1BP3 in essential tremor and Parkinson disease. *Neurology* 65, 651–2 (2005).
64. Zahorakova, D. *et al.* No association with the ETM2 locus in Czech patients with familial essential tremor. *Neuro Endocrinol. Lett.* 31, 549–52 (2010).
65. Novelletto, A. *et al.* Linkage exclusion in Italian families with hereditary essential tremor. *Eur. J. Neurol.* 18, e118–20 (2011).
66. Aridon, P. *et al.* Further evidence of genetic heterogeneity in familial essential tremor. *Parkinsonism Relat. Disord.* 14, 15–18 (2008).
67. Stefansson, H. *et al.* Variant in the sequence of the LINGO1 gene confers risk of essential tremor. doi:10.1038/ng.299
68. Deng, H., Gu, S. & Jankovic, J. LINGO1 variants in essential tremor and Parkinson's disease. *Acta Neurol. Scand.* 125, 1–7 (2012).
69. Zhou, Z.-D., Sathiyamoorthy, S. & Tan, E.-K. LINGO-1 and Neurodegeneration: Pathophysiologic Clues for Essential Tremor. *Tremor Other Hyperkinet. Mov. (N. Y.)* 2, (2012).
70. Thier, S. *et al.* Polymorphisms in the glial glutamate transporter SLC1A2 are associated with essential tremor. *Neurology* 79, 243–8 (2012).
71. Merner, N. D. *et al.* Exome Sequencing Identifies FUS Mutations as a Cause of Essential Tremor. *Am. J. Hum. Genet.* 91, 313–319 (2012).
72. Wu, Y.-R. *et al.* Identification of a novel risk variant in the FUS gene in essential

- tremor. *Neurology* 81, 541–4 (2013).
73. Rajput, A. *et al.* Identification of FUS p.R377W in essential tremor. *Eur. J. Neurol.* 21, 361–363 (2014).
 74. Rajput, A. *et al.* VPS35 and DNAJC13 disease-causing variants in essential tremor. *Eur. J. Hum. Genet.* 23, 887–8 (2015).
 75. Unal Gulsuner, H. *et al.* Mitochondrial serine protease HTRA2 p.G399S in a kindred with essential tremor and Parkinson disease. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18285–90 (2014).
 76. Hegde, R. *et al.* Identification of Omi/HtrA2 as a mitochondrial apoptotic serine protease that disrupts inhibitor of apoptosis protein-caspase interaction. *J. Biol. Chem.* 277, 432–8 (2002).
 77. Jones, J. M. *et al.* Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of mnd2 mutant mice. *Nature* 425, 721–7 (2003).
 78. Sood, P. *et al.* Characterization of transgenic mice overexpressing wild type and G399S mutant HtrA2/Omi – Implications for PD. *Basal Ganglia* 3, 41 (2013).
 79. Hor, H. *et al.* Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum. Mol. Genet.* 24, 5677–86 (2015).
 80. Liu, X. *et al.* Identification of candidate genes for familial early-onset essential tremor. *Eur. J. Hum. Genet.* 24, 1009–1015 (2016).
 81. Deuschl, G., Raethjen, J., Hellriegel, H. & Elble, R. Treatment of patients with essential tremor. *Lancet Neurol.* 10, 148–161 (2011).
 82. Nahab, F. B., Peckham, E. & Hallett, M. Essential tremor, deceptively simple.... *Pract. Neurol.* 7, 222–33 (2007).
 83. Charles, P. D., Esper, G. J., Davis, T. L., Maciunas, R. J. & Robertson, D. Classification of tremor and update on treatment. *Am. Fam. Physician* 59, 1565–72 (1999).
 84. Bhidayasiri, R. Differential diagnosis of common tremor syndromes. *Postgrad. Med. J.* 81, 756–62 (2005).

85. Thanvi, B., Lo, N. & Robinson, T. Essential tremor-the most common movement disorder in older people. *Age Ageing* 35, 344–9 (2006).
86. Lader, M. Effectiveness of benzodiazepines: do they work or not? *Expert Rev. Neurother.* 8, 1189–91 (2008).
87. Flora, E. Della, Perera, C. L., Cameron, A. L. & Maddern, G. J. Deep brain stimulation for essential tremor: a systematic review. *Mov. Disord.* 25, 1550–9 (2010).
88. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33 Suppl, 228–37 (2003).
89. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* 30, 365–366 (2002).
90. Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* 245, 1073–80 (1989).
91. Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–70 (1987).
92. Li, W. & Olivier, M. Current analysis platforms and methods for detecting copy number variation. *Physiol. Genomics* 45, 1–16 (2013).
93. Duerr, R. H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 314, 1461–3 (2006).
94. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745–755 (2011).
95. Morgan, T. H. SEX LIMITED INHERITANCE IN DROSOPHILA. *Science* 32, 120–2 (1910).
96. Tom Strachan & Andrew P. Read. *Human Molecular Genetics* 4. (Garland Science/Taylor & Francis Group, 2011). at https://books.google.com.tr/books/about/Human_Molecular_Genetics_4.html?id=FIItEmQEACAAJ&redir_esc=y
97. Chakravarti, A. & Nei, M. Utility and efficiency of linked marker genes for

- genetic counseling. II. Identification of linkage phase by offspring phenotypes. *Am. J. Hum. Genet.* 34, 531–51 (1982).
98. Sellick, G. S. *et al.* Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res.* 32, e164 (2004).
 99. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24 (2012).
 100. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–9 (2005).
 101. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* 456, 728–31 (2008).
 102. Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–70 (1987).
 103. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–7 (1977).
 104. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674–9
 105. Zimmermann, J., Voss, H., Schwager, C., Stegemann, J. & Ansorge, W. Automated Sanger dideoxy sequencing reaction protocol. *FEBS Lett.* 233, 432–436 (1988).
 106. Venter, J. C. *et al.* The sequence of the human genome. *Science* 291, 1304–51 (2001).
 107. Weber, J. L. & Myers, E. W. Human Whole-Genome Shotgun Sequencing. *Genome Res.* 7, 401–409 (1997).
 108. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
 109. Jasny, B. R. & Roberts, L. Building on the DNA Revolution. *Science* (80-.). 300, 277–277 (2003).
 110. Thorisson, G. A., Smith, A. V., Krishnan, L. & Stein, L. D. The International

- HapMap Project Web site. *Genome Res.* 15, 1592–1593 (2005).
111. Aldridge, S. *et al.* 1000 Genomes project. *Nat. Biotechnol.* 26, 256–256 (2008).
 112. Venter, J. C., Levy, S., Stockwell, T., Remington, K. & Halpern, A. Massive parallelism, randomness and genomic advances. *Nat. Genet.* 33 Suppl, 219–27 (2003).
 113. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728–32 (2005).
 114. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–80 (2005).
 115. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* 26, 1135–1145 (2008).
 116. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38 (2013).
 117. Chen, J.-H., Qiu, J., Chen, H., Pang, C. P. & Zhang, M. Rapid and cost-effective molecular diagnosis using exome sequencing of one proband with autosomal dominant congenital cataract. *Eye (Lond)*. 28, 1511–6 (2014).
 118. Bao, R. *et al.* Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Cancer Inform.* 2014, 67–82 (2014).
 119. Human Genome Sequencing Consortium, I. Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945 (2004).
 120. Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat. Methods* 7, 111–118 (2010).
 121. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276 (2009).
 122. Sulonen, A.-M. *et al.* Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.* 12, R94 (2011).
 123. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray

- hybridization. *Nat. Methods* 4, 903–905 (2007).
124. Bodi, K. *et al.* Comparison of commercially available target enrichment methods for next-generation sequencing. *J. Biomol. Tech.* 24, 73–86 (2013).
 125. Asan *et al.* Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol.* 12, R95 (2011).
 126. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19096–101 (2009).
 127. Goh, G. *et al.* Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics Inform.* 10, 214 (2012).
 128. Julie, G., Hamdan, F. F. & Rouleau, G. A. A strategy to identify de novo mutations in common disorders such as autism and schizophrenia. *J. Vis. Exp.* (2011). doi:10.3791/2534
 129. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nat. Rev. Genet.* 13, 565–75 (2012).
 130. Bilgüvar, K. *et al.* Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 467, 207–10 (2010).
 131. Wang, J. L. *et al.* TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* 133, 3510–8 (2010).
 132. Novarino, G. *et al.* Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* 343, 506–11 (2014).
 133. Züchner, S. *et al.* Whole-Exome Sequencing Links a Variant in *DHDDS* to Retinitis Pigmentosa. *The American Journal of Human Genetics* 88, (2011).
 134. Vissers, L. E. L. M. *et al.* A de novo paradigm for mental retardation. *Nat. Genet.* 42, 1109–12 (2010).
 135. Guerreiro, R. J. *et al.* Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer’s disease. *Neurobiol. Aging* 33, 1008.e17–1008.e23 (2012).
 136. Ramagopalan, S. V *et al.* Rare variants in the CYP27B1 gene are associated with

- multiple sclerosis. *Ann. Neurol.* 70, 881–6 (2011).
137. Zimprich, A. *et al.* A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. *The American Journal of Human Genetics* 89, (2011).
 138. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–81 (2009).
 139. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2745–7 (2015).
 140. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–9 (2010).
 141. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39, e118 (2011).
 142. Marie Schwarz, J., Rödelberger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. (2010). doi:10.1038/nmeth0810-575
 143. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* 12, 228 (2011).
 144. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 42, 30–35 (2010).
 145. Lupski, J. R. *et al.* Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. *Genome Med.* 5, 57 (2013).
 146. Bowen, M. E. *et al.* Loss-of-Function Mutations in PTPN11 Cause Metachondromatosis, but Not Ollier Disease or Maffucci Syndrome. *PLoS Genet.* 7, e1002050 (2011).
 147. Hoischen, A. *et al.* De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* 42, 483–5 (2010).
 148. Walsh, T. *et al.* Whole exome sequencing and homozygosity mapping identify

- mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *Am. J. Hum. Genet.* 87, 90–4 (2010).
149. Pierce, S. B. *et al.* Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am. J. Hum. Genet.* 87, 282–8 (2010).
 150. Krawitz, P. M. *et al.* Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat. Genet.* 42, 827–9 (2010).
 151. Gilissen, C. *et al.* Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am. J. Hum. Genet.* 87, 418–23 (2010).
 152. Byun, M. *et al.* Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J. Exp. Med.* 207, 2307–12 (2010).
 153. Musunuru, K. *et al.* Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* 363, 2220–7 (2010).
 154. Haack, T. B. *et al.* Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat. Genet.* 42, 1131–4 (2010).
 155. Bolze, A. *et al.* Whole-exome-sequencing-based discovery of human FADD deficiency. *Am. J. Hum. Genet.* 87, 873–81 (2010).
 156. Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* 68, 857–64 (2010).
 157. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* 17, 9–18 (2015).
 158. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* 42, 790–793 (2010).
 159. Worthey, E. A. *et al.* Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* 13, 255–262 (2011).
 160. Çalışkan, M. *et al.* Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome

- 19p13. *Hum. Mol. Genet.* 20, 1285–9 (2011).
161. Becker, J. *et al.* Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* 88, 362–71 (2011).
 162. Norton, N. *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am. J. Hum. Genet.* 88, 273–82 (2011).
 163. Simpson, M. A. *et al.* Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat. Genet.* 43, 303–5 (2011).
 164. Isidor, B. *et al.* Truncating mutations in the last exon of NOTCH2 cause a rare skeletal disorder with osteoporosis. *Nat. Genet.* 43, 306–8 (2011).
 165. Glazov, E. A. *et al.* Whole-Exome Re-Sequencing in a Family Quartet Identifies POP1 Mutations As the Cause of a Novel Skeletal Dysplasia. *PLoS Genet.* 7, e1002027 (2011).
 166. O’Sullivan, J. *et al.* Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome. *Am. J. Hum. Genet.* 88, 616–20 (2011).
 167. Vissers, L. E. L. M. *et al.* Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, encoding the Golgi-resident nucleotide phosphatase, gPAPP. *Am. J. Hum. Genet.* 88, 608–15 (2011).
 168. Puente, X. S. *et al.* Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. *Am. J. Hum. Genet.* 88, 650–6 (2011).
 169. Götz, A. *et al.* Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am. J. Hum. Genet.* 88, 635–42 (2011).
 170. Klein, C. J. *et al.* Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* 43, 595–600 (2011).
 171. O’Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders

- identifies severe de novo mutations. *Nat. Genet.* 43, 585–589 (2011).
172. Garrod, A. E. About Alkaptonuria. *Med. Chir. Trans.* 85, 69–78 (1902).
 173. Bittles, A. Consanguinity and its relevance to clinical genetics. *Clin. Genet.* 60, 89–98 (2001).
 174. Bittles, A. H. & Black, M. L. Consanguinity, human evolution, and complex diseases. *Proc. Natl. Acad. Sci.* 107, 1779–1786 (2010).
 175. Gulsuner, S. *et al.* Homozygosity mapping and targeted genomic sequencing reveal the gene responsible for cerebellar hypoplasia and quadrupedal locomotion in a consanguineous kindred. *Genome Res.* 21, 1995–2003 (2011).
 176. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).
 177. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* 428, 529–535 (2004).
 178. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–9 (2009).
 179. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–2 (2010).
 180. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–303 (2010).
 181. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6, 80–92
 182. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010).
 183. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–6 (2011).
 184. Johansson, M. U. *et al.* Defining and searching for structural motifs using

DeepView/Swiss-PdbViewer. *BMC Bioinformatics* 13, 173 (2012).

185. Ra, H.-J. & Parks, W. C. Control of matrix metalloproteinase catalytic activity. *Matrix Biol.* 26, 587–96 (2007).
186. Clark, I. M., Swingler, T. E., Sampieri, C. L. & Edwards, D. R. The regulation of matrix metalloproteinases and their inhibitors. *Int. J. Biochem. Cell Biol.* 40, 1362–78 (2008).
187. Sadowski, T., Dietrich, S., Koschinsky, F. & Sedlacek, R. Matrix metalloproteinase 19 regulates insulin-like growth factor-mediated proliferation, migration, and adhesion in human keratinocytes through proteolysis of insulin-like growth factor binding protein-3. *Mol. Biol. Cell* 14, 4569–80 (2003).
188. van Horssen, J., Bö, L., Vos, C. M. P., Virtanen, I. & de Vries, H. E. Basement membrane proteins in multiple sclerosis-associated inflammatory cuffs: potential role in influx and transport of leukocytes. *J. Neuropathol. Exp. Neurol.* 64, 722–9 (2005).
189. MURPHY, G. *et al.* Evaluation of Some Newer Matrix Metalloproteinases. *Ann. N. Y. Acad. Sci.* 878, 25–39 (1999).
190. Pendas, A. M. *et al.* Identification and Characterization of a Novel Human Matrix Metalloproteinase with Unique Structural Characteristics, Chromosomal Location, and Tissue Distribution. *J. Biol. Chem.* 272, 4281–4286 (1997).
191. Cossins, J., Dudgeon, T. J., Catlin, G., Gearing, A. J. H. & Clements, J. M. Identification of MMP-18, a Putative Novel Human Matrix Metalloproteinase. *Biochem. Biophys. Res. Commun.* 228, 494–498 (1996).
192. Kim, Y.-S. & Joh, T. H. Matrix metalloproteinases, new insights into the understanding of neurodegenerative disorders. *Biomol. Ther. (Seoul)*. 20, 133–43 (2012).
193. Cañete Soler, R., Gui, Y. H., Linask, K. K. & Muschel, R. J. MMP-9 (gelatinase B) mRNA is expressed during mouse neurogenesis and may be associated with vascularization. *Brain Res. Dev. Brain Res.* 88, 37–52 (1995).
194. Vaillant, C. *et al.* MMP-9 deficiency affects axonal outgrowth, migration, and apoptosis in the developing cerebellum. *Mol. Cell. Neurosci.* 24, 395–408

(2003).

195. Larsen, P. H., DaSilva, A. G., Conant, K. & Yong, V. W. Myelin formation during development of the CNS is delayed in matrix metalloproteinase-9 and -12 null mice. *J. Neurosci.* 26, 2207–14 (2006).
196. Larsen, P. H., Wells, J. E., Stallcup, W. B., Opdenakker, G. & Yong, V. W. Matrix metalloproteinase-9 facilitates remyelination in part by processing the inhibitory NG2 proteoglycan. *J. Neurosci.* 23, 11127–35 (2003).
197. Backstrom, J. R., Miller, C. A. & Tökés, Z. A. Characterization of Neutral Proteinases from Alzheimer-Affected and Control Brain Specimens: Identification of Calcium-Dependent Metalloproteinases from the Hippocampus. *J. Neurochem.* 58, 983–992 (1992).
198. Avolio, C. *et al.* Serum MMP-2 and MMP-9 are elevated in different multiple sclerosis subtypes. *J. Neuroimmunol.* 136, 46–53 (2003).
199. Lorenzl, S., Albers, D. S., Narr, S., Chirichigno, J. & Beal, M. F. Expression of MMP-2, MMP-9, and MMP-1 and their endogenous counterregulators TIMP-1 and TIMP-2 in postmortem brain tissue of Parkinson's disease. *Exp. Neurol.* 178, 13–20 (2002).
200. Gu, Z. *et al.* A highly specific inhibitor of matrix metalloproteinase-9 rescues laminin from proteolysis and neurons from apoptosis in transient focal cerebral ischemia. *J. Neurosci.* 25, 6401–8 (2005).
201. Rosenberg, G. A. *et al.* Immunohistochemistry of matrix metalloproteinases in reperfusion injury to rat brain: activation of MMP-9 linked to stromelysin-1 and microglia in cell cultures. *Brain Res.* 893, 104–12 (2001).
202. van Horssen, J. *et al.* Matrix metalloproteinase-19 is highly expressed in active multiple sclerosis lesions. *Neuropathol. Appl. Neurobiol.* 32, 585–593 (2006).
203. Lettau, I. *et al.* Matrix metalloproteinase-19 is highly expressed in astroglial tumors and promotes invasion of glioma cells. *J. Neuropathol. Exp. Neurol.* 69, 215–23 (2010).
204. Fudge, N. J. *et al.* Extracellular matrix-associated gene expression in adult sensory neuron populations cultured on a laminin substrate. *BMC Neurosci.* 14,

15 (2013).

205. Louis, E. D. 'Essential tremor' or 'the essential tremors': is this one disease or a family of diseases? *Neuroepidemiology* 42, 81–9 (2014).
206. Acharya, S., Rayborn, M. E. & Hollyfield, J. G. Characterization of SPACR, a sialoprotein associated with cones and rods present in the interphotoreceptor matrix of the human retina: immunological and lectin binding analysis. *Glycobiology* 8, 997–1006 (1998).
207. Bork, P. & Patthy, L. The SEA module: a new extracellular domain associated with O-glycosylation. *Protein Sci.* 4, 1421–5 (1995).
208. Macao, B., Johansson, D. G. A., Hansson, G. C. & Härd, T. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nat. Struct. Mol. Biol.* 13, 71–6 (2006).
209. Hageman, G. S. & Johnson, L. V. Chapter 9 Structure, composition and function of the retinal interphotoreceptor matrix. *Prog. Retin. Res.* 10, 207–249 (1991).
210. IMPG1 (interphotoreceptor matrix proteoglycan 1) | Gene Report | BioGPS. at <<http://biogps.org/#goto=genereport&id=3617>>

APPENDIX A

Primer List

Table A1. Primers used for PCR amplification of ET-5, ET-49 and ET-31 families.

Primer Name	Direction	Sequence
Segregation		
MMP19	F	ATTGTGTCTGTGGGTGAGCA
	R	CTTCAGCAGCTACCCCAAAC
TMEM139	F	CATGGTGCCAATGCACTTAC
	R	GTCTGCTGTGGGCAAAAGAT
ATXN3	F	TCAGACTAACTGCTCTTGCATTC
	R	CAAAAATCACATGGAGCTCGT
COL24A1	F	GCACATGGCAAGTGCTTAAT
	R	GAATGCAAGGCTGAACCAACA
ST7L	F	ACTACATCCCAGAAGTGGCT
	R	TTTGCTTACTTTGTTTCTCGTATCT
OTOF	F	AATGCCGGCTCTACAAGGTG
	R	ATCAATCCAGGGCAGGTTGG
IMPG1	F	CTGGCTTGGCGGTTTGTTTC
	R	GCAGGTGGCGAAGATATGGT
ITGA10	F	CCCCACAAATGTTGGTTAAATTATTCT
	R	GCCTCATCTCTAGCCAGCAG
IQCE	F	GGAGGCTGTTCACTGACCAG
	R	GCTCCAGTTTCCGCCACATG
SLC34A1	F	TGACAGGAGTTGTGGTGGTG
	R	CAGCAGGACCAGCACTGAC

