IDENTIFICATION AND CHARACTERIZATION OF EXONIC VARIANTS RELATED WITH FAMILIAL ESSENTIAL TREMOR

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

NEUROSCIENCE

By

İSLAM OĞUZ TUNCAY

July, 2017

IDENTIFICATION AND CHARACTERIZATION OF EXONIC VARIANTS RELATED WITH FAMILIAL ESSENTIAL TREMOR

By İslam Oğuz Tuncay,
July, 2017

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

Ayşe Begüm Tekinay (Advisor)

Michelle Marie Adams

Fatma Nazlı Durmaz Çelik

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan

Director of the Graduate School of Engineering and Science

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF EXONIC VARIANTS RELATED WITH FAMILIAL ESSENTIAL TREMOR

İslam Oğuz Tuncay
M.Sc. in Neuroscience
Advisor: Ayşe Begüm Tekinay
July, 2017

Essential tremor (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. In this study, whole exome sequencing and pedigree analyses were performed in unrelated ET families from Anatolia. Whole exome sequencing analysis of family members resulted in the identification of *MMP19* p.R456Q in families ET-5 and ET-49. Expression analysis in mice showed a possible developmental pattern for expression of MMP-19 as well as a tissue-specific expression pattern showing high levels of expression in the brain for this gene. Two other families, ET-17 and ET-19 were also analyzed; however the results were not able to identify variant cosegregating with ET in these families. Identification of the new genes related with ET will provide invaluable insights into the underlying mechanism of thıs most common movement disorder and will potentially open new avenues for its treatment.

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

ÖZET

AİLESEL ESANSİYEL TREMOR İLE İLİŞKİLİ EKZONİK VARYANTLARIN BELİRLENMESİ VE KARAKTERİZASYONU

İslam Oğuz Tuncay
Nörobilim Yüksek Lisans
Tez Danışmanı: Ayşe Begüm Tekinay
Temmuz, 2017

Esansiyel tremor (ET), insanlardaki en yaygın hareket bozukluğudur. Kalıtsallığı ve frekansı oldukça yüksek olmasına rağmen, ET'nin genetik temeli ve patofizyolojisi tam olarak anlaşılamamıştır. Bu çalışmada, Anadolu dört ila beş kuşak boyunca otozomal dominant kalıtım modeline uygun düzende ET vakaları gözlenen Anadolu kökenli ailelerde tüm ekzom dizileme ve soyağacı analizi gerçekleştirilmiştir. Yapılan analizler sonucu, hastalıkla birlikte nesilde nesle aktarılan, protein seviyesinde zarar verici bir mutasyon belirlendi: ET-5 ve ET-49 ailelerinde *MMP19* p.R456Q. Farelerde yapılan ifade analizleri, beyinde yüksek seviyede ekspresyon gösteren dokuya özgü bir ifade şablonunun yanı sıra, MMP19 ifadesi için olası bir gelişimsel düzen gösterdi. İki ayrı aile, ET-17 ve ET-19 da aynı analiz yöntemlerinden geçirildi, ancak bu ailelerde ET ile birlikte aktarılan bir mutasyon belirlenemedi. ET ile ilgili yeni genlerin tanımlanması, en yaygın hareket bozukluğunun altında yatan mekanizma hakkında paha biçilemez bilgiler sağlayacak ve potansiyel olarak tedavisi için yeni yollar açacaktır.

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

ACKNOWLEDGEMENTS

I would like to start by thanking my advisor Asst. Prof. Dr. Ayşe B. Tekinay for her guidance, motivation and support. I feel lucky to have been able to work with her for the last two years. I would like to thank Prof. Dr. Tayfun Özçelik for his continuous support, his uncanny ability to teach a valuable lesson in every interaction, and sharing his immense knowledge of genetics.

I want to thank Prof. Dr. Cenk Akbostancı and Prof. Dr. Haluk Topaloğlu for the identification and recruitment of patients. I would like to thank Dr. Emre Onat for his help in my experiments and data analysis. I would like to thank Asst. Prof. Dr. Nazlı Durmaz Çelik, Dr. Çağrı Ulukan, Dr. Eda Aslanbaba and Dr. Adem Demir for their help with the clinical assessment of patients. I'd like to thank Dr. Peren Karagin for all her help. I am thankful to all essential tremor patients, relatives and other participants for their cooperation in this study.

I would like to thank Dr. Seher Yaylacı and Merve Şen for their help, support and collaboration that went way beyond a work friendship. I have learned so much from them, and I will forever cherish the memories of us running to meetings and carrying way too many files. I would also like to thank the youngest member of our team, little miss Neva Yaylacı for simply being adorable. I would like to thank Nuray Gündüz for being by my side and helping (and sharing my frustrations) with *in situ* hybridization for the last couple months. I also would like to thank Melike Sever for her help with the qRT-PCR experiments and her friendship. I would also like to thank our intern Umut Taşdelen for his help with PCR and 3D modelling.

I would like to thank Özge Uysal for being by my side, literally since my first class in college. I would not be able to get through last six years without her friendship and her notes – which were always a joy to read thanks to her impeccable handwriting. I would also like to thank Göksemin Şengül, for her erratic yet strangely calming presence. Both Göksemin and Özge gave me a push whenever I needed one, and I appreciate them greatly. I would like to thank Zeynep Orhan for teaching me how to use bacteria, the finicky little creatures that brought me to the brink of insanity. I would like to thank Fatih Yergöz for giving me competition for how chill one can be. I would like to thank Nurcan Haştar for accompanying me whenever I wanted to sing

a random Turkish pop song, and I would like to thank İdil Uyan for showing me how it actually should be sung. I would like to thank Mustafa Beter for being an amazing desk mate and showing how to discuss passionately. I would like to Canelif Yılmaz for humbling me about my baking skills. I want to thank Çağla Eren for her infectious laughter and joy. I'd like to thank the immunology duo, Şehmus and Burak for their friendship. I want to thank NBT-BML lab members Asst. Prof. Mustafa Özgür Güler, Gökhan Günay, Begüm Dikeçoğlu, Dr. Gülistan Tansık, İbrahim Çelik, Elif Arslan, Dr. Berna Şentürk, Ahmet Emin Topal, Alper Devrim Özkan and Dr. Gözde Uzunallı for all their help, and for creating such a warm working environment.

I have to thank my dearest friend, Salih Aksoy, for putting up with me for 10 years, and always being there for me. I truly don't know what I would do without him. I'd like to thank Murat Demirbüken for all the late night trips to get waffles. I'd like to thank Muammer Yaman for being the loveliest roommate one could have. I'd like to thank Ömer Fatih Konar for watching movies and laughing with me, even though we both knew we should've been studying instead, and being the younger brother I never had. I want to thank Enes Aybar for teaching me basics of American football.

I'd like to thank Alper Duranel for all the buckets of KFC we shared while watching sit-coms, and for being okay with me liking chocolate ice cream over vanilla. I'd like to thank my text-chain friends Ali Fuat Geyik, Buğrahan Şahin, Fatih Yiğit, Mustafa Yılmaz, Muhammed Tanır and Sefa Aydemir for enduring all the memes I have shared. I would also like to thank two of my oldest friends, Göktuğ Kalender and Abdullah Topçuoğlu for always being supportive of me and always being just a phone call away.

I'd like to thank my London trip crew; Dr. Tuğrul Nalbantoğlu for all his insults that are too funny to get mad at, Aykut Argun for all the late night walks in Bilkent, Alper İnecik for letting me watch Modern Family from his cellphone which was definitely too small for the job, and Yasin Kaya for his surprising ability to hit a volleyball with the back of his hand, and our hosts Burak Şimşek, Cihad Öge and Yasin Kadıoğlu for taking Turkish hospitality to a whole new level. What happened in London may stay in London, but our friendship is forever.

I’d like to thank Dr. Bilal Uyar for our shared car rides, along with his support and guidance. I would like to thank my young friends who inspired me not just be a better scientist but a better person as well; Kemal, Emin, Furkan, Alperen, Elif, Selim, Sümeyye, Şevval, Tolga, Alp Eren, Bedriye, Buğra Han and Sinem.

Last but not the least, I want to thank my family; my late father Eyyup, my amazing mom Nesrin for being the strongest person I know, my aunt Ayşe who is like a second mom to me, my older sister Neslihan for always encouraging me to reach for the stars, my younger sister Nurnihan for giving me the idea to pursue a carrier in genetics, and my nephews İhsan, Furkan and Tahir and my niece Zeynep for being the joy of my life. My family always supported me, loved me, and believed in me, and I’m thankful for everything they have done for me.

Thank you,

İslam Oğuz Tuncay

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abbreviations

BWA	Burrows-Wheeler Aligner
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
ISH	*In Situ* Hybridization
MAF	Minor Allele Frequency
MMP19	Matrix Metallopeptidase 19
NGS	Next Generation Sequencing
PD	Parkinson's Disease
RT-PCR	Real Time Polymerase Chain Reaction
SAMtools	Sequence Alignment/Map Tools
SNP	Single Nucleotide Polymorphism
VCF	Variant Call File
WES	Whole Exome Sequencing

CHAPTER 1

Introduction

1.1 Essential Tremor

1.1.1 Clinical Features and Pathophysiology

Essential Tremor (ET [OMIM 190300]) is a chronic, progressive neurological disease, and with a prevalence of 0.9% in general population it is often regarded as the most common adult movement disorder[1]. Different tremors are seen as a symptom in a number of neurological and motor conditions (Table 1). ET's characteristic motor symptom is a 4-12 Hz kinetic tremor of hands and arms observed when performing voluntary movements[2]. ET patients may develop additional motor symptoms, such as resting tremor, the spreading of tremor to legs, neck, voice and other organs[3,4], as well as a number of non-motor symptoms including mild cognitive deficits, psychiatric impairments (e.g. anxiety, depression), partial loss of hearing and sleeping problems[5,6,7]. Most prominent problem caused by ET is the result of its effect on the upper limbs, which a staggering 95% of the affected individuals suffer. This effect causes hardship while performing day-to day actions such as holding a pen, drinking, or eating[3,4,8].

Although it is a very common disease, there is no consensus about the pathophysiology of ET, or its classification as a functional or a neurodegenerative disease[9]. Cerebellar involvement, including dendrite swellings and Purkinje cell heterotopias, has been reported in several clinical, physiological and neuroimaging studies[10,11]. In addition to pathophysiology, ET shows heterogeneity in terms of its etiology[12], age of onset[13], clinical features[14] and pharmacological response

phenotype[15,16], supporting the idea that ET might not be a single disease, rather a family of diseases sharing a key feature, kinetic tremor of arms[17].

1.1.2 Etiology and Diagnostics

Etiologically, ET can be divided into three subsections. Patients with age of onset above 65 years of age are considered as senile ET. In sporadic ET, patients are defined as fulfilling the consensus criteria when under 65 years old with no family history of the disease. If the patient has another family member diagnosed with non-senile ET, they are categorized under hereditary ET[18].

Frequency of ET can increase up to 6.3% and 21.7% among individuals aged ≥60-65 and ≥90, respectively[19]. ET patients with family history of the disease is in the 30-70% range according to population studies, and twin studies estimates the heritability of ET to be between 45% and 90%[20,21]. Environmental factors that have been associated with sporadic ET cases include higher blood levels of β-carboline alkaloids and lead[22–24].

Diagnosis of ET is solely dependent on clinical assessment and patient's medical history. Subjectivity of this methodology is the main reason why incorrect diagnosis is abnormally common, estimated between 30 to 50%, and a consensus criteria has therefore determined by the Movement Disorder Society (MDS)[18].

	Resting Tremor	Action Tremor			
		Kinetic Tremor			Postural Tremor
		Intentional tremor	Basic Kinetic Tremor	Task specific tremor	
Observed When	No voluntary movement, arms fully supported	Performing any voluntary movement	Performing any voluntary movement	Performing a specific voluntary movement	While maintaining a static position
Tremor Amplitude	High	Increases when nearing a target	Stable	Varies	Low
Tremor Frequency	Low to moderate	Below 3 Hz	Varies (3 to 10 Hz)	Varies (4 to 10 Hz)	Moderate to high
Affiliated Conditions	Parkinson's disease, Drug-induced Parkinsonism	Multiple Sclerosis, Stroke		Writer's tremor, Musician's tremor	Essential tremor, metabolic disorders

Table 1 Types of tremor.

1.1.3 Genetics of Essential Tremor

Despite the prevalence of the disease and evident contribution of hereditary factors, genetic studies on ET are limited. Three genome-wide linkage studies, all performed on either Northern American or Icelandic populations, have been published to date. Genome-wide scan of 75 ET patients from 16 Icelandic families resulted in the identification of first ET locus, *ETM1* [OMIM 190300] (then called *FET1*, for *F*amilial *E*ssential *T*remor locus 1). Combined logarithm of odds (LOD) score for the proposed model of dominant inheritance was 3.71; however no single family score was above the significance threshold for a monogenic disorder marker to be mapped, as single-family LOD scores were all calculated to be ≤1.29[25]. Located on 2p22-p25, the *ETM2* [OMIM 602134] locus was identified in a Czech-American family. LOD score was calculated to be 5.92 for the autosomal dominant model of inheritance[26]. The third genome-wide linkage study for ET was performed on a much larger cohort compared to the previous two, recruiting 325 affected individuals from 7 different families, all from North America. The locus revealed was located on 6p23, and was named *ETM3* [OMIM 611456][27]. A consistent problem for all three loci is the lack of reproducibility, as various linkage and association studies on different cohorts have failed to show a significant relation between these loci and ET[28–34].

Potential disease-causing variants within ETM loci have been the subject of several studies. A study on 30 families of French descent resulted in the identification of *DRD3* p.Ser9Gly variant, located in *ETM1*, present in 23 of the said families. Studies on Italian and Asian families, however, didn't support these findings as results showed no significant association of the Ser9Gly variant with ET[35,36]. *HS1-BP3* 828C→G variant was identified by a systematic screening of an established minimal critical region within *ETM2*[37,38]. Consequent studies were not able to confirm these

results[27,39]. 15 genes have been sequenced and analyzed by Shatunov *et al*. to find a pathologic variant within *ETM3*, but none was found[27].

Genome-wide association studies resulted in the identification of two genes in relation to ET. *LINGO1* was identified as a risk factor on a cohort of American and European families. Identified cosegregating variants of *LINGO1* are intronic, but suggested to be potentially disruptive in conjunction with environmental factors[40–42]. In another study recruiting a European cohort *SLC1A2* was identified as a risk factor, but the variants were deemed non-causative by a meta-analysis study[43,44].

For the last couple of years, a growing number of studies utilized whole exome sequencing to identify novel variants related to ET. A 2012 study was the first of this trend, identifying *FUS/TLS* (fused in sarcoma/translated in liposarcoma) c.868C>T nonsense mutation cosegregating with ET in a large French-Canadian kindred[45]. Following up, ET cohort screenings revealed M329I and R377W missense variants, further supporting the case for *FUS* as a ET-related gene[46,47]. In 2014, our group has identified *HTRA2* p.G399S missense mutation in a six-generation consanguineous Turkish kindred cosegregating with both ET and PD[48]. The genetic correlation between ET and PD was further supported by a 2015 report by Rajput *et al* which showed that *DNAJC13* c.2564A>G, a variant previously identified in PD patients, was present in 2 affected individuals from a 571-patient ET cohort[49]. Again in 2015, two novel missense variants of *TENM4* were identified to be cosegregating with ET in a Spanish family[50]. In another study featuring a Spanish family, Nav1.4 p.Gly1537Ser located on *SCN4A* gene was found to be cosegregating with ET. A possible relation between ET and epilepsy was suggested as the mutation affects the ion selectivity of Nav1.4, which is a voltage-gated sodium channel[51]. Liu *et al* reported five novel missense variants of four different genes five families with early onset ET; *NOS3*

p.Gly16Ser and p.Pro55Leu, *KCNS2* p.Asp379Glu, *HAPLN4* p.Gly350Arg and *USP46* p.Ala133Val[52]. Latest report by Leng *et al* identifies *SCN11A* p.Arg225Cys as a putative genetic contributor to early-onset familial episodic pain and adult onset hereditary ET in a four-generation Chinese family[53].

1.2 Identification of Disease-Related Genes

1.2.1 Linkage and Association Studies

Although their popularity has declined especially in the last decade due to the rise of next generation sequencing, mapping-based methods have been –and still are – useful tools for the identification of disease-related genes.

Karyotyping have been useful in identifying several developmental syndromes[54], but since chromosomal abnormalities are often *de novo* cases, this method is not particularly useful in identifying genetic factors behind inherited traits. Morgan's studies on fruit flies showed the role of linkage in the inheritance of Mendelian traits[55]. Traits that map closely will be less likely to be subject to recombination, and therefore usually cosegregate within families, therefore recombination studies can be utilized to link traits with encoding genes[56]. Logarithm of odds (LOD) score is used for the statistical verification of the association between traits and genes[57]. When it comes to polygenic traits, however, linkage studies are inefficient[58].

Genome wide association studies (GWAS) examine the nonrandom association of common SNPs and disease phenotypes at the population level. Based on the common disease-common variant hypothesis, GWAS studies assume diseases with a high prevalence will be associated with genetic factors that also have a high prevalence[59]. The International HapMap project was crucial for the development of GWAS, since the genotype data from various populations helped identify variation across genome and determine correlations between common variant, so that phenotype studies in different populations have the proper strategy and don't result in the collection of redundant data[60].

Sequencing data for the common SNPs is used to determine linkage disequilibrium (LD), which identifies a lack of random disassociation, similar to chromosomal linkage. Linkage disequilibrium of SNPs can help identify genetic factors related to certain traits via either direct or indirect association. In direct association, the SNP that shows high LD is associated with said trait. In indirect association, the SNP that shows high LD is not the influential SNP, but it is a common SNP that is linked to the influential SNP[61]. For genotyping, GWAS studies utilize chip-based microarray technologies, mainly including Illumina (San Diego, CA) or Affymetrix (Santa Clara, CA) products[62]. The idea GWAS adapted into family studies in the shape of homozygosity mapping, where the common SNP arrays of affected and unaffected family members are used to identify homozygous regions inherited within a family[63]. Main shortcoming of GWAS is that it's unsuitable for studying rare conditions, as it is based on utilizing *common* variants[64].

1.2.2 Sequencing-Based Methods

1.2.2.1 Early Methods and Development of Next Generation Sequencing

First method for DNA sequencing was based on chain termination with dideoxy nucleotides. Described by Sanger *et al*. in 1977, this approach would be the basis for the later developed Sanger sequencing, using fluorescence detection for automated DNA sequence analysis[65–67]. A major caveat of this method is that it is a relatively slow method if you want to sequence massive sequences such as the human genome, as it can't sequences longer than ~1000 base pairs.

During the beginning phase of Human Genome Project (HGP), the lack of a high-throughput sequencing method was the biggest obstacle, and the development of shotgun sequencing 1997 was what resolved this problem[68]. With a divide-and-

conquer approach, shotgun sequencing involved creating short fragments of DNA using enzymatic or mechanic methods. These fragments are then cloned into vectors and sequenced individually. The overlapping sequences between fragments are used for the alignment and assembly of the complete sequence[69]. This parallel sequencing approach would later be the basis of next-generation sequencing (NGS).

NGS uses the same divide-and-conquer approach, but instead of using vectors, fragmented DNA bits are ligated to pre-designated adapters[70]. This method also has its caveats, primarily in terms of data management and analysis. Each NGS reaction results in sequence data about millions of short fragments that only partially overlap, therefore developing computational methods that can effectively and accurately create the complete sequence is needed[71].

1.2.2.2 Whole Exome Sequencing

The human genome is a massive sequence, containing nearly 3 billion bases. The protein-coding portion of the genome, known as the exome, takes up roughly 1% of the genome. Assuming most of the functional information that we can get from a genomic sequence is within the exome, at least in terms of the proportion of information with respect to the length of the sequence, whole exome sequencing (WES) offers a rapid and cost-effective approach for medical genetics purposes. It also lowers the burden on data analysis as the outcome is smaller amount of sequenced fragments that involve fewer repeats[72,73]. In WES, similar traditional NGS methods, the DNA samples are broken down to fragments ligated to adapters, but the adapters that are used in this method are specifically designed probes that hybridize only with exonic fragments, thus “capturing” the exome. Then the probes are either bound to magnetic beads and amplified through PCR for enrichment before high-throughput sequencing (solution-based WES) or bound to a high-density microarray

and sequenced without the need for enrichment (array-based WES) [74,75]. Starting in 2009, familial studies for a wide variety of conditions utilizing WES have been published.

WES approach have proved particularly useful for the identification of novel variants linked with rare conditions[76]. Variant databases made available via broad-scope exome studies like 1000 Genomes Project, Greater Middle East Variome Project (GME) and National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) are used for the exclusion of common variants. Possible functional effects of the variants can be predicted using *in silico* tools such as SIFT, PolyPhen, CADD and PROVEAN[77–80]. Main shortcoming of WES is the loss of information either caused by specific enrichment which eliminates non-exonic sequences that might have important roles in terms of protein expression such as microRNAs, promoters, or the inefficiency of enrichment that may result in the inability of capturing some parts of the exome[81].

CHAPTER 2

Material and Methods

2.1 Subjects

The study was approved by the institutional ethical review boards for studies with human subjects at Bilkent, Hacettepe and Ankara Universities. All participants signed an informed consent form in concordance with the guidelines of Turkish Ministry of Health. Proband of ET-5 was first evaluated at Hacettepe University Medical School, while other members of the family were followed at Hacettepe University Medical School and Ankara University Medical School. Probands of ET-17, ET-19 and ET-49 were first evaluated at Ankara University Medical School, and the other members of these families were followed at Hacettepe University Medical School and Ankara University Medical School. Assessment of essential tremor is done according to Washington Heights-Inwood Genetic Study of Essential Tremor (WHIGET)[82] and Consensus Statement of the Movement Disorder Society on Tremor (MDS)[18]. Helsinki Declaration was regarded during all examinations[83]. Severity of resting and postural tremors for each participant was graded at a scale of 0 to +3: 0 meaning no visible tremor, +1 low amplitude tremor, +2 moderate amplitude tremor and +3 high amplitude tremor. For the evaluation of kinetic tremor, participants were asked to perform four distinct tasks: drawing spirals, pouring water, finger-to-nose movement, and drinking water. Severity of tremor was assessed during each task, again in a scale of 0 to +3. Bradykinesia, muscular rigidity and postural instability assessments were also done according to UK Parkinson Disease Society Brain Bank[84], in effort to distinguish between pure ET cases and ET cases with Parkinsonism.

2.2 Whole Exome Sequencing

Two affected individuals (II-1 and IV-3) and one control (III-2) from ET-5, four affected individuals (III-2, IV-2, IV-4, V-2 and V-3) and one control (IV-1) from family ET-49, two affected family members (II-4 and IV-3) and one control (IV-1) from ET-19, and three affected individuals (IV-9, V-1 and V-5) from ET-17 were selected for whole exome sequencing. Following kits were used according to manufacturer's instructions: Nucleospin Blood Kit (Macherey-Nagel), Illumina TruSeq DNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA), SeqCap EZ Exome Capture Kit (Roche), QIAquick PCR Purification Kit (Qiagen), for the isolation of DNA from blood, library construction, exome capture and exome cleaning, respectively. ABI system KAPA Illumina Library Quantification Kit was used with RT-PCR to determine the quality and the concentration of the exomes. Illumina HiSeq2500 was used for the sequencing of the libraries, and Illumina Real Time Analysis Software (Illumina, Inc., San Diego, CA, USA) for determining the quality of the sequences.

2.3 Bioinformatics

2.3.1 Initial Analysis of WES Output

Sequence data were first converted to .bcl files, and then using Illumina CASAVA software (Illumina, Inc., San Diego, CA, USA) to FASTQ files. Fragments were aligned to the reference genome (UCSC hg19) using Burrows-Wheeler Aligner (BWA, v0.6.1-r104)[85]. Sequence Alignment/Map Tools (SAMtools) software package[86] was used to remove PCR duplicates. BEDtools[87] was used to get sequence depths for exonic regions. Genome Analysis Tool Kit (GATK; v3.0–0-g6bad1c6)[88] was used for the recalibration of base quality scores and SNP and indel variants.

SnpEff (4.2, 2015-12-05)[89] and ANNOVAR (2016Feb01)[90] were used for the annotation of variants annotated based on position and function.

2.3.2 Filtration, Prioritization and Segregation

First round of filtration was according to the impact annotation, as determined by SnpEff. Variants that were considered as having low impact, including synonymous, start-retained and stop-retained mutations, were excluded as they are not protein-altering. Variants that were considered as having high or moderate impact were further examined. The moderate-impact group included inframe indels, splice region variants and missense variants while the high-impact group included frameshift, splice donor/acceptor and stop gain/loss variants.

To create a shortlist of potentially ET-linked variants for each family, an initial round of filtering was done based on three main criteria: first, the variant would have to be on a protein-coding region; second, the minor allele frequency for the allele recorded in ExAC (Exome Aggregation Consortium), the 1000 Genomes Project (1000genomes.org), the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project, and our in-house database would have to be below 0.02%; and third, the out of the individuals that went through WES, the variant should be carried by all of the affected.

We used the genotyping data of the ET cohort to exclude variants. We excluded variants that were in homozygous or compound heterozygous form in any of the control samples. Variants were defined and prioritized as potentially damaging if caused a nonsense mutation, or a missense mutation that was predicted to be damaging by at least two *in silico* prediction tools. Score cut-offs for the prediction tools were ≤2.5 for PROVEAN, ≤0.05 for SIFT, ≥0.453 for Polyphen2 HDIV, ≥0.447

for Polyphen2 HVAR, MutationTaster prediction of A (disease causing automatic), D (disease causing), and P (possibly disease causing), >1.94 for MutationAssessor, ≥10 for CADD and ≥3 for GERP [77–80,91–94].

Cosegregation of the variants with ET within the families ET-5, -31 and -49 was performed using the primers listed on Table A1. Primer3Plus[95] was used to design the primers. Chromas Lite (Technelysium Pty Ltd) and FinchTV was used for the analysis of sequencing data.

2.4 Analysis of Protein Expression and Function

2.4.1 Quantitative Real Time - PCR

All experimental procedures involving animals were approved by the Animal Ethics Committee of Bilkent University (Protocol # 2013/25). RNA was isolated from C57BL/6 mouse tissues using TRIzol (Invitrogen) according to the manufacturer's instructions. Yield and purity of extracted RNA were assessed by Nanodrop 2000 (Thermo Scientific). cDNA synthesis from RNA and qRT-PCR were performed using SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit according to the manufacturer's instructions. Reaction conditions were briefly as follows: 55 °C for 5 min, 95 °C for 5 min, 40 cycles of 95 °C for 15 s, 58 °C for 30 s, and 40 °C for 1 min, followed by a melting curve analysis to confirm product specificity. For analysis of the expression data, primary gene expression data was normalized by the expression level of GAPDH. A comparative Ct method (Pfaffl Method) was used to analyze the results.

2.4.2 *In situ* Hybridization

The expression pattern of Mmp19 gene in the mouse brain was analyzed with *in situ* hybridization as described previously[96]. All experimental procedures involving

animals were approved by the Animal Ethics Committee of Bilkent University (Protocol # 2013/25). Mmp19 gene PCR product was prepared with Q5® High-Fidelity DNA Polymerase from C57BL/6 mouse genomic DNA. PCR products were run in low melting agarose gel and purified with PureLink® Quick Gel Extraction Kit according to the manufacturer's instructions. The purified PCR products were cloned into zero blunt pCR4-TOPO vectors (Invitrogen) and selected colonies were sequenced. The riboprobes were synthesized by using Digoxigenin (Dig)-labeled NTPs (Roche) with RNA transcription kit (NEB). P7 and adult mouse brains were obtained from C57BL/6 mice, which were housed in a 12-h dark, 12-h light cycle, and fed ad libitum. Adult brain sections were prepared as described[97]. Twenty-micrometer sagittal sections were taken with a cryostat (Leica). Sections were incubated at 60 °C overnight in hybridization buffer containing 50% formamide, 5X SSC, 5X Denhardt's reagent, 50 mg/mL heparin, 500 mg/mL herring sperm DNA, and 250 mg/mL yeast tRNA. Hybridized sections were washed for 90 min with 50% formamide and 2X SSC at 60 °C. Probes were detected with anti-Dig Fab fragments conjugated to alkaline phosphatase and NBT/BCIP substrate mixture[96].

2.4.3 Conservation Analysis and 3D Modelling

Sequences for conservation analysis were obtained from NCBI HomoloGene database. Clustal Omega[98] was used for the alignment of obtained sequences. Predicted 3D structure models of the proteins were obtained from Swiss Model, Raptor X or Phyre2. Swiss-PdbViewer (DeepView, v4.1) was used for manipulation and analysis of the 3D models[99–102].

CHAPTER 3

Results and Discussion

3.1 Variant Search in Family ET-17

3.1.1 Clinical Features of ET-17

The clinical diagnosis of ET was made initially in the proband (V-1) of a six-generation consanguineous family, ET-17. ET-17 is of Turkish origin and ET is observed multiple generations. 9 individuals from this family were clinically assessed based on criteria of both consensus statement of MDS[18] and WHIGET[82], and 3 individuals were diagnosed with ET (Figure 1). Archimedes spiral drawing tests are shown in figure 2.

Family ET-17 is from Mardin, located in southeastern Anatolia. Proband of family ET-17 (V-1) was 42 at the time of diagnosis, and her parents declared she has shown symptoms since the age of 21. Detailed clinical assessment of IV-3 showed resting tremor with mild amplitude on both hands, as well as severe kinetic tremor on both hands and moderate action tremor on her neck. She also had mild bradykinesia on her left hand and mild rigidity on her right hand. Other affected individuals were her mother, IV-9 (13, 80) and her sister, V-5 (48, 55), having severe and moderate kinetic tremors, respectively. One of the maternal aunt of the proband, IV-2, is diagnosed with Parkinson's disease.

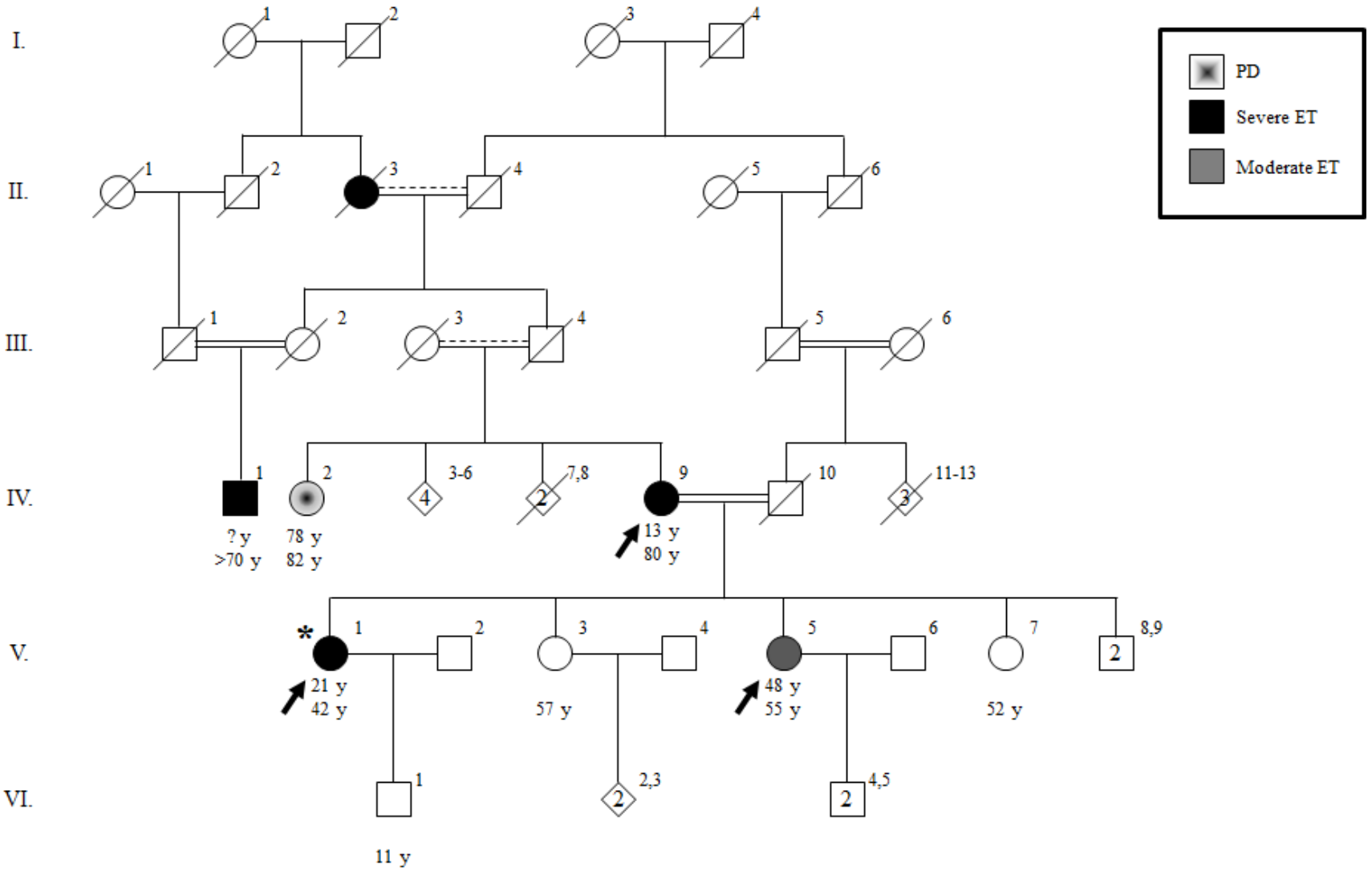


Figure 1 Pedigree of ET-17. Age at onset of tremor for affected individuals and current ages are indicated in this order under the symbols. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

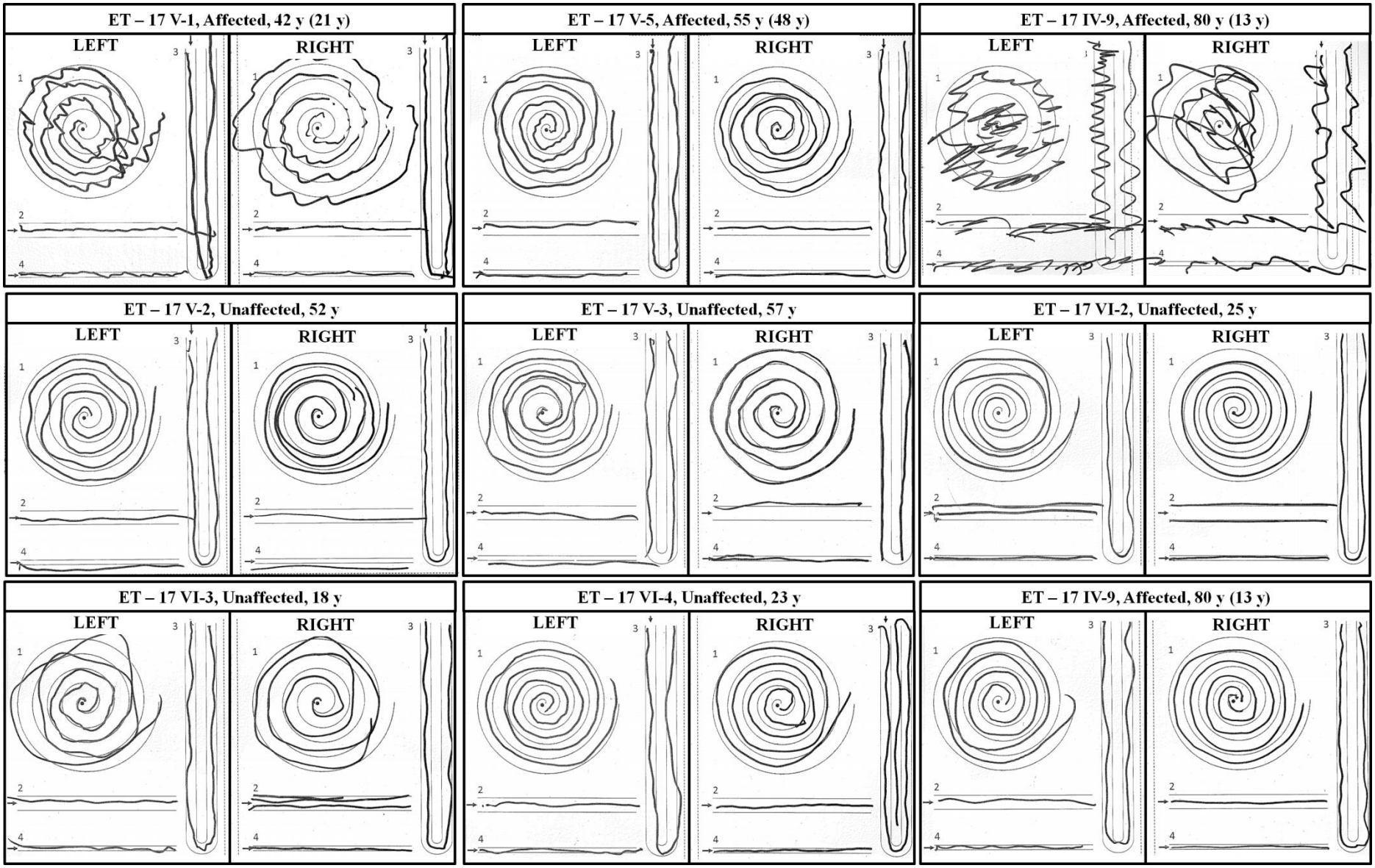


Figure 2 Archimedes spiral drawing test results for members of family ET-17.

3.1.2 Whole Exome Sequencing

Affected family members IV-9, V-1 and V-5 were selected for WES from ET-17. NanoDropTM ND-1000 Spectrophotometer (NanoDrop Technologies, Inc, DE, USA) was used for the purity and concentration measurements. Possible degradation of DNA samples was also checked by running them on agarose gel.

Using Burrows-Wheeler Aligner, WES reads were mapped to the reference genome, UCSC hg19. SAMtools were used to remove duplicate reads. Exonic coverage was calculated with BEDtools.

SnpEff was used for the position and impact annotation of variants, using reference genome UCSC hg19. For the functional annotation, impact of each variant were categorized under either “high”, “moderate”, “modifier” and “low” by SnpEff. Low impact variants and modifier variants were filtered, leaving high impact variants which includes frameshift variants, splice acceptor/donor variants and stop loss variants, and moderate impact variants which includes inframe indels, missense variants and splice region variants.

SnpEff annotation and subsequent filtering was followed with annotation by ANNOVAR software package. Three types of annotations done by ANNOVAR are (1) gene-based, (2) region-based and (3) filter-based annotation. Gene-based and region-based annotations identify the genes and chromosomal regions which the variants are located on, respectively. Filter-based annotation uses a series of databases to call information about the variant in terms population frequency, dbSNP ID, protein damage prediction and evolutionary conservation. Population frequency data was obtained from Genome Aggregation Database (gnomAD), Exome Aggregation Consortium (ExAC), Exome Sequencing Project (ESP), Kaviar (~Known VARiants)

Genomic Variant Database (Kaviar), Greater Middle East Variome Project (GME), Complete Genomics (CG) and 1000 Genomes Project (1000G) databases. Evolutionary conservation scores were obtained from GERP++, phastCons, PhyloP and SiPhy databases.

Protein damage prediction data was obtained from SIFT, PolyPhen HDIV, PolyPhen HVAR, MutationAssessor, M-CAP, MutationTaster, LRT, PROVEAN, FATHMM, MetaSVM, MetaLR and CADD databases. MetaSVM showed the highest percentage of predicted benign mutations with 59.36%, while M-CAP showed the lowest value of 23.01%. For the predicted damaging mutations, highest percentage was 42.93%, predicted by CADD, and the lowest was MetaSVM's 4.09% (Figure 3). 45.5% of the mutations were predicted to be benign by all 12 databases, while 41.9% were predicted to be damaging by more than 2 databases and 15.5% were predicted to be damaging by at least half of the databases (Figure 4).

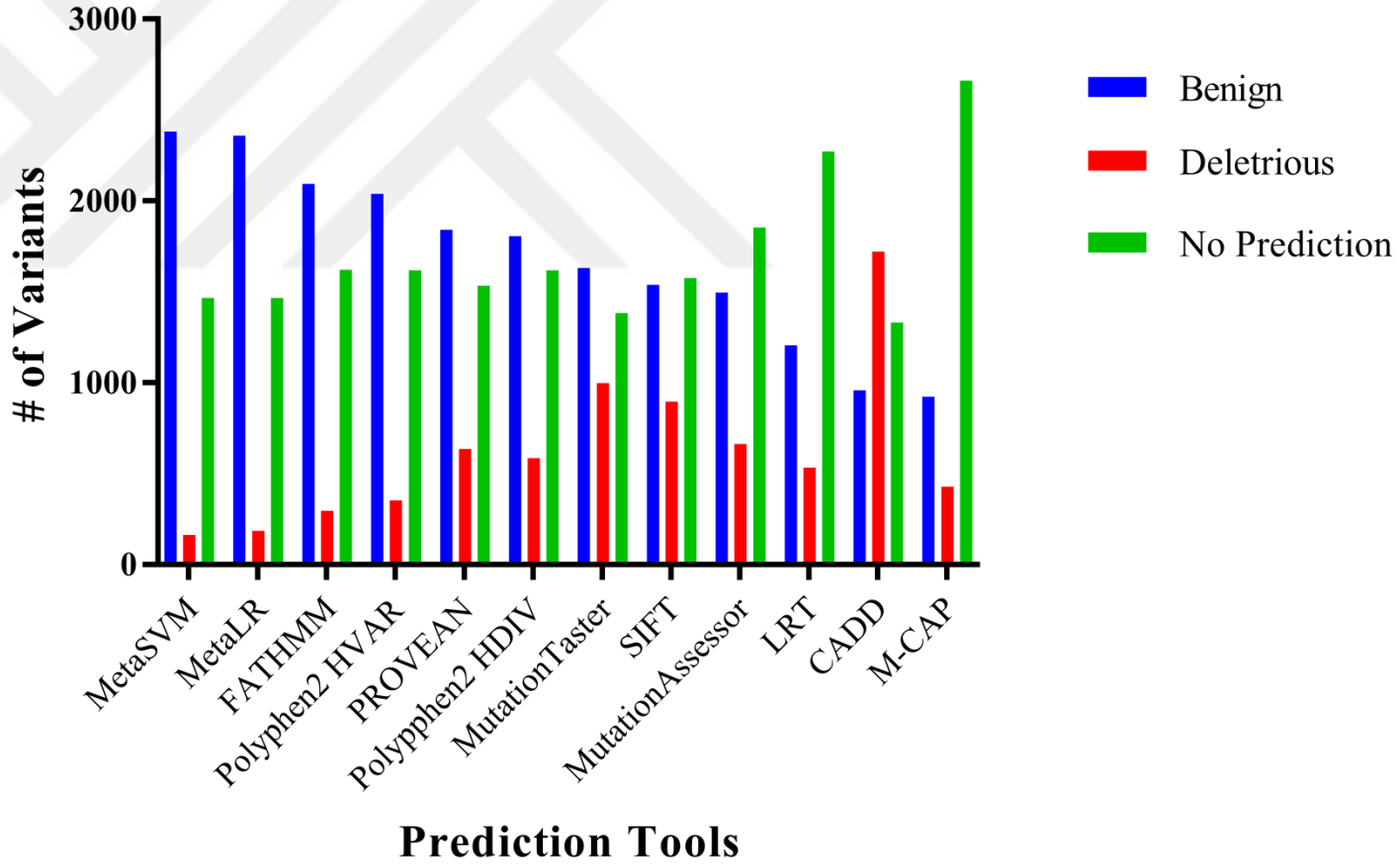


Figure 3 Protein damage prediction for ET-17 variants (1). This plot shows the number of variants each tool predicted to be benign, predicted to be deletrious or possible deletrious, or had no prediction about their effect.

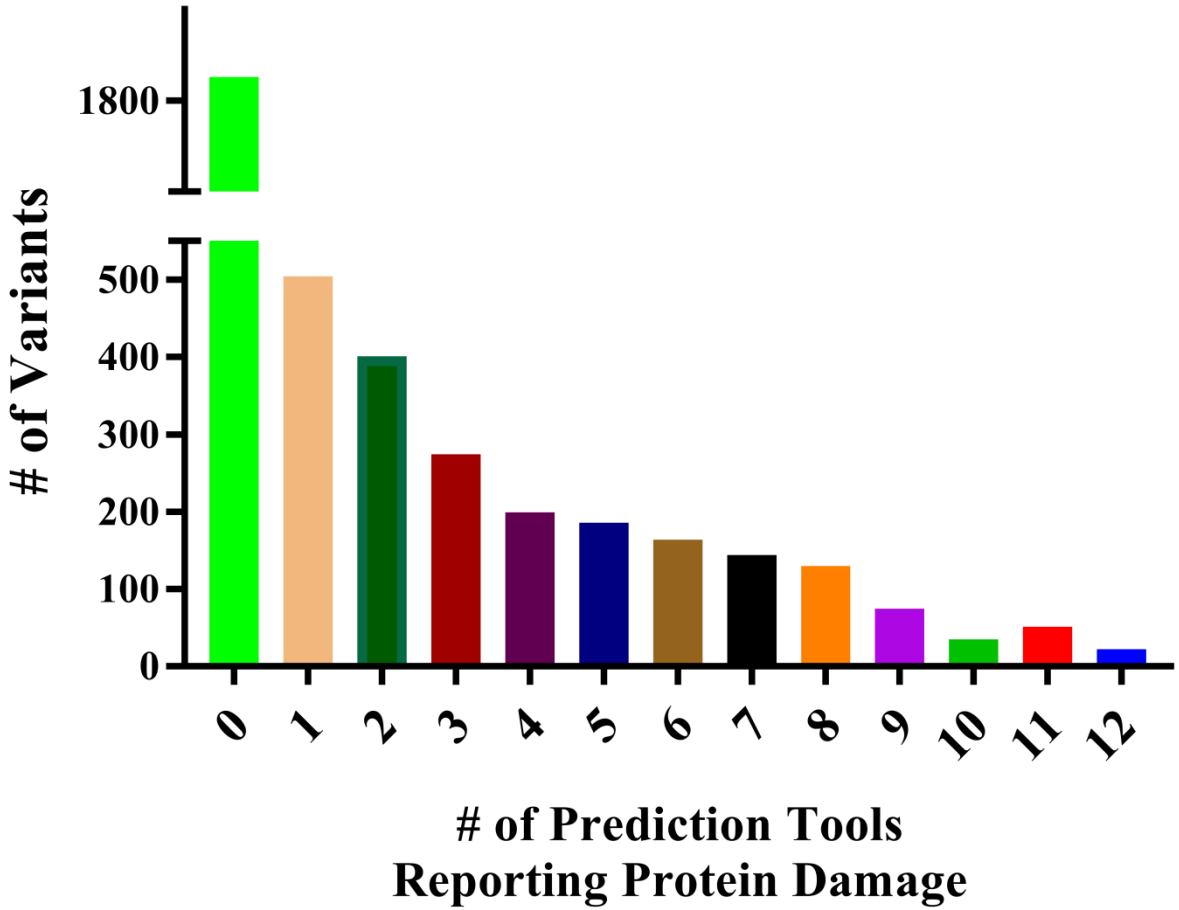


Figure 4 Protein damage prediction for ET-17 variants (2). This plot shows how many of the variants were predicted to be damaging or possibly damaging by how many of the prediction tools out of 12.

3.1.3 Identification of Candidate Variants

After the annotation process, filtration and prioritization of the variants were implemented in order to get a list of candidate variants for the subsequent segregation analysis (Figure 5).

Total variants count for all the members of ET-17 whose DNA went through WES was 5,895,636. First step of filtration was excluding variants annotated as modifier variants or low-impact variants by SnpEff, which decreased the variant count 1,011,279, almost one-sixth of the starting count. This was followed by the exclusion of variants that were present in the ExAC_ALL cohort with minor allele frequency (MAF) above 0.02. Remaining 946,856 variants filtered according to protein altering affect. In this step, variants that weren't predicted to be damaging by at least two *in silico* tools were excluded, which cut down the number of variants to 29,197. Final step of filtration was the exclusion variants that were found to be in a homozygous state in control samples, leaving 357 variants (Figure 5).

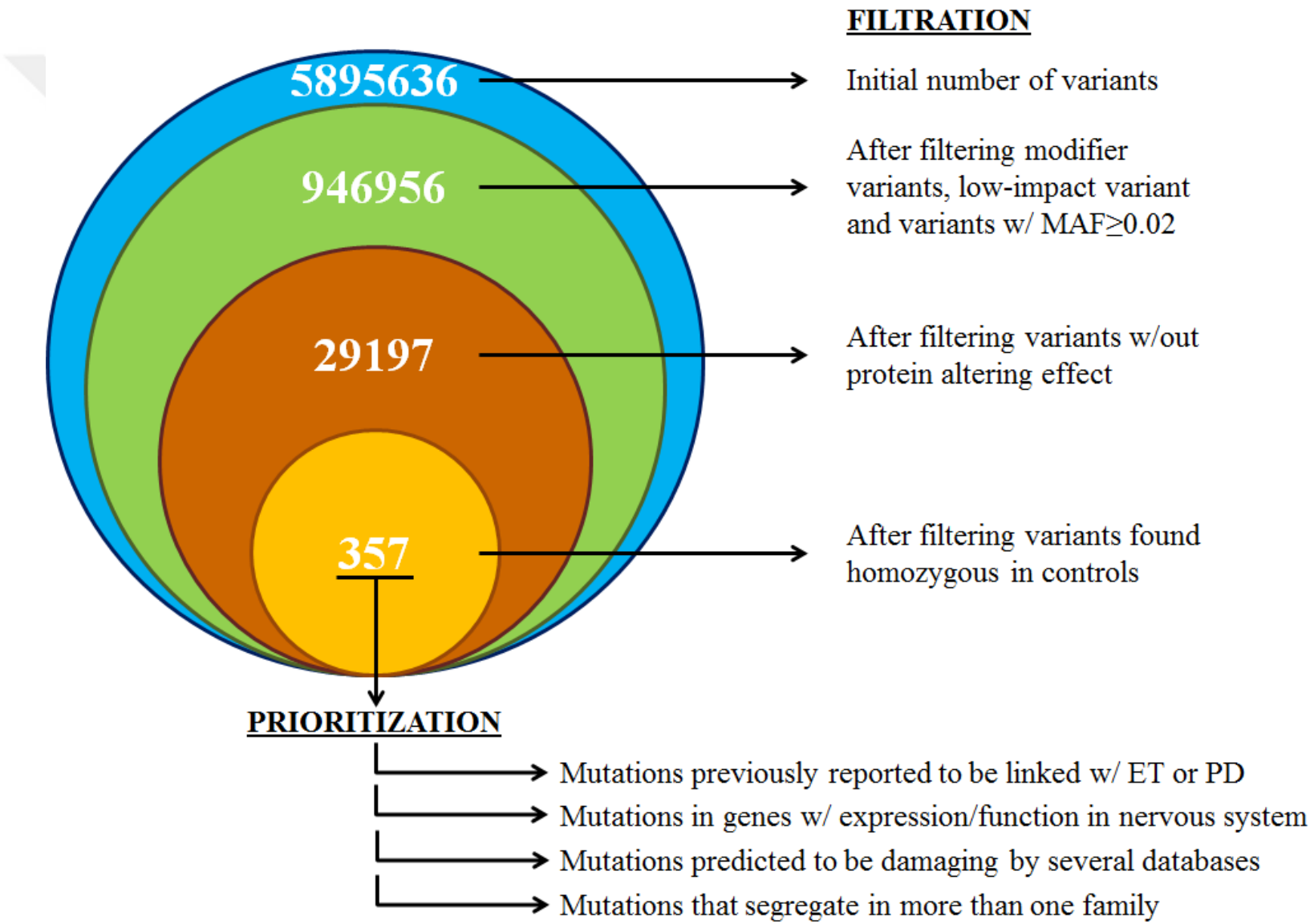


Figure 5 Pipeline for filtration and prioritization of variants. Area of each circle is logarithmically proportionate with the number of variants.

For prioritization, several factors were taken into account. Variants effecting sequences that are not protein-coding, such as retained introns, pseudogenes and lincRNA genes, were deprioritized. We also excluded variants using cohort data, eliminating ones that were homozygous in over 20 samples in ExAC cohort and over 10 samples in our 190-sample in-house database. Variants that had MAF≥0.02 in 1000G, ESP, gnomAD, GME, Kaviar or CG datasets were also excluded.

For the protein damage prediction data, a cut-off value was set for each dataset and variants that scored above the cut-off value in more than one databases were prioritized. Variants scoring below the cut-off on all databases were excluded. Cut-off values (or damage predictions) for each dataset are listed in Table 2.

	Score	Prediction
SIFT	≤2.5	D (Damaging)
Polyphen2 HDIV	≥0.957	D (Probably Damaging)
	≥0.453	P (Possibly Damaging)
Polyphen2 HVAR	≥0.909	D (Probably Damaging)
	≥0.447	P (Possibly Damaging)
LRT	-	D (Damaging)
MutationTaster	-	A (disease causing automatic)
	-	D (disease causing)
MutationAssessor	≥3.505	H (High)
	≥1.945	P (Moderate)
FATHMM	≤-2.5	D (Damaging)
PROVEAN	≤0.05	D (Damaging)
MetaSVM	≤-1.5	D (Damaging)
MetaLR	-	D (Damaging)
M-CAP	-	D (Damaging)
CADD	≥20	-

Table 2 Prioritization criteria for protein damage prediction databases.

After all the prioritization steps were completed, we examined the list of variants to find ones that may fit an inheritance model within each family. Variants that were homozygous or compound heterozygous in control samples were excluded, and variants that were shared by all affected individuals in the family were selected. To search for a variant that is possibly inherited in a recessive manner, we listed variants that were homozygous in probands. Tables 3 and 4 summarize the data about said variants.

Out of the 21 prioritized variants that were homozygous for the proband, 17 were eliminated as they had MAF≥0.02 in different population frequency databases. The 4 remaining variants, *FAM47E-STBD1* p.Pro187Leu, *MSLNL* p.Gly579Arg, *OR8G1* p.Ser289Ile, and *TREML3P* n.93+1delG, were found to be present in several individuals, affected and unaffected, who were a part of our in-house ET cohort WES database consisting of 52 samples from 15 families (Table 3). As a result, we decided to search for a candidate variant that could fit autosomal dominant inheritance, and checked for variants that were heterozygous for the proband. After the elimination using MAF values and the ET WES database, remaining 3 candidates were *TLL2* p.Thr495Met, *SNCAIP* p.Arg853His, and *SRFBP1* p.Leu31Phe (Table 4).

To investigate the cosegregation of the candidate variants with ET in these families, Sanger sequencing and segregation analysis were performed, using primers designed using Primer3Plus (Table A1). Results showed that none of the three candidates cosegregate with ET in this family (Figures 6-8).

Variant Information							Frequency and Protein Damage Data								ET - 17		
Chr	Position	Ref	Alt	Annotation	Gene	Alteration	CG	ESP	GME	SIFT	Polyphen	LRT	MutTas	PROVEAN	V-1	V-5	IV-9
4	77184996	C	T	missense	FAM47E-STBD1	p.Pro187Leu	0.011	0.007	0	T	B	N	D	N	Hom	Hom	Het
16	824837	C	T	missense	MSLNL	p.Gly579Arg	.	0.0096	0.0061	D	D	D	D	D	Hom	Hom	Hom
11	124121288	G	T	missense	OR8G1	p.Ser289Ile	.	.	.	.	.	.	.	.	Hom	Het	Het
6	41190289	C	-	splice	TREML3P	n.93+1delG	.	.	.	.	.	.	.	.	Hom	Hom	Hom
12	51740415	A	C	missense	CELA1	p.Val3Gly	.	.	0.0774	D	B	D	N	N	Hom	Het	Het
17	45400908	G	A	splice	EFCAB13	c.-86+1G>A	0.75	.	.	.	.	.	.	.	Hom	Het	Het
6	32359389	G	A	splice	HCG23	n.245-1G>A	0.054	.	.	.	.	.	.	.	Hom	Hom	Het
6	33373341	C	T	missense	KIFC1	p.Ala490Val	0.043	0.0072	0.0122	T	B	N	D	N	Hom	Hom	Het
22	31325984	C	T	splice	MORC2-AS1	n.186+2C>T	0.43	.	.	.	.	.	.	.	Hom	Hom	Hom
4	4204184	T	C	missense	OTOP1	p.Ile241Val	.	0.0202	0	T	B	N	D	N	Hom	Het	Het
14	67862269	G	A	missense	PLEK2	p.Thr80Met	.	0.0204	0.025	T	P	N	D	N	Hom	Het	Het
9	35752124	G	A	missense	RGP1	p.Gly352Ser	0.033	0.0002	0.0031	T	B	D	D	N	Hom	Het	Het
18	29136518	A	C	splice	RP11-75N4.2	n.349+2T>G	0.55	.	.	.	.	.	.	.	Hom	Hom	Hom
4	5439808	T	A	splice	STK32B	n.163+2T>A	0.5	.	.	.	.	.	.	.	Hom	Hom	Het
22	17265194	G	A	missense	XKR3	p.Pro232Leu	0.42	.	.	T	B	N	P	D	Hom	Hom	Hom
22	17264565	G	T	missense	XKR3	p.His442Asn	0.86	.	.	T	P	U	P	N	Hom	Hom	Hom
3	75779769	C	G	missense	ZNF717	p.Val114Leu	0.48	.	.	D	.	.	N	N	Hom	Het	Het
3	75790513	T	C	missense	ZNF717	p.Tyr64Cys	0.83	.	0.1	T	B	.	P	N	Hom	Hom	Hom

Table 3 List of prioritized variants that were homozygous for the proband of family ET-17. Gray shading: Variants that had MAF≥0.02. Yellow shading: Protein damage prediction "Damaging". Purple Shading: Variants that was present in several individuals from the in-house ET cohort. Hom: Homozygous. Het: Heterozygous. Chr: Chromosome. Pos: Position. Ref: Reference Allele. Alt: Altered Allele. MutTas: MutationTaster.

Variant Information							Frequency and Protein Damage Data								ET - 17		
Chr	Position	Ref	Alt	Annotation	Gene	Alteration	ExAC	ESP	100G	SIFT	Polyphen	LRT	MutTas	PROVEAN	V-1	V-5	IV-9
10	98155678	G	A	missense	TLL2	p.Thr495Met	0.016	0.0173	0.0164	D	D	N	D	D	Het	Het	Het
5	121786959	G	A	missense	SNCAIP	p.Arg853His	0.009	0.0109	0.0064	D	D	D	D	N	Het	Het	Het
5	121309945	C	T	missense	SRFBP1	p.Leu31Phe	0.0021	0.0004	0.0012	D	P	N	D	D	Het	Het	Het
5	140794253	C	A	missense	PCDHGA10	p.Pro504His	0.0662	0.0001	.	D	D	.	D	D	Het	Het	Het

Table 4 List of prioritized variants that were heterozygous for the proband of family ET-17. Gray shading: Variants that had MAF≥0.02. Yellow shading: Protein damage prediction "Damaging". Het: Heterozygous. WT: Wild Type. Chr: Chromosome. Pos: Position. Ref: Reference Allele. Alt: Altered Allele. MutTas: MutationTaster.

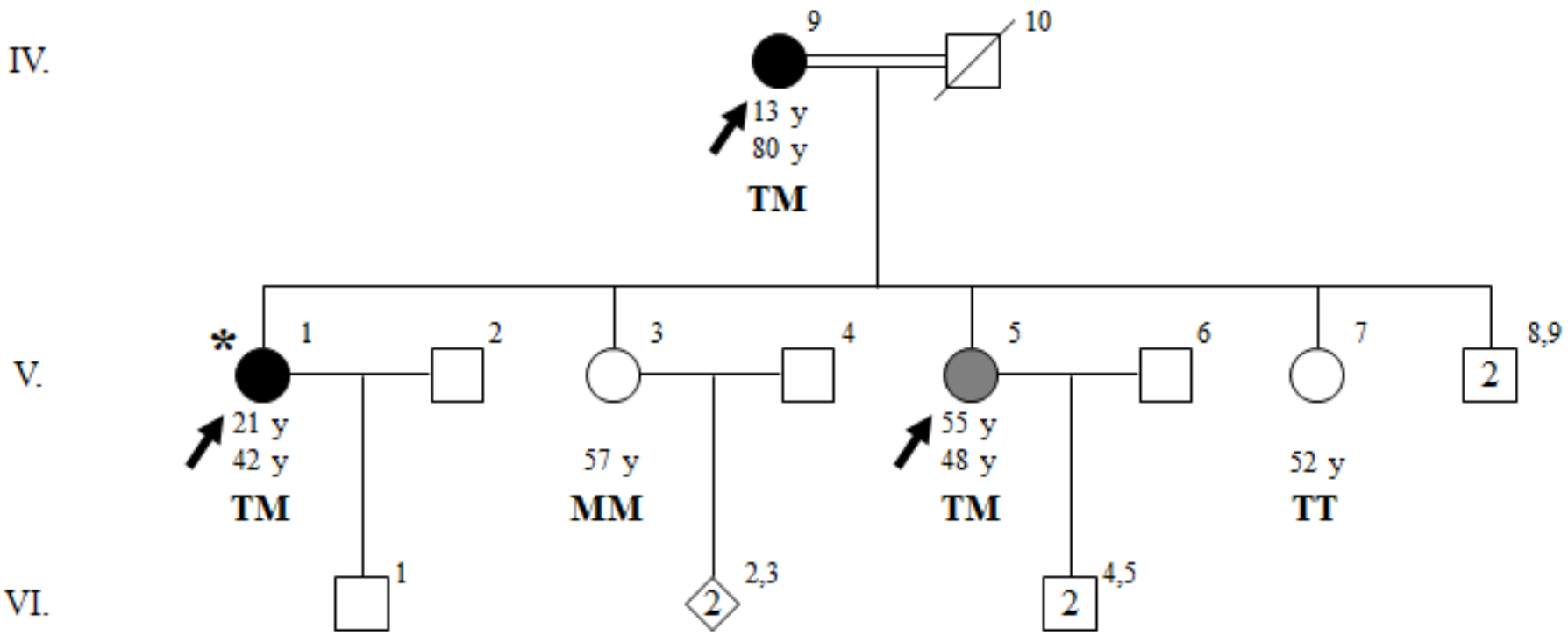


Figure 6 Pedigree of family ET-17 with genotypes at *TLL2* p.T495M. Age at onset of tremor for affected individuals, current ages, and genotypes at *TLL2* p.T495M are indicated in this order under the symbols. T indicates the wild-type allele, threonine; M indicates the variant allele, methionine, at *TLL2* p.T495M. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

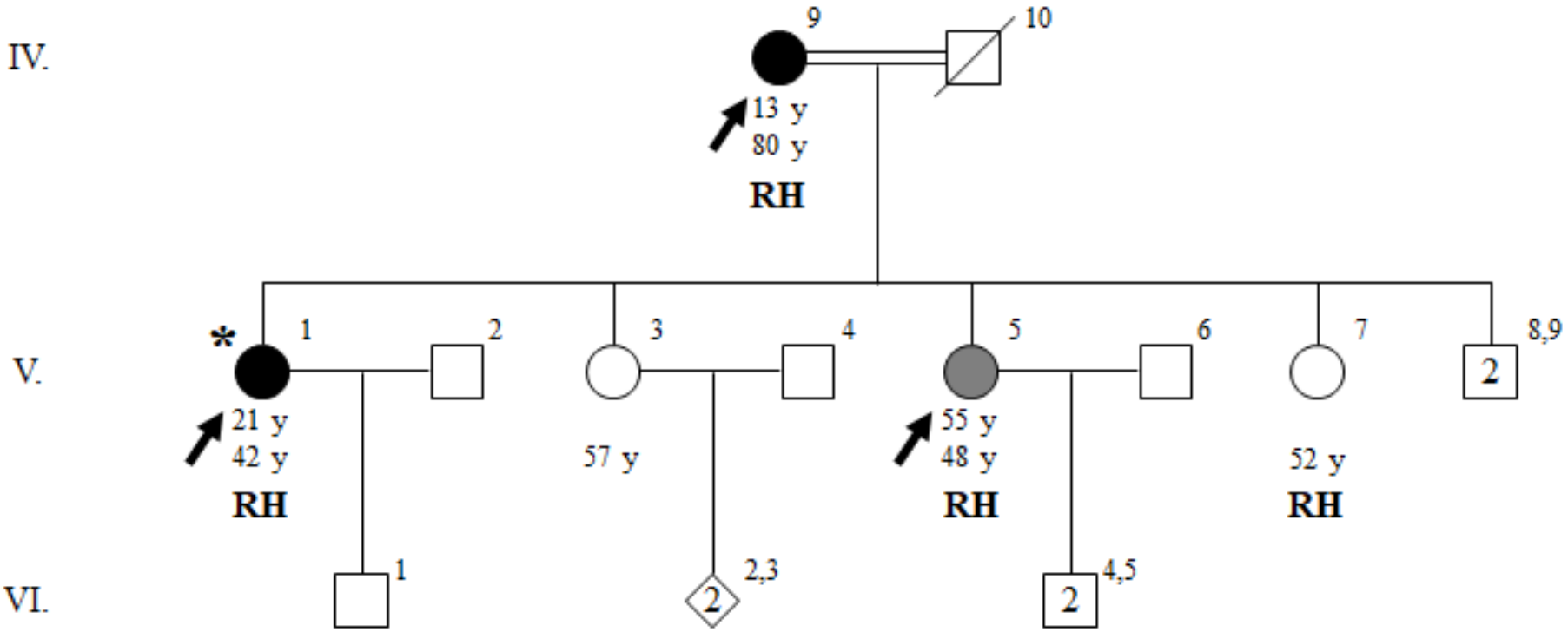


Figure 7 Pedigree of family ET-17 with genotypes at *SNCAIP* p.R853H. Age at onset of tremor for affected individuals, current ages, and genotypes at *SNCAIP* p.R853H are indicated in this order under the symbols. R indicates the wild-type allele, arginine; H indicates the variant allele, histidine, at *SNCAIP* p.R853H. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

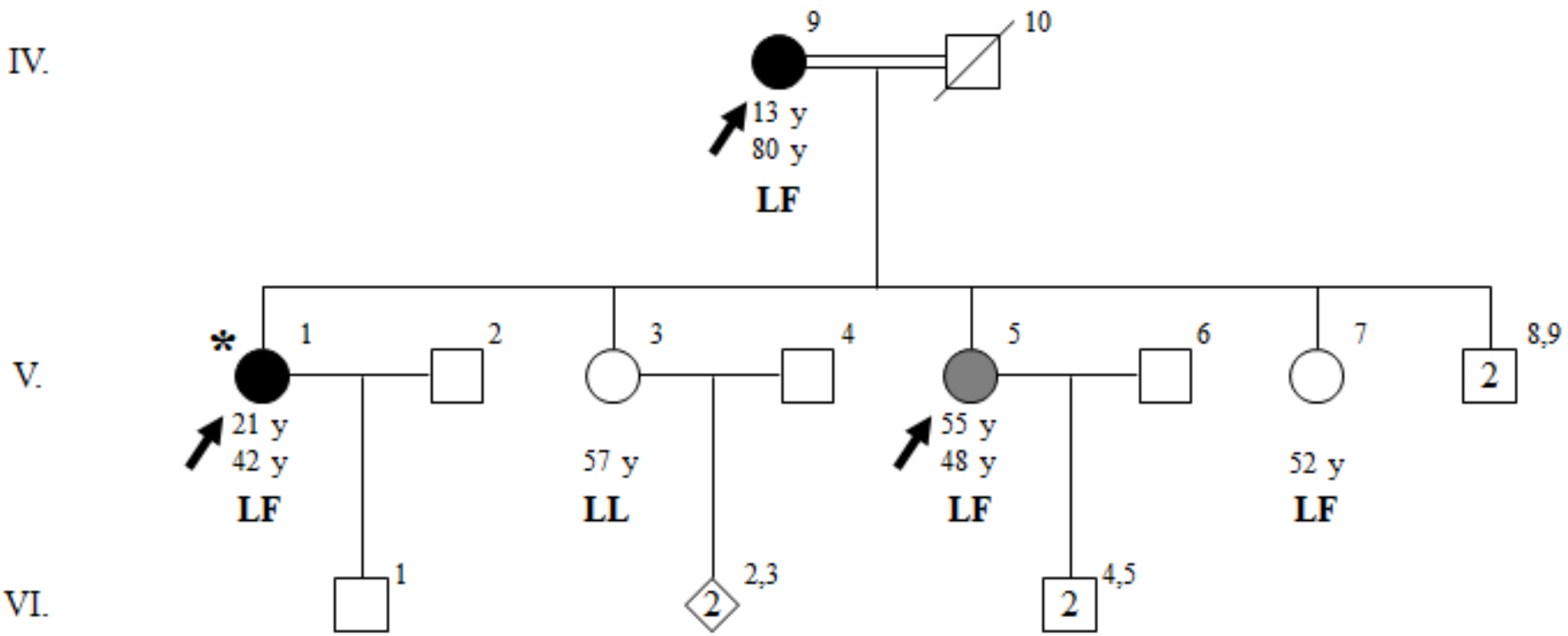


Figure 8 Pedigree of family ET-17 with genotypes at *SRFBP1* p.L31F. Age at onset of tremor for affected individuals, current ages, and genotypes at *SRFBP1* p.L31F are indicated in this order under the symbols. L indicates the wild-type allele, leucine; F indicates the variant allele, phenylalanine, at *SRFBP1* p.L31F. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

3.2 Variant Search in Family ET-19

3.2.1 Clinical Features of ET-19

The clinical diagnosis of ET was made initially in the proband (IV-3) of a four-generation endogamous family, ET-19. ET-19 is of Turkish origin and ET is observed multiple generations. 9 individuals from this family were clinically assessed based on criteria of both consensus statement of MDS[18] and WHIGET[82], and 7 individuals were diagnosed with ET (Figure 9). Archimedes spiral drawing tests are shown in figure 10.

Family ET-19 is from Elazığ, located in east-central Anatolia. Proband of family ET-19 (IV-3) was 25 at the time of diagnosis, and she has shown symptoms since the age of 20. Other affected individuals were her father (III-6) and her paternal grandfather(II-4), both having definite ET, her paternal uncle (III-7) who has moderate ET, and her cousin (IV-4), her half-uncle (III-8) and her half-cousin (IV-5) who have mild ET. The proband also reported that her sister (IV-2) shows tremor symptoms, but this individual didn't want to participate in the study.

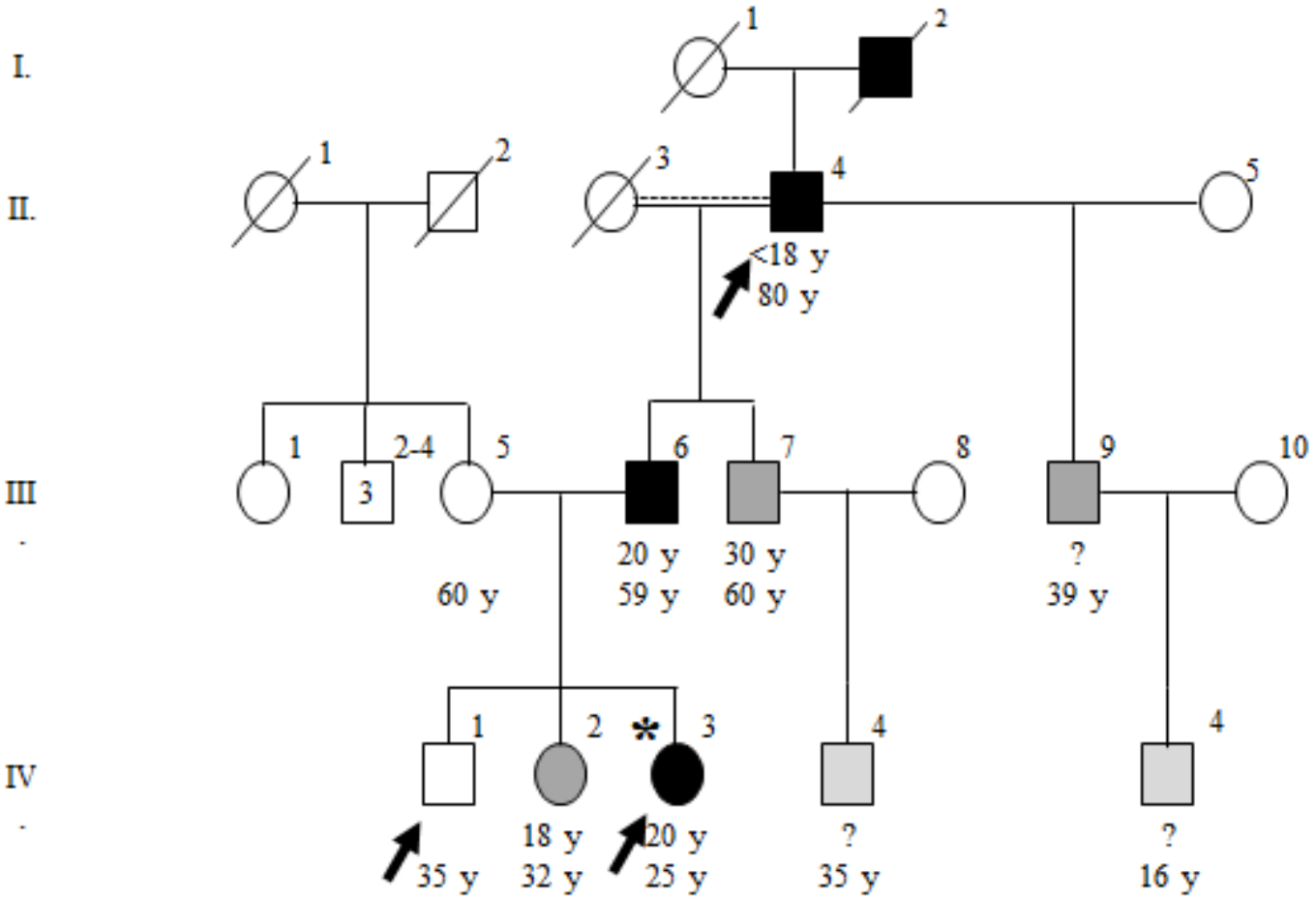


Figure 9 Pedigree of ET-19. Age at onset of tremor for affected individuals and current ages are indicated in this order under the symbols. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

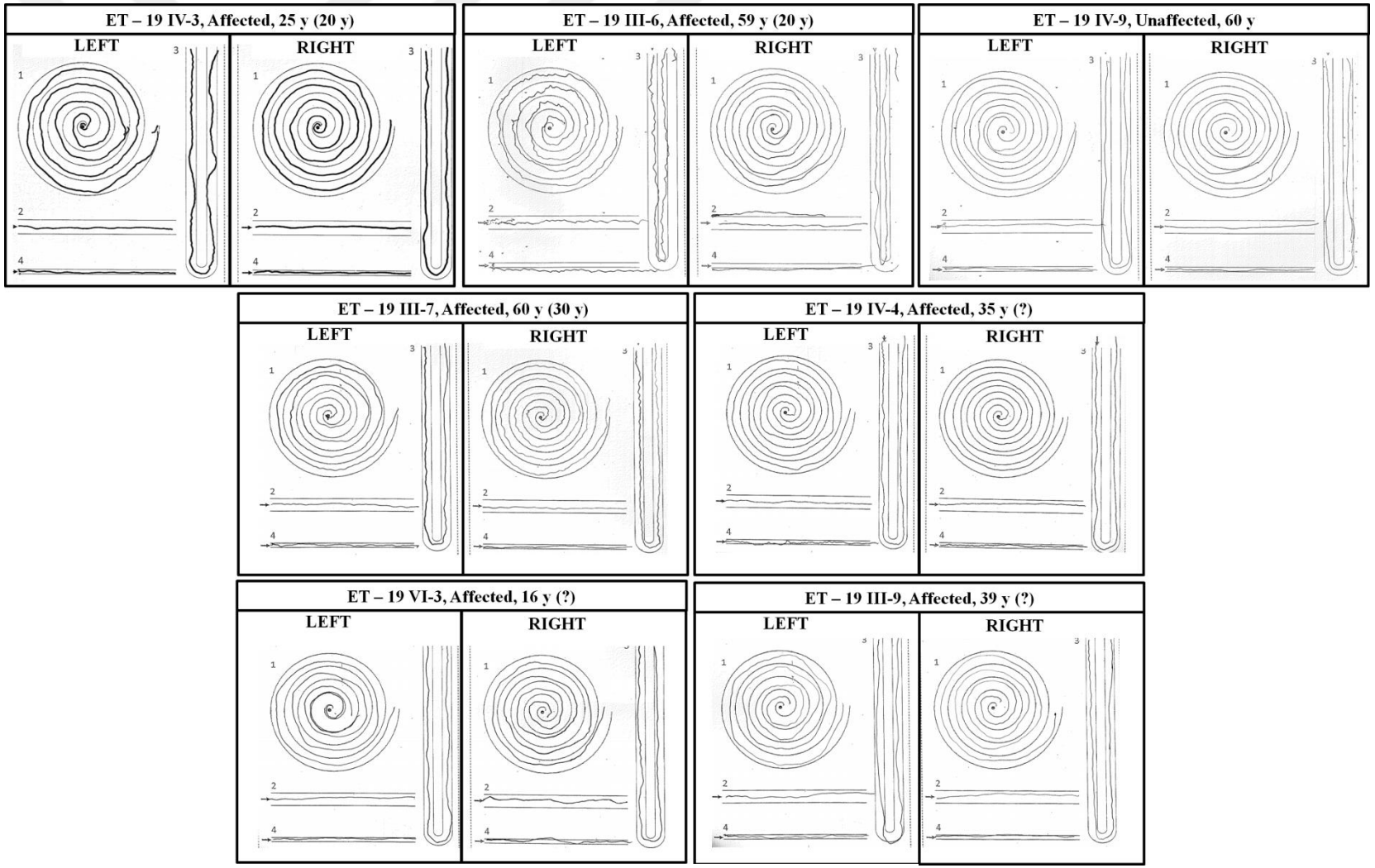


Figure 10 Archimedes spiral drawing test results for members of family ET-19.

3.2.2 Whole Exome Sequencing

Affected family members II-4 and IV-3 along with unaffected IV-1 were selected for WES from ET-19. NanoDropTM ND-1000 Spectrophotometer (NanoDrop Technologies, Inc, DE, USA) was used for the purity and concentration measurements (Table 5). Possible degradation of DNA samples was also checked by running them on agarose gel (Figure 11).

WES resulted in over 49 million reads for each sample. Using Burrows-Wheeler Aligner, ≥99.98% of the reads was mapped to the reference genome, UCSC hg19. SAMtools were used to remove duplicate reads, which consisted ≤1.67% of the total. Exonic coverage was calculated via using BEDtools. Percentage of exonic regions with at least 5-fold coverage was ≥97.48% for each sample (Table 6).

Measurement / DNA Sample	A260/A280 Ratio	A260/A230 Ratio	DNA Concentration (ng/µL)
II-4 (ET-19)	1.92	2.49	255.3
IV-1 (ET-19)	1.91	2.51	325.3
IV-3 (ET-19)	1.88	2.47	195.3

Table 5 Purity and concentration measurements for ET-19 DNA samples.

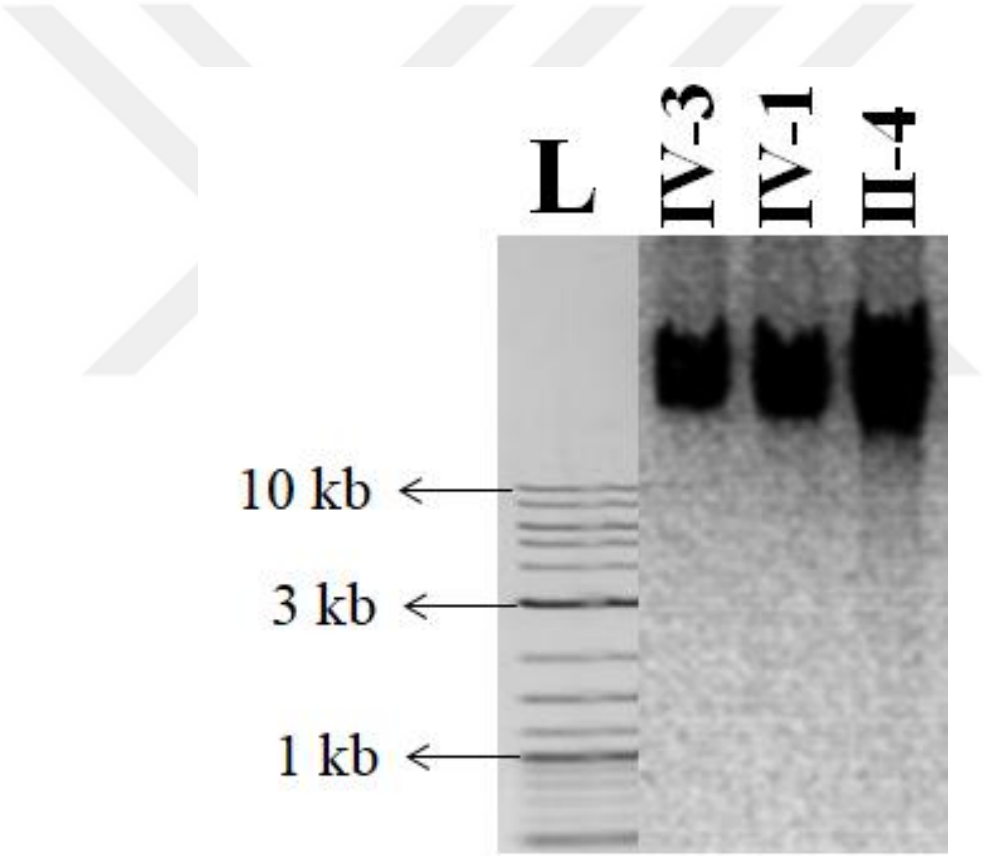


Figure 11 DNA density evaluations for ET-19 WES samples. After samples were diluted 1:5 in TE, they were run on 1% agarose gel for 50 minutes under 70 V. BioRad Gel Doc 2000 was used to capture the image. L: NEB 2-Log Ladder.

DNA Sample	Number of Reads	Mapped Reads	Duplicate Reads	Exonic Regions w/ ≥5x Coverage
II-4 (ET-19)	50913415	50904719 (99.98%)	1.64%	97.72%
IV-1 (ET-19)	54884415	54873761 (99.98%)	1.61%	97.88%
IV-3 (ET-19)	49359617	49352717 (99.99%)	1.57%	97.48%

Table 6 Statistics of whole exome sequencing results for ET-19 samples.

Protein damage prediction data was obtained from SIFT, PolyPhen HDIV, PolyPhen HVAR, MutationAssessor, M-CAP, MutationTaster, LRT, PROVEAN, FATHMM, MetaSVM, MetaLR and CADD databases. MetaSVM showed the highest percentage of predicted benign mutations with 69.35%, while CADD showed the lowest value of 20.76%. For the predicted damaging mutations, highest percentage was 58.51%, predicted by CADD, and the lowest was MetaSVM's 6.05% (Figure 12). 34.7% of the mutations were predicted to be benign by all 12 databases, while 50.8% were predicted to be damaging by more than 2 databases and 19.7% were predicted to be damaging by at least half of the databases (Figure 13).

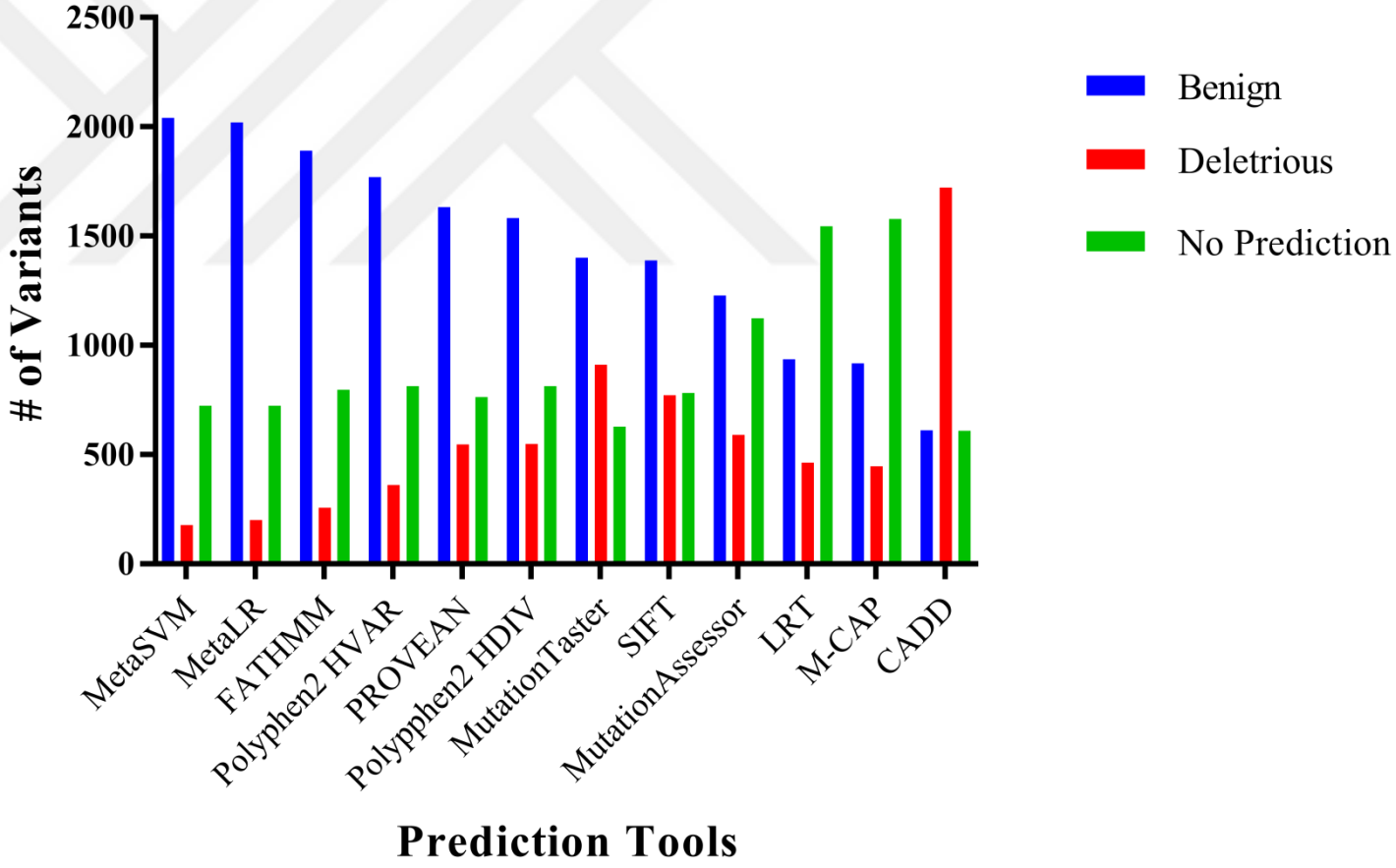


Figure 12 Protein damage prediction for ET-19 variants (1). This plot shows the number of variants each tool predicted to be benign, predicted to be deletrious or possible deletrious, or had no prediction about their effect.

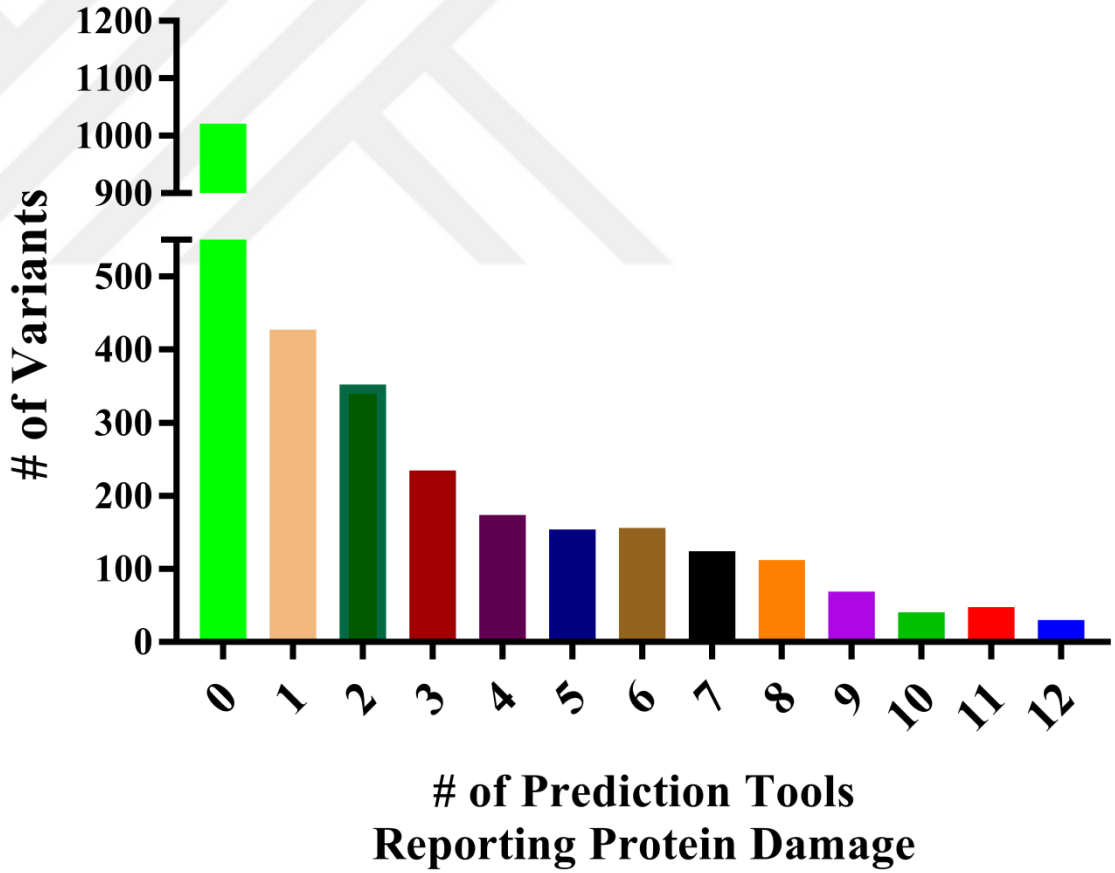


Figure 13 Protein damage prediction for ET-19 variants (2). This plot shows how many of the variants were predicted to be damaging or possibly damaging by how many of the prediction tools out of 7.

3.2.3 Identification of Candidate Variants

After filtration and prioritization steps were completed as explained in Section 3.1.3, candidate variant lists were created. Tables 7 and 8 summarize the data about said variants.

Out of the 4 prioritized variants that were homozygous for the proband, 3 were eliminated as they had MAF≥0.02 in different population frequency databases, and the remaining 1 variant was eliminated due to it being present in several individuals, affected and unaffected, who were a part of our in-house ET cohort WES database (Table 7). As a result, we decided to search for candidate variants that were heterozygous for the proband. After the elimination using MAF values and the ET WES database, remaining 5 candidates were *TCP10L2* p.Arg320*, *ARHGEF4* p.Gly67Trp, *EPHA8* p.Arg879Gln, *SPEN* p.His3315Gln and *TMEM230* p.Arg171Cys (Table 8).

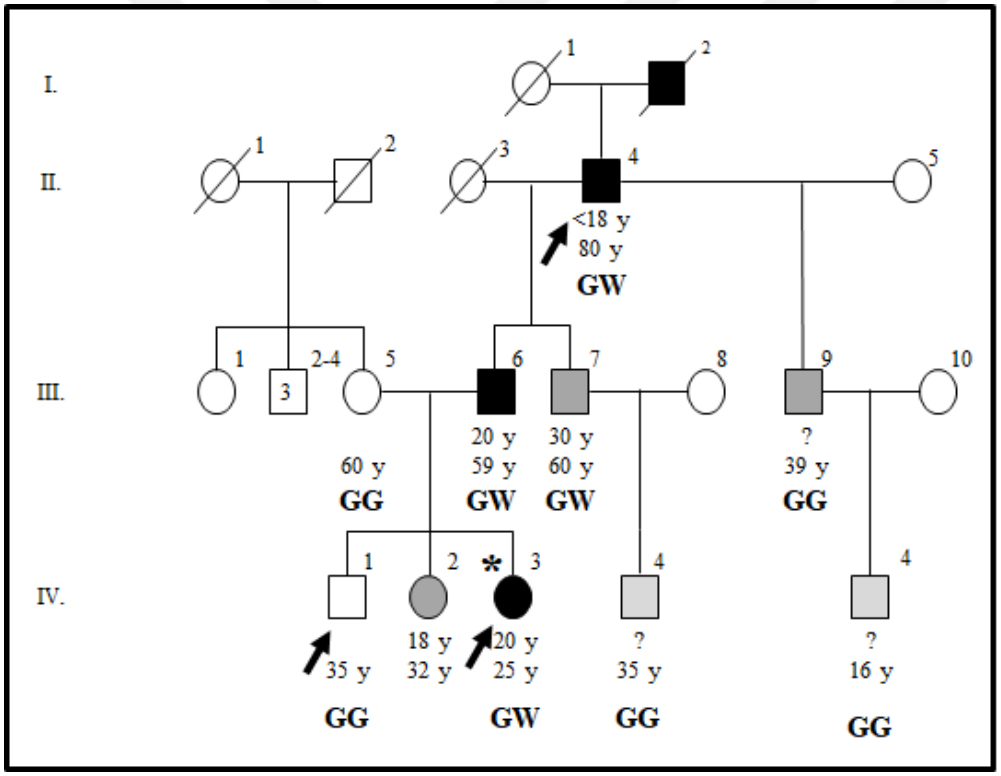
To investigate the cosegregation of the candidate variants with ET in these families, Sanger sequencing and segregation analysis were performed. Results showed 4 of the 5 candidate variants did not cosegregate with the disease in ET-19 (figures 14-15). Only cosegregating variant was *TCP10L2* p.Arg320*. To further verify the relation of the variant with ET, a cohort screening for 62 unrelated ET patients were done. Out of these 62 individuals, all of whom the proband of a different family with hereditary ET, two had the *TCP10L2* p.Arg320* variant in a heterozygous state, probands of families ET-39 and ET-108. We didn't have any samples except the proband's for ET-39, since the family did not want to participate in the study. We performed segregation analysis for ET-108, only to find out the variant did not cosegregate with the disease, thus eliminating it as a candidate (Figure 16).

Variant Information							Frequency and Protein Damage Data								ET - 19		
Chr	**Position**	**Ref**	**Alt**	**Annotation**	**Gene**	**Alteration**	**ExAC**	**GME**	**1000G**	**SIFT**	**Polyphen**	**LRT**	**MutTas**	**PROVEAN**	**IV-3**	**II-4**	**IV-1**
6	31324207	A	C	missense	HLA-B	p.Leu119Arg	.	.	0.1851	D	D	U	L	D	Hom	Hom	Het
19	464140	C	G	missense	ODF3L2	p.Ala192Pro	.	0.1638	.	D	D	N	L	N	Hom	Hom	Het
22	38039746	C	T	missense	SH3BP1	p.Thr190Met	0.0049	0.0237	0.0018	T	B	N	L	N	Hom	Hom	Het
16	64497	C	T	missense	WASH4P	p.Gly440Arg	0	.	.	.	.	.	.	.	Hom	Hom	Het

Table 7 List of prioritized variants that were homozygous for the proband of family ET-19. Gray shading: Variants that had MAF≥0.02. Purple Shading: Variants that was present in several individuals from the in-house ET cohort. Yellow shading: Protein damage prediction "Damaging". Hom: Homozygous. Het: Heterozygous. Chr: Chromosome. Pos: Position. Ref: Reference Allele. Alt: Altered Allele. MutTas: MutationTaster.

Variant Information							Frequency and Protein Damage Data								ET - 19		
Chr	**Position**	**Ref**	**Alt**	**Annotation**	**Gene**	**Alteration**	**ExAC**	**GME**	**1000G**	**SIFT**	**Polyphen**	**LRT**	**MutTas**	**PROVEAN**	**IV-3**	**II-4**	**IV-1**
6	167595300	C	T	stop-gained	TCP10L2	p.Arg320*	0.0089	0.017	0.0078	.	.	.	D	.	Het	Hom	WT
2	131769466	G	T	missense	ARHGEF4	p.Gly67Trp	0.0006	0.0011	0.0003	D	.	.	N	N	Het	Het	WT
1	22927488	G	A	missense	EPHA8	p.Arg879Gln	0.0016	0.011	0.0015	D	D	D	D	D	Het	Het	WT
1	16262680	C	G	missense	SPEN	p.His3315Gln	.	.	.	D	D	.	N	N	Het	Het	WT
20	5081478	G	A	missense	TMEM230	p.Arg171Cys	0.003	0.0098	0.0025	D	D	D	D	D	Het	Het	WT
19	49622210	C	T	missense	C19orf73	p.Val24Met	0.0058	0.018	0.006	.	P	.	N	D	Het	Het	WT
17	39274291	T	C	missense	KRTAP4-11	p.Met93Val	0.02063	0.0001	0.0004	T	B	.	N	N	Het	Het	WT
6	151671747	A	C	missense	AKAP12	p.Ser741Arg	0.0057	0.035	0.0051	D	B	D	N	D	Het	Het	WT
19	46807262	G	A	missense	HIF3A	p.Arg45His	0.0431	0.0001	0.0146	T	D	N	D	N	Het	Het	WT
22	30857611	G	A	missense	SEC14L3	p.Ser281Leu	0.0001	0.0002	0.0775	D	P	D	D	D	Het	Het	WT
1	16069664	T	C	missense	TMEM82	p.Val104Ala	0.0056	0.082	0.0045	D	P	N	D	D	Het	Hom	WT

Table 8 List of prioritized variants that were heterozygous for the proband of family ET-19. Gray shading: Variants that had MAF≥0.02. Purple Shading: Variants that was present in several individuals from the in-house ET cohort. Yellow shading: Protein damage prediction "Damaging". Hom: Homozygous. Het: Heterozygous. WT: Wild Type Chr: Chromosome. Ref: Reference Allele. Alt: Altered Allele. MutTas: MutationTaster.



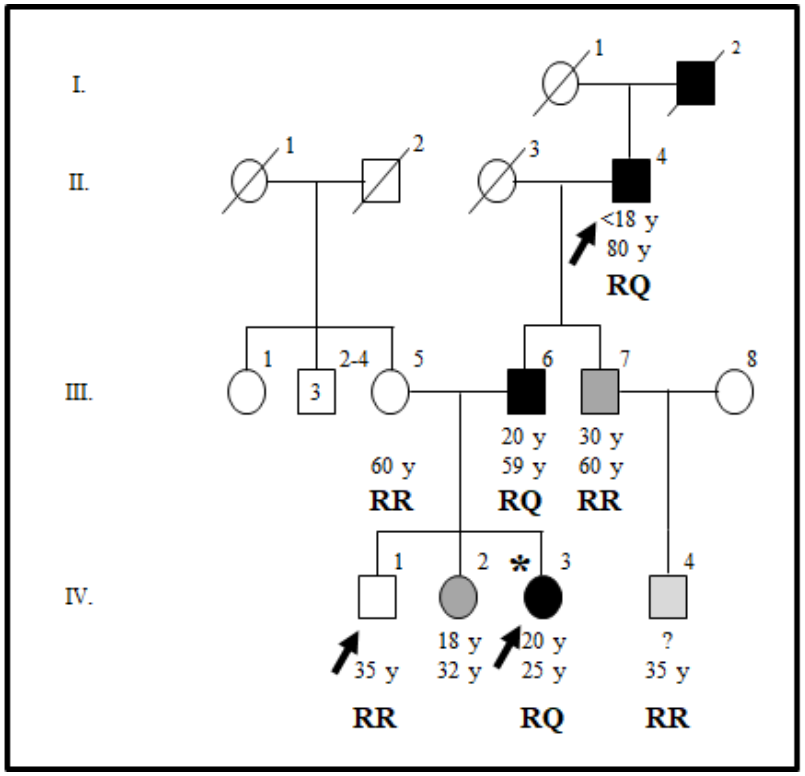


Figure 14 Pedigree of family ET-19 with genotypes at *ARHGEF4* p.Gly67Trp (left) and *EPHA8* p.Arg879Gln (right). Age at onset of tremor for affected individuals, current ages, and genotypes are indicated in this order under the symbols. For *ARHGEF4* p.Gly67Trp, G indicates the wild-type allele, glycine; W indicates the variant allele, tryptophane. For *EPHA8* p.Arg879Gln, R indicates the wild-type allele, arginine; Q indicates the variant allele, glutamine. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

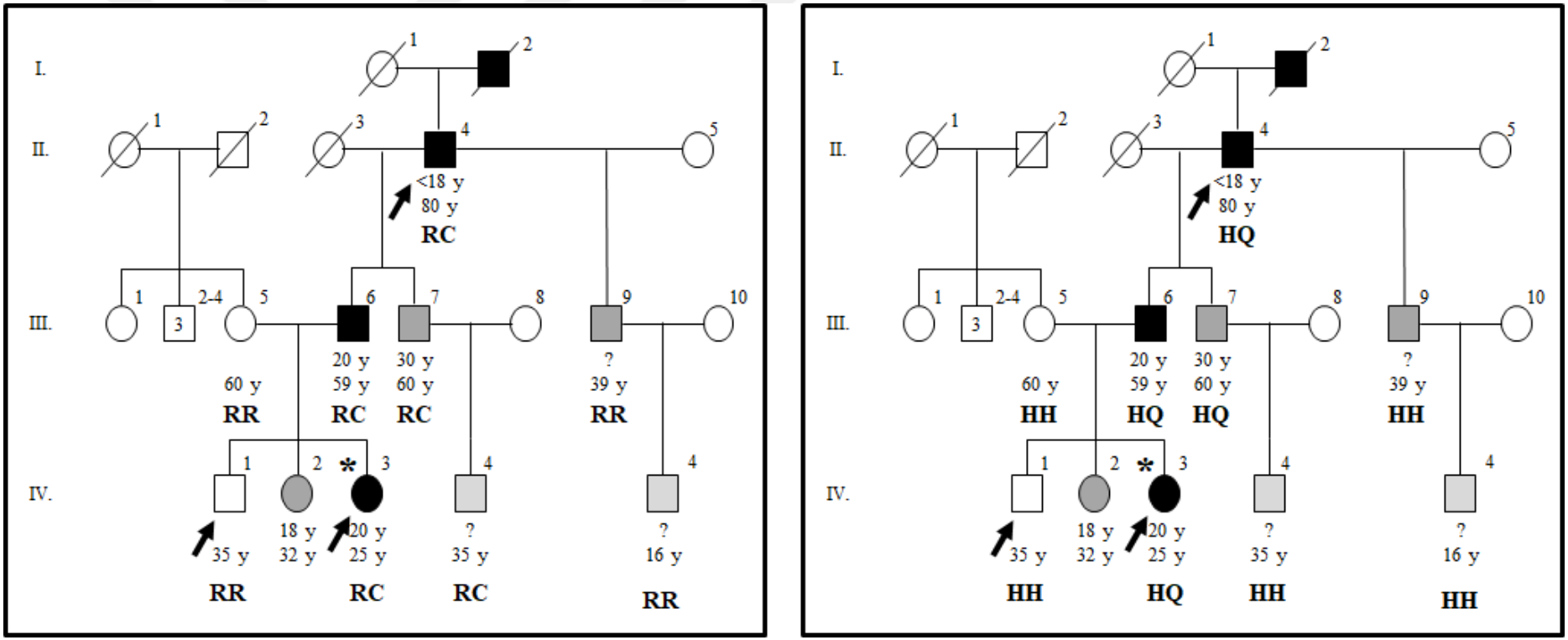


Figure 15 Pedigree of family ET-19 with genotypes at *TMEM230* p.Arg171Cys (left) and *SPEN* p.His3315Gln (right). Age at onset of tremor for affected individuals, current ages, and genotypes are indicated in this order under the symbols. For *TMEM230* p.Arg171Cys, R indicates the wild-type allele, arginine; C indicates the variant allele, cysteine. For *SPEN* p.His3315Gln, H indicates the wild-type allele, histidine; Q indicates the variant allele, glutamine. Individuals who underwent exome sequencing are indicated with arrows. Proband is indicated with an asterisk.

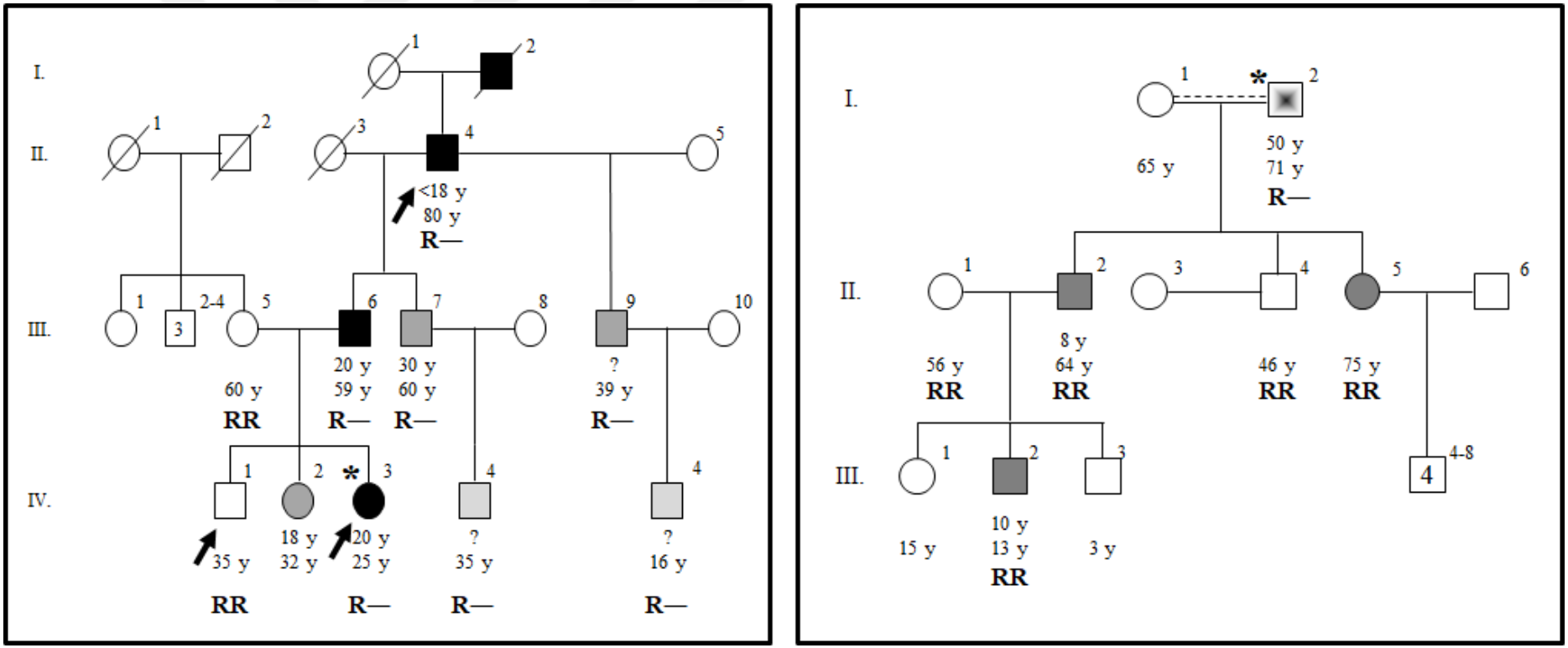


Figure 16 Pedigrees of families ET-19 (left) and ET-108 (right) with genotypes at *TCP10L2* p.Arg320*. Age at onset of tremor for affected individuals, current ages, and genotypes are indicated in this order under the symbols. R indicates the wild-type allele, arginine; "—" indicates the nonsense variant. Individuals who underwent exome sequencing are indicated with arrows. Probands are indicated with asterisks.

3.3 *MMP19* p.R456Q as the Putative Disease Causing Variant in ET-5 and ET-49

3.3.1 Identification of *MMP19* p.R456Q as the Putative Disease Causing Variant in ET-5 and ET-49

A missense variant, *MMP19* p.R456Q, have been previously identified as the putative disease causing variant in two consanguineous and/or endogamous families with ET cases in multiple generations, ET-5 and ET-49 (Figure 17)[103].

MMP19 p.R456Q is located at chr12: 56,230,980 (hg19; c.1367G>A), resulting in an arginine (Arg, R) to glutamine (Gln, Q) substitution at the 456th amino acid residue, coded by of exon 9 of the matrix metalloproteinase-19 (MMP-19, RASI, MMP-18, ENSG00000123342, ENST00000322569) gene in the HX (Hemopexin) superfamily domain (Figure 18b-c). Multiple-sequence alignment of MMP-19 illustrated that the p.R456 residue is conserved in Euteleostomi (Figure 18a), which was also supported by evolutionary conservations scores with a PhastCon score of 0.999 and a GERP++ score of 3.88. Alteration of arginine to glutamine was predicted to be damaging by *in silico* analysis tools, Polyphen2 HDIV, Polyphen2 HVAR, MutationTaster, MutationAssessor and CADD. MAF values showed low frequency in several databases, including ESP (0.0035), 1000G (0.0014), ExAC (0.0037), Kaviar (0.0037), GME (0.0040), PopFreqMax (0.0077) and gnomAD (0.0038). Turkish Peninsula cohort of the GME study also showed a low MAF value with 0.0091. We also genotyped 62 unrelated ET patients, all probands of different families, and none had the *MMP19* p.R456Q variant.

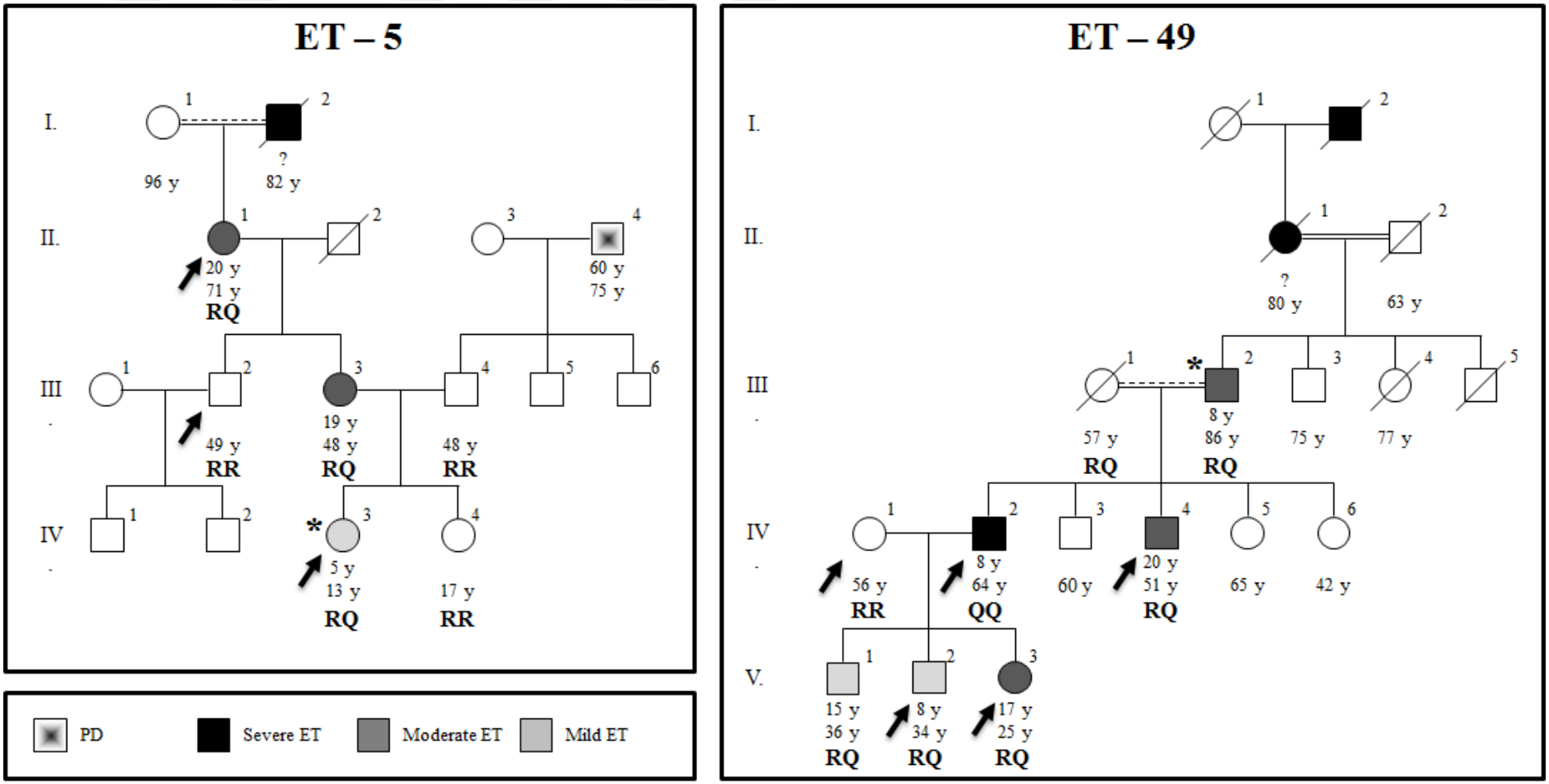


Figure 17 Pedigrees of families ET-5 and ET-49 segregating essential tremor, with genotypes at *MMP19* p.R456Q. Age at onset of tremor for affected individuals, current ages, and genotypes at *MMP19* p.R456Q are indicated in this order under the symbols. R indicates the wild-type allele, arginine; Q indicates the variant allele, glutamine, at *MMP19* p.R456Q. Individuals who underwent exome sequencing are indicated with arrows. Probands are indicated with asterisks.

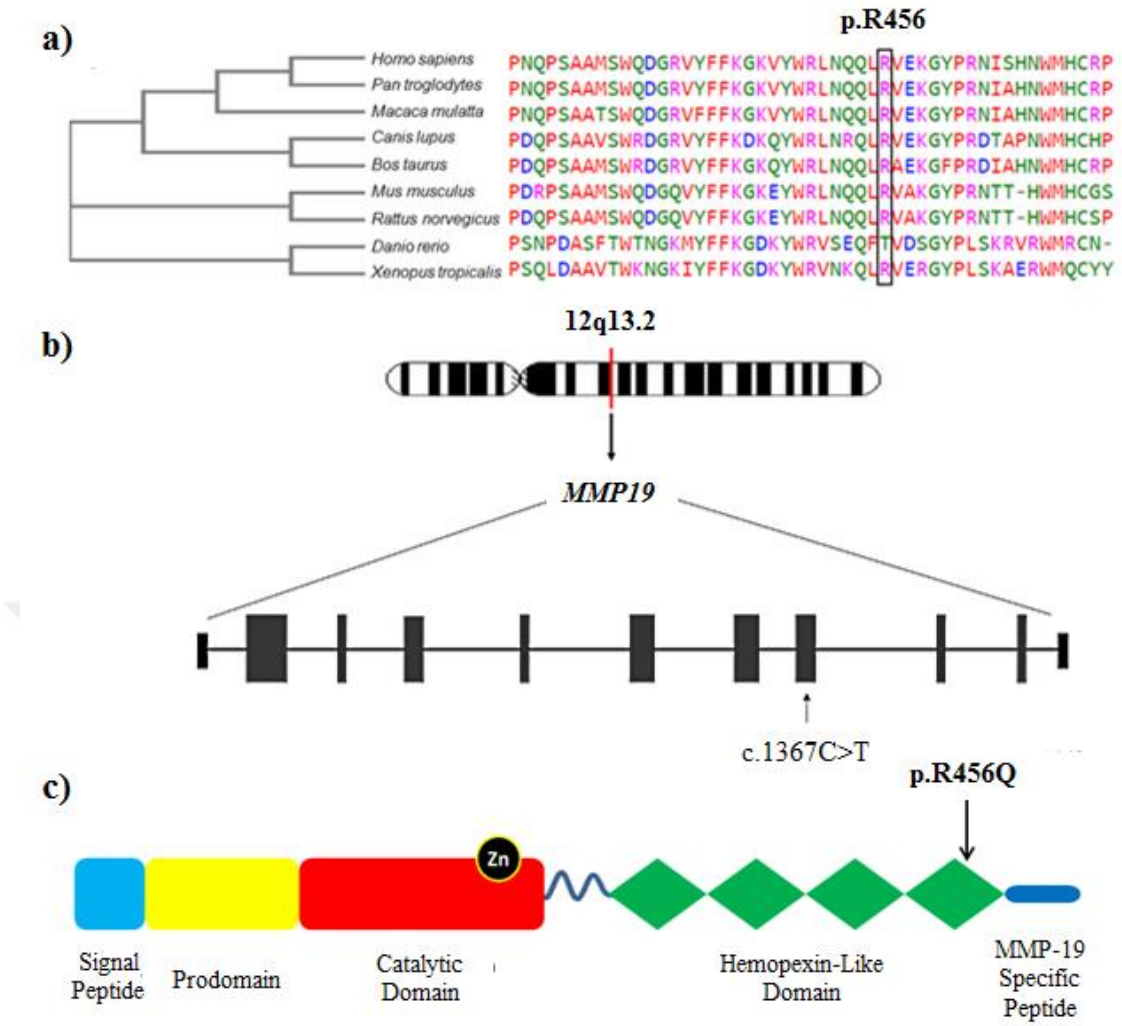


Figure 18 Summary of the alterations in the MMP-19 protein (from Şen 2016 with permission[103]). a) Sequence homology of MMP-19 protein p.R456 region among various species. The box indicates the mutant amino acid p.R456. b) Chromosomal location (top) and gene structure (middle) of *MMP19*. c) The domain structure of the MMP-19 protein, including the signal peptide, the pro-domain and catalytic domain containing zinc ion binding site, hinge region and hemopexin domain (bottom). The p.R456Q missense alteration is in the hemopexin domain.

Effect of the variant on the protein structure was examined by constructing a 3D model (Figure 19). Prediction showed an alteration on the electrostatic interactions on the molecular surface of the protein, specifically the forming of a new hydrogen bond between Gln454 and Gln456 residues, latter being the new residue caused by the mutation. Replacement of the positively charged arginine side chain with the polar neutral glutamine side chain seems to open the possibility of hydrogen bonding with other polar side chains.

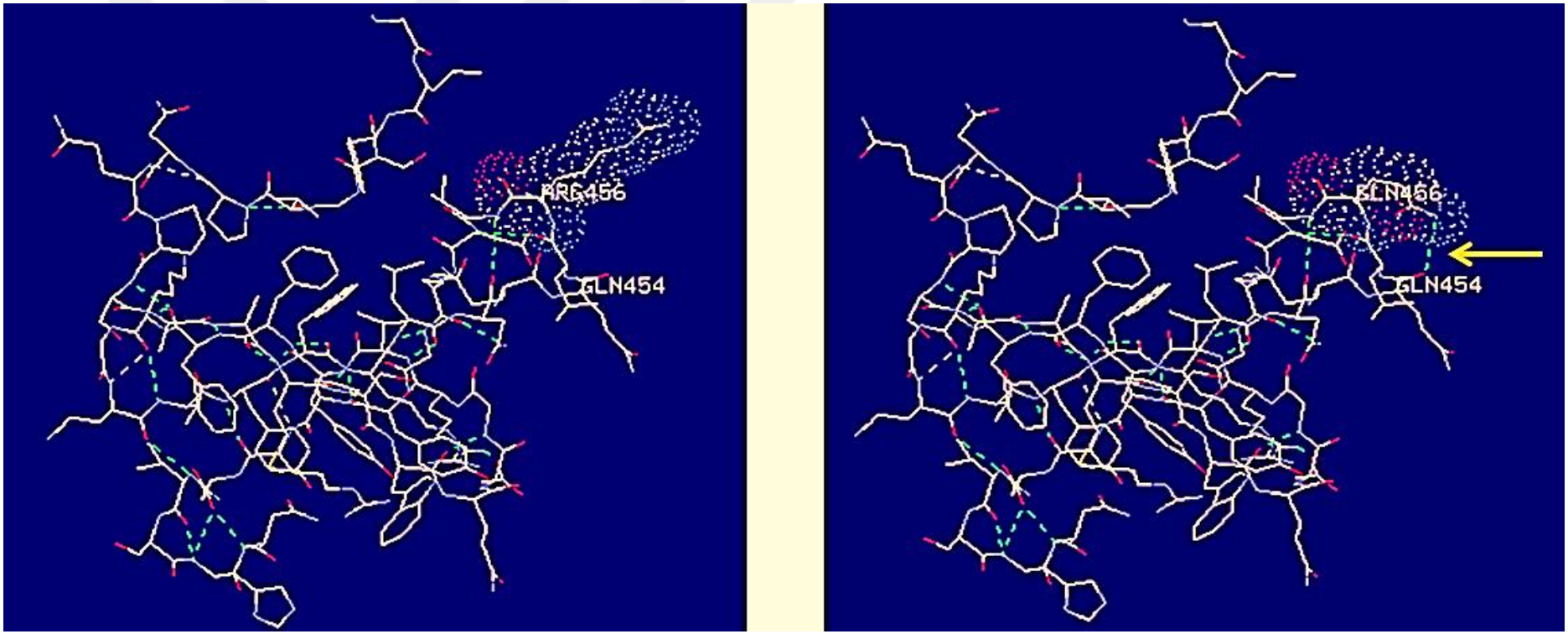


Figure 19 Model 3D protein structure of MMP-19 protein (from Şen 2016 with permission[103]). (Left) Predicted 3D structure of wildtype MMP-19 protein. (Right) Predicted 3D structure of mutant MMP-19 protein. Conversion of p.R456Q results in the formation of a hydrogen bond (arrow) between Gln456 and Gln454.

3.3.2 Expression Analysis

In order to analyze mRNA expression levels from multiple C57B16 mouse tissues (brain, eye, kidney, liver, lung, pancreas, spleen and stomach) and brain tissues of animals from different developmental stages (E15, P1, P7, P11, P15, young adult and aged), gene expression profiles were assessed by quantitative RT-PCR. MMP-19 mRNA was most highly expressed in brain which showed statistically significant difference compared to other tissues (Figure 20a). Moderate level MMP-19 expression was also observed in kidney and liver. Compared to brain tissue, eye, lung, pancreas, spleen and stomach showed little expression. MMP-19 expression was further evaluated among brain tissues from different developmental stages. MMP-19 mRNA expression appeared to be developmentally regulated since it was expressed in high amounts at E15, while showing a significant decline at P1, and increased expression until adulthood. Nevertheless, MMP-19 expression declined when the animals reached adulthood and was downregulated in aged animals (Figure 20).

MMP-19 expression pattern within the brain was examined by *in situ* hybridization. Expression of *Mmp19* was observed with anti-sense probe in granular layer of cerebellum and some regions of hippocampus and thalamus of adult mouse brain sections (Figure 21).

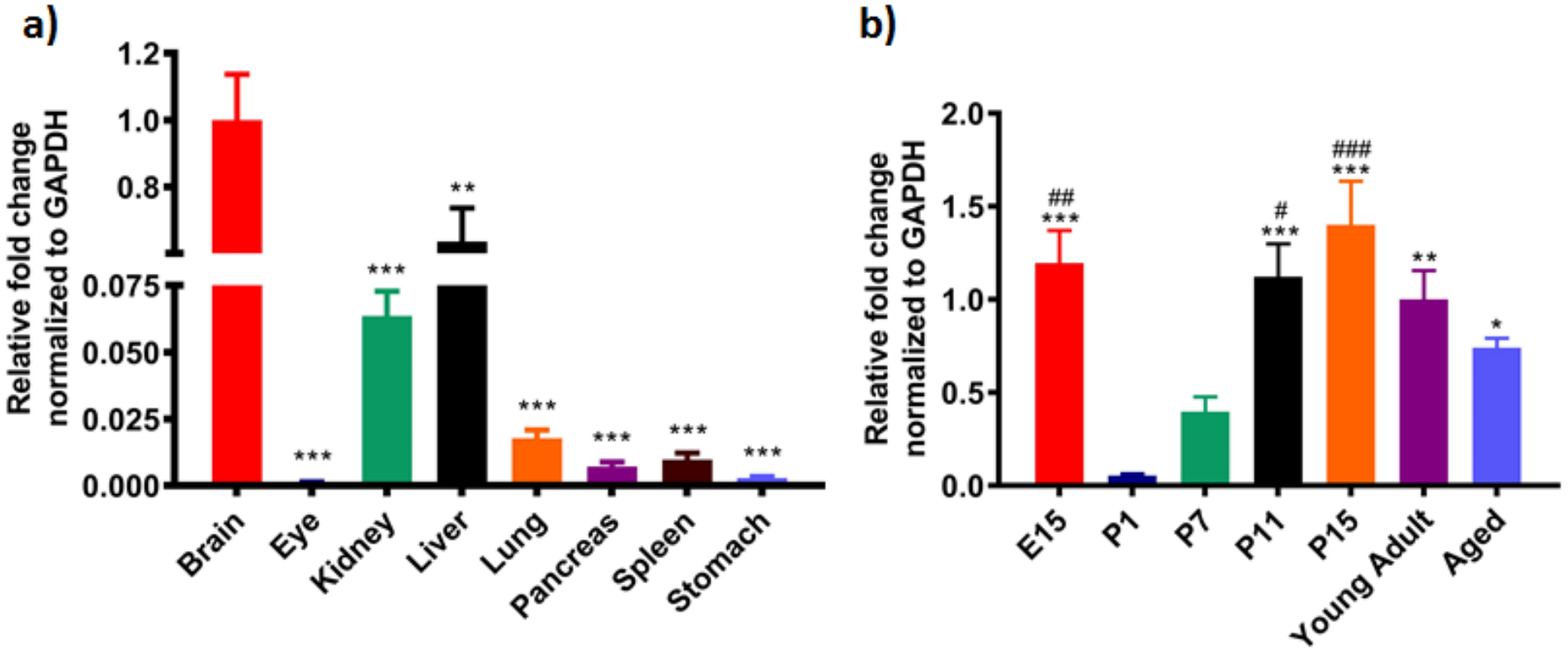


Figure 20 Expression pattern of MMP19. (a) Gene expression analysis of MMP-19 in different organs of adult C57B16 mouse. The expression level of MMP-19 was normalized to GAPDH. Values represent mean ± SEM. Statistical analyses were performed by one way ANOVA followed by Bonferroni posttest (**p<0.01 and ***p<0.001 vs. brain). (b) Gene expression analysis of MMP-19 in different developmental stages of C57B16 mouse. Values represent mean ± SEM. Statistical analyses were performed by one way ANOVA followed by post Bonferroni test (*p<0.05, **p<0.01 and ***p<0.001 vs. P1; #p<0.05, ##p<0.01 and ###p<0.001 vs. P7).

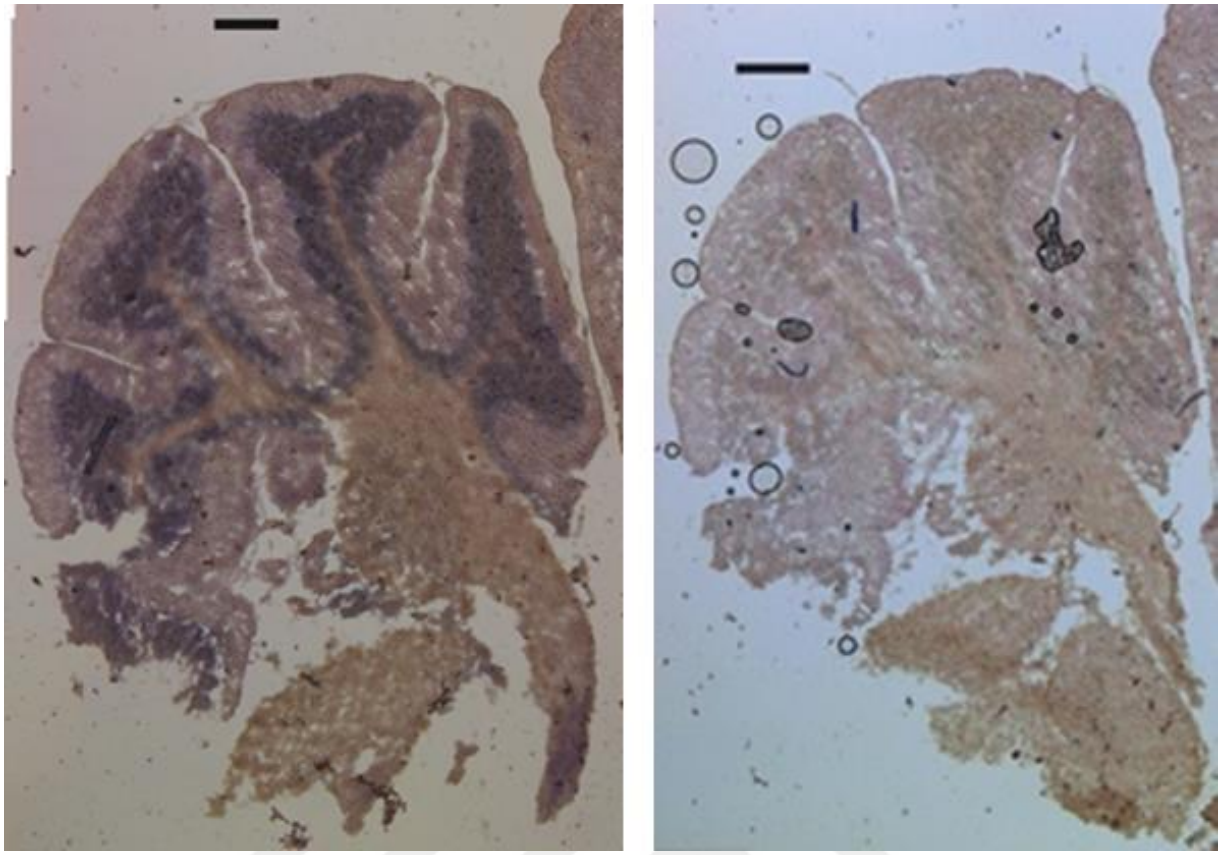


Figure 21 Expression of *Mmp19* in adult mouse brain. (left) *In situ* hybridization of adult mouse brain sections revealing increased expression of *Mmp19* in granular of cerebellum and some parts of hippocampus (right) No hybridization was observed with the sense probe.

CHAPTER 4

Conclusion and Future Perspectives

Despite being most prevalent one, ET is one of the least studies movement disorders, partly due to heterogeneity of its phenotype and etiology. The rising popularity of WES studies in recent years lead to an increase in the number of studies on the genetics of ET, however the complete picture has not yet been revealed and there is still much to be discovered. This thesis summarizes our efforts to identify the novel exonic variants that potentially have a causal link with ET. We utilized WES to examine the inheritance of ET in three multi-generation families.

For ET-17 and ET-19, we were unable to identify a variant that cosegregate with ET. This is most likely be caused by one of two reasons: either we are losing the causative variant during filtration and prioritization steps, or the genetic factors in these families are non-exonic and we are losing them because of the WES approach. For the first possible reason, re-canvassing of the prioritization process is the first measure that can be taken, as this is the most subjective step of the variant selection. By casting a wider net by preparing a longer "short-list", we can increase the possibility of finding the causative variant. We can also re-do the filtration steps, as mistakes on these steps are also possible and might cause the loss of valuable variants from the candidate group. For the second possible reason, the solution is to use a different approach. Whole genome sequencing should be the first choice in this case, as it is similar to WES in its core, but is much more thorough.

ET-5 and ET-49 are from western and central Anatolia, respectively. Both families show ET across multiple generations. We identified a rare mutation, *MMP19*

p.R456Q, which shows cosegregation with ET across generations. We identified *MMP19* c.1367C>T [p.R456Q] in all affected members of these families, whether in a homozygous or a heterozygous form. The rare missense variant is reported in variome databases with low frequency, with MAF=0.003692 in ExAC, 0.0014 in 1000G and 0.0035 in ESP databases. Altered R456 residue is conserved within Tetrapoda, multiple-sequence alignment showed. Alteration causes the loss of arginine side chain containing positively charged guanidium group. 3D modeling analysis suggested a change in the hydrogen-bonding pattern of the molecule, supporting prediction of *in silico* tools as to mutation being damaging to the protein structure.

MMP19 encodes matrix metalloproteinase-19 (MMP-19), a zinc-dependent extracellular matrix endopeptidase. As a member of the MMP family, MMP-19 degrades extracellular matrix (ECM) proteins such as type-IV collagen, gelatin, fibronectin and aggrecan, as well as laminin, a protein prominent in brain vascular matrix[104,105].

Matrix metalloproteinases (MMPs) are the extracellular proteases essential for the breakdown of extracellular matrix components under normal conditions such as embryonic development and tissue remodeling[106]. These roles are reflected in their involvement with the nervous system, taking part in developmental processes such as axonal guidance and neurogenesis as well as synaptic plasticity and repairing mechanisms. They are also reported to have involvement in learning and memory[107–110]. They are also involved in several brain-related conditions, Alzheimer's disease, multiple sclerosis (MS), ischemia, as well as PD, recent reports suggest[111–115]. MMP-19's involvement MS is also suggested as it is shown to be substantial within MS lesions by an immunohistochemistry study[116]. Cell culture studies show MMP-19

expression in brain cells, such as primary microglia and astrocytes, astrocytoma/glioma cells and adult sensory neuron populations[117,118].

Along with a signal peptide for secretion and the zinc-binding catalytic domain, MMP-19 has a hemopexin-like domain that has a role in substrate recognition, similar to most MMPs. This domain also has a role in the regulation of the catalytic activity of MMPs by tissue inhibitors of metalloproteinases (TIMPs). TIMPs -2, -3 and -4, have a strong inhibitory on MMP-19, while TIMP-1's effect is less efficient[119]. Located in the hemopexin-like domain, alteration of R456 is expected to have an effect on the regulation of the catalytic activity of MMP-19 as well as its substrate recognition activity; especially because this alteration results in the loss of a charged group, and this ought to effect the noncovalent interaction between MMP-19 and other proteins, including TIMPs and ECM components. Predictive 3D models of the mutant protein suggested the formation of a new hydrogen bond, further supporting the effect on protein-protein interaction.

Expression of MMP-19 was detected in several mouse tissues with varying degrees in our study while the highest expression was observed in brain. The detection of relatively high levels of MMP-19 expression in a wide variety of normal young adult mouse tissues suggests that this enzyme could play a role in matrix remodeling processes in all tissues. Also, our results showed that MMP-19 mRNA expression was found to be fluctuated during different developmental stages. Due to the predicted effect of MMP-19 in matrix remodeling, variation of mRNA expression may be referred to the role of MMP-19 in development.

Our *in situ* hybridization results showed expression in granular layer of cerebellum, as well as some parts of hippocampus and thalamus of adult mouse brain sections.

Cerebellar expression is in parallel with the results of previous pathophysiology studies, showing cerebellar involvement in ET[10,11]. Expression in hippocampus and thalamus suggests a possible link to memory. We suggest that cognitive and memory decline be included in the clinical assessment of ET patients in future studies.

Studies on the function and expression of MMP-19 are limited. Although the *in silico* predictive data suggest protein damage caused by this mutation, functional studies are need to be done for further proving the causal link between the identified variant and ET.

BIBLIOGRAPHY

1. 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).

2. Louis, E. D., Gerbin, M. & Galecki, M. Essential tremor 10, 20, 30, 40: clinical snapshots of the disease by decade of duration. *Eur. J. Neurol.* **20,** 949–954 (2013).

3. Elble, R. J. Diagnostic criteria for essential tremor and differential diagnosis. *Neurology* **54,** S2-6 (2000).

4. 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).

5. Chandran, V. & Pal, P. K. Essential tremor: Beyond the motor features. *Park. Relat. Disord.* **18,** 407–413 (2012).

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. Deuschl, G., Elble, R. & Elbe, R. Essential tremor-neurodegenerative or nondegenerative disease towards a working definition of ET. *Park. Mov Disord* **24,** 2033–2041 (2009).

10. Quattrone, A. *et al.* Essential head tremor is associated with cerebellar vermisatrophy: a volumetric and voxel-based morphometry MR imaging study. *AJNR Am J Neuroradiol* **29,** 1692–1697 (2008).

11. Louis, E. D. From neurons to neuron neighbothoods: the rewiring of the cerebellar cortex in essential tremor. *Cerebellum* **13,** 501–512 (2014).

12. Louis, E. D. *et al.* Risk of tremor and impairment from tremor in relatives of patients with essential tremor: a community-based family study. *Ann Neurol* **49,** 761–769 (2001).

13. Louis, E. D., Hernandez, N., Rabinowitz, D., Ottman, R. & Clark, L. N. Predicting age of onset in familial essential tremor: how much does age of onset run in families? Neuroepidemiology. 2013;40(4):269–273.Louis ED, Hernandez N, Rabinowitz D, Ottman R, Clark LN. Predi, Clark LN. Predicting age of onset in familial essential. *Neuroepidemiology* **40,** 269–273 (2013).

14. Louis, E. D. Factor analysis of motor and nonmotor signs in essential tremor: are these signs all part of the same underlying pathogenic process? *Neuroepidemiology* **33,** 41–46 (2009).

15. Zesiewicz, T. A. *et al.* A double-blind placebo-controlled trial of zonisamide (zonegran) in the treatment of essential tremor. *Mov Disord* **22,** 279–282 (2007).

16. Gironell, A. *et al.* A randomized placebo-controlled comparative trial of gabapentin and propranolol in essential tremor. *Arch. Neurol.* **56,** 475–80 (1999).

17. Louis, E. D. 'Essential tremor' or 'the essential tremors': is this one disease or a family of diseases? *Neuroepidemiology* **42,** 81–89 (2014).

18. Deuschl, G., Bain, P. G. & Brin, M. F. Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. *Mov. Disord.* **13 Suppl 3,** 2–23 (1998).

19. Louis, E. D. & Ferreira, J. J. How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. *Mov. Disord.* **25,** 534–541 (2010).

20. Tanner, C. M. *et al.* Essential tremor in twins: an assessment of genetic vs environmental determinants of etiology. *Neurology* **57,** 1389–1391 (2001).

21. Lorenz, D. *et al.* High concordance for essential tremor in monozygotic twins of old age. *Neurology* **62,** 208–211 (2004).

22. 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,** 1783–1794 (2010).

23. Louis, E. D., Zheng, W., Applegate, L., Shi, L. & Factor-Litvak, P. Blood harmane concentrations and dietary protein consumption in essential tremor. *Neurology* **65,** 391–396 (2005).

24. 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,** 1170–1177 (2005).

25. Gulcher, J. R. *et al.* Mapping of a familial essential tremor gene, FET1, to chromosome 3q13. *Nat. Genet.* **17,** 84–87 (1997).

26. Higgins, J. J., Pho, L. T. & Nee, L. E. A gene (ETM) for essential tremor maps to chromosome 2p22-p25. *Mov. Disord.* **12,** 859–864 (1997).

27. Shatunov, A. *et al.* Genomewide scans in North American families reveal genetic linkage of essential tremor to a region on chromosome 6p23. *Brain* **129,** 2318–2331 (2006).

28. Kovach, M. J. *et al.* Genetic heterogeneity in autosomal dominant essential tremor.

29. Illarioshkin, S. N. *et al.* Molecular Genetic Analysis of Essential Tremor. *Russ. J. Genet.* **38,** 1447–1451 (202AD).

30. Lucotte, G., Lagarde, J.-P., Funalot, B. & Sokoloff, P. Linkage with the Ser9Gly DRD3 polymorphism in essential tremor families. *Clin. Genet.* **69,** 437–440 (2006).

31. 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,** 972–977 (1998).

32. Aridon, P. *et al.* Further evidence of genetic heterogeneity in familial essential tremor. *Parkinsonism Relat. Disord.* **14,** 15–18 (2008).

33. Novelletto, A. *et al.* Linkage exclusion in Italian families with hereditary essential tremor. *Eur. J. Neurol.* **18,** e118-20 (2011).

34. Zahorakova, D. *et al.* No association with the ETM2 locus in Czech patients with familial essential tremor. *Neuro Endocrinol. Lett.* **31,** 549–552 (2010).

35. Tan, E.-K. *et al.* DRD3 variant and risk of essential tremor. *Neurology* **68,** 790–791 (2007).

36. Vitale, C. *et al.* DRD3 Ser9Gly variant is not associated with essential tremor in a series of Italian patients. *Eur. J. Neurol.* **15,** 985–987 (2008).

37. Higgins, J. J. *et al.* HS1-BP3 gene variant is common in familial essential tremor. *Mov. Disord.* **21,** 306–309 (2006).

38. Higgins, J. J. *et al.* A variant in the HS1-BP3 gene is associated with familial essential tremor. *Neurology* **64,** 417–421 (2005).

39. Deng, H. *et al.* Extended study of A265G variant of HS1BP3 in essential tremor and Parkinson disease. *Neurology* **65,** 651–652 (2005).

40. Stefansson, H. *et al.* Variant in the sequence of the LINGO1 gene confers risk of essential tremor. *Nat. Genet.* **41,** 277–279 (2009).

41. Deng, H., Gu, S. & Jankovic, J. LINGO1 variants in essential tremor and Parkinson's disease. *Acta Neurol. Scand.* **125,** 1–7 (2012).

42. 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).

43. Thier, S. *et al.* Polymorphisms in the glial glutamate transporter SLC1A2 are associated with essential tremor. *Neurology* **79,** 243–248 (2012).

44. Kuhlenbäumer, G., Hopfner, F. & Deuschl, G. Genetics of essential tremor: meta-analysis and review. *Neurology* **82,** 1000–1007 (2014).

45. Merner, N. D. *et al.* Exome Sequencing Identifies FUS Mutations as a Cause of Essential Tremor. *Am. J. Hum. Genet.* **91,** 313–319 (2012).

46. Wu, Y.-R. *et al.* Identification of a novel risk variant in the FUS gene in essential tremor. *Neurology* **81,** 541–544 (2013).

47. Rajput, A. *et al.* Identification of FUS p.R377W in essential tremor. *Eur. J. Neurol.* **21,** 361–363 (2014).

48. 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,** 18285–18290 (2014).

49. Rajput, A. *et al.* VPS35 and DNAJC13 disease-causing variants in essential tremor. *Eur. J. Hum. Genet.* **23,** 887–888 (2015).

50. Hor, H. *et al.* Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum. Mol. Genet.* **24,** 5677–5686 (2015).

51. Bergareche, A. *et al.* SCN4A pore mutation pathogenetically contributes to autosomal dominant essential tremor and may increase susceptibility to epilepsy. *Hum. Mol. Genet.* (2015). doi:10.1093/hmg/ddv410

52. Liu, X. *et al.* Identification of candidate genes for familial early-onset essential tremor. *Eur. J. Hum. Genet.* **24,** 1009–1015 (2016).

53. Leng, X.-R., Qi, X.-H., Zhou, Y.-T. & Wang, Y.-P. Gain-of-function mutation p.Arg225Cys in SCN11A causes familial episodic pain and contributes to essential tremor. *J. Hum. Genet.* (2017).

54. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30,** 365–366 (2002).

55. Morgan, T. H. SEX LIMITED INHERITANCE IN DROSOPHILA. *Science* **32,** 120–122 (1910).

56. Strachan, T. & Read, A. P. *Human Molecular Genetics 4*. (Garland Science/Taylor & Francis Group, 2011).

57. 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,** 531–551 (1982).

58. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

59. Reich, D. & Lander, E. On the allelic spectrum of human disease. *Trends Genet* (2001).

60. Bush, W. S. & Moore, J. H. Genome-Wide Association Studies. *PLoS Comput. Biol.* (2012). doi:10.1371/journal.pcbi.1002822

61. Hirschhorn, J. & Daly, M. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6,** 95–108 (2005).

62. Distefano, J. K. & Taverna, D. M. Technological issues and experimental design of gene association studies. *Methods Mol. Biol.* **700,** 3–16 (2011).

63. Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236,** 1567–1570 (1987).

64. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456,** 728–731 (2008).

65. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467 (1977).

66. Zimmermann, J., Voss, H., Schwager, C., Stegemann, J. & Ansorge, W. Automated Sanger dideoxy sequencing reaction protocol. *FEBS Lett.* **233,** 432–436 (1988).

67. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679

68. Weber, J. L. & Myers, E. W. Human Whole-Genome Shotgun Sequencing. *Genome Res.* **7,** 401–409 (1997).

69. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).

70. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).

71. 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).

72. 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,** 1511–1516 (2014).

73. 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).

74. Sulonen, A.-M. *et al.* Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.* **12,** R94–R94 (2011).

75. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4,** 903–905 (2007).

76. Sawyer, S. L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: Time to address gaps in care. *Clinical Genetics* **89,** 275–284 (2016).

77. 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,** 1073–1081 (2009).

78. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31,** 2745–2747 (2015).

79. Marie Schwarz, J., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease- causing potential of sequence alterations. (2010). doi:10.1038/nmeth0810-575

80. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).

81. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12,** 745–755 (2011).

82. Louis, E. D. *et al.* The Washington Heights-Inwood Genetic Study of Essential Tremor: methodologic issues in essential-tremor research. *Neuroepidemiology* **16,** 124–133 (1997).

83. WMA. *Declaration of Helsinki. Lance* **353,** (1974).

84. Hughes, A., Daniel, S., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis

of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* **55,** 181–184 (1992).

85. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

86. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

87. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

88. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

89. 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

90. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164–e164 (2010).

91. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).

92. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39,** e118–e118 (2011).

93. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15,** 901–913 (2005).

94. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).

95. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35,** (2007).

96. Tekinay, A. B. A. *et al.* A role for LYNX2 in anxiety-related behavior. *Proc.*

Natl. Acad. Sci. U. S. A. **106,** 4477–82 (2009).

97. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425,** 917–925 (2003).

98. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).

99. Johansson, M. U. *et al.* Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* **13,** 173 (2012).

100. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31,** 3381–3385 (2003).

101. Kallberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7,** 1511–1522 (2012).

102. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).

103. Şen, M. Identification of candidate genes for familial essential tremor. (Bilkent University, 2016).

104. 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,** 4569–4580 (2003).

105. 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,** 722–729 (2005).

106. Page-McCaw, A., Ewald, A. J. & Werb, Z. Matrix metalloproteinases and the regulation of tissue remodelling. *Nat. Rev. Mol. Cell Biol.* **8,** 221–33 (2007).

107. 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).

108. Vaillant, C. *et al.* MMP-9 deficiency affects axonal outgrowth, migration, and apoptosis in the developing cerebellum. *Mol. Cell. Neurosci.* **24,** 395–408 (2003).

109. 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,** 2207–2214 (2006).

110. 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,** 11127–11135 (2003).

111. 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).

112. Avolio, C. *et al.* Serum MMP-2 and MMP-9 are elevated in different multiple sclerosis subtypes. *J. Neuroimmunol.* **136,** 46–53 (2003).

113. 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).

114. 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,** 6401–6408 (2005).

115. 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,** 104–112 (2001).

116. van Horssen, J. *et al.* Matrix metalloproteinase-19 is highly expressed in active multiple sclerosis lesions. *Neuropathol. Appl. Neurobiol.* **32,** 585–593 (2006).

117. Lettau, I. *et al.* Matrix metalloproteinase-19 is highly expressed in astroglial

tumors and promotes invasion of glioma cells. *J. Neuropathol. Exp. Neurol.* **69,** 215–223 (2010).

118. 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).

119. 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,** 1362–1378 (2008).

APPENDIX A

Primer Name	Direction	Sequence
TLL2	Forward	CCTCCCCAACCTTCCTTC
	Reverse	TGGGCTTGAGAATAGCCTCT
SNCAIP	Forward	CCATCTCCCACCTCAGAGAG
	Reverse	GATCCCCCGATTCGTTACTT
SRFBP1	Forward	GGTTTGATTTCATCAAGCTACG
	Reverse	CCGCATACAAAACAAAAATGG
EPHA8	Forward	CCCTACTGGAACATGACCAA
	Reverse	CAGATCTGGGGCAATACTGG
TCP10L2	Forward	CTTAGAGGCAGAGGCTGCAA
	Reverse	AAACTGGCAGTCAACAATGCT
SPEN	Forward	CTGTCCCTGTCCCCCTTC
	Reverse	CTTGGCTTAAAAGCCTGCTG
ARHGEF4	Forward	AGAGGAAATGAGGCCAGATG
	Reverse	TTGGTTATCCTGACGCCTGT
TMEM230	Forward	AAGGATTAGGTCTGTGACGTTTT
	Reverse	TGCAAGCTGCAGAATTCCTT
MMP19	Forward	ATTGTGTCTGTGGGTGAGCA
	Reverse	CTTCAGCAGCTACCCCAAAC

Table A1 Primer list.