A search for novel genes for juvenile myoclonic epilepsy

A submission for the partial fulfillment of MS of the Integrated-PhD programme

By

Pooja Barak

Molecular biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
(A Deemed University)
Jakkur, Bangalore 560 064, India

April 2014

DECLARATION

I hereby declare that this thesis titled "**A search for novel genes for juvenile myoclonic epilepsy**" is an authentic record of research work carried out by me under the guidance of Prof. Anuranjan Anand in Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore. This work has not been submitted in another place for the award of any other degree.

In keeping the norm of reporting scientific observations, due acknowledgements have been made wherever the work described here has been based on the findings of other investigators. Any omission, which might have occurred by oversight or misjudgment, is apologized for.

Pooja Barak

Place: JNCASR, Bangalore

Date:

CERTIFICATE

This is to certify that the work described in this thesis entitled "**A search for novel genes for juvenile myoclonic epilepsy**" is the result of the investigations carried out by Ms. Pooja Barak in the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, under my guidance.

The results presented in this thesis have not previously formed the basis for the award of any other diploma, degree or fellowship.

Anuranjan Anand

Place: JNCASR, Bangalore

Date:

Acknowledgements

I take this opportunity to express my special appreciation and thanks to my advisor Prof. Anuranjan Anand for providing me an opportunity to work with him and for being a tremendous mentor. I would like to thank him for encouraging my research and for allowing me to grow as a research student. His advice on both research as well as on my career have been priceless.

I would like to thank the faculty members, Professors Tapas Kundu, Kaustav Sanyal, Namita Surolia, Ranga Uday Kumar, Hemalatha Balram, Maneesha Inamdar, Ravi Manjithaya, Jayanta Haldar, G Ramesh, Balasubramanian, Shivaprasad, Shobhana Narsimha, Amitabh Joshi, Chandrabhas Narayan, Umesh Wahgmare, Ms. Nabonita Guha for very exciting course work which helped me get my basics strong before I could embark on journey of research work.

I wish to thank the members of Human Molecular Genetics laboratory: Praveen, Manpreet, Kalpita, Shalini, Shveta, Somya, Sourav, Rammurthy, Mohan, Dr. Vikas and Chandrashkhara for making my stay in the lab pleasant and in helping troubleshoot the problems that I encountered in my work.

Special thanks goes to my friends Anubha, Pratik, Mariyam, Kirthana and Sutanuka for helping me stay positive and focused by making me forget all my problems and for standing by my side through thick and thin.

I acknowledge the support from library, academic section, administration, computer lab, security, chandraya, hostel and mess. I would like to thank JNCASR for the fellowship.

Above all, I would like to thank my family. Words cannot express how grateful I am to my parents for all the love and support that they have always provided me.

Abbreviations

°C	Degree Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
ABI	Applied Biosystems
bp	Base pairs
BWA	Burrows-Wheeler Aligner
CA	California
CACNB4	Calcium channel, voltage-dependent, beta 4 subunit
CDC20B	Cell division cycle 20B
cm	Centimeters
cM	Centimorgan
dbSNP	Single Nucleotide Polymorphism database
DES	Desmin
DNA	Deoxyribonucleic acid
dNTP	Doxynucleotide triphosphate
EEG	Electroencephalogram
EFHC1	EF hand domain (C-terminal)-containing 1
EJM	Epilepsy, Juvenile Myoclonic
EtBr	Ethidium bromide
EVS	Exome Variant Server
GABR1	Gamma-aminobutyric acid (GABA) A receptor, alpha-1
GABRD	Gamma-aminobutyric acid (GABA) A receptor, delta
GAIIx	Illumina Genome Analyzer IIx platform
GRch37	Genome Reference Consortium Human genome build 37
GTCS	Generalized Tonic-Clonic Seizures
ILAE	International League Against Epilepsy
JME	Juvenile Myoclonic Epilepsy

kb	Kilobases
KCl	Potassium chloride
LOD	Logarithm of Odds
MAF	Minor Allele Frequency
Mb	Megabases
mg	Milligram
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Milliliter
mM	Millimolar
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
ng	Nanogram
NJ	New Jersey
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
pmol	Picomole
s	Seconds
TAE	Tris-Acetate-EDTA
Taq	*Thermus aquaticus*
Tris HCl	Tris Hydrochloride
U	Units
USA	United States of America
UTR	Untranslated region
WI	Wisconsin

Contents

Chapter 1

Introduction

1.1. Epilepsy

Epilepsy is a common neurological disorder characterized by recurrent epileptic seizures usually unprovoked by any immediately identifiable cause. One of the first descriptions of epileptic seizures can be traced back to 2,000 B.C. in ancient Akkadian texts, a language widely used in the region of Mesopotamia. The author described a patient with symptoms resembling epilepsy: *his neck turns left, his hands and feet are tense and his eyes wide open, and from his mouth froth is flowing without having any consciousness.* There are many reports which show that epilepsy was known in the ancient times but was not understood until mid-1950s. Modern advances in the understanding of epilepsy came during the 18th and 19th century. During this period, John Hughlings Jackson studied epilepsy in detail and published his first paper on epilepsy in 1861. He is considered the father of modern epileptology (Novel aspects on epilepsy, 2011).

Seizures, hallmark of epilepsy, are brief episodes of involuntary shaking which may involve a part of the body or the whole body and may be accompanied by loss of consciousness. An epileptic seizure is a clinical manifestation of an abnormal and excessive discharge of a set of neurons in the brain. Seizures may arise in a localized area of the brain or may involve the whole brain. The signs and symptoms depend on which area of the brain is affected. The hippocampal formation and cerebral cortex are considered the most epileptogenic regions of the brain (Bozzi et al 2012).

1.1.1. Epilepsy classification

The classification of the epilepsies is an evolving process for which the ultimate goal is to establish a clinically relevant, scientifically-based classification. Classification of epilepsy was first published in 1960 with official updates made in 1981 for seizures (Commission

on Classification and Terminology of the International League Against Epilepsy [ILAE], 1981) and 1989 for epilepsies (Commission on Classification and Terminology of the International League Against Epilepsy, 1989). In 2010, the first major modernization of the epilepsy classification was recommended. It was based on the new insights gained in the field of epilepsy based on major scientific advances such as modern neuroimaging, genomic technologies and concepts in molecular biology. Significantly, it presented an Organization rather than a Classification to denote that knowledge is still inadequate to have a fundamental framework on which to base a scientific classification (Berg et al 2010).

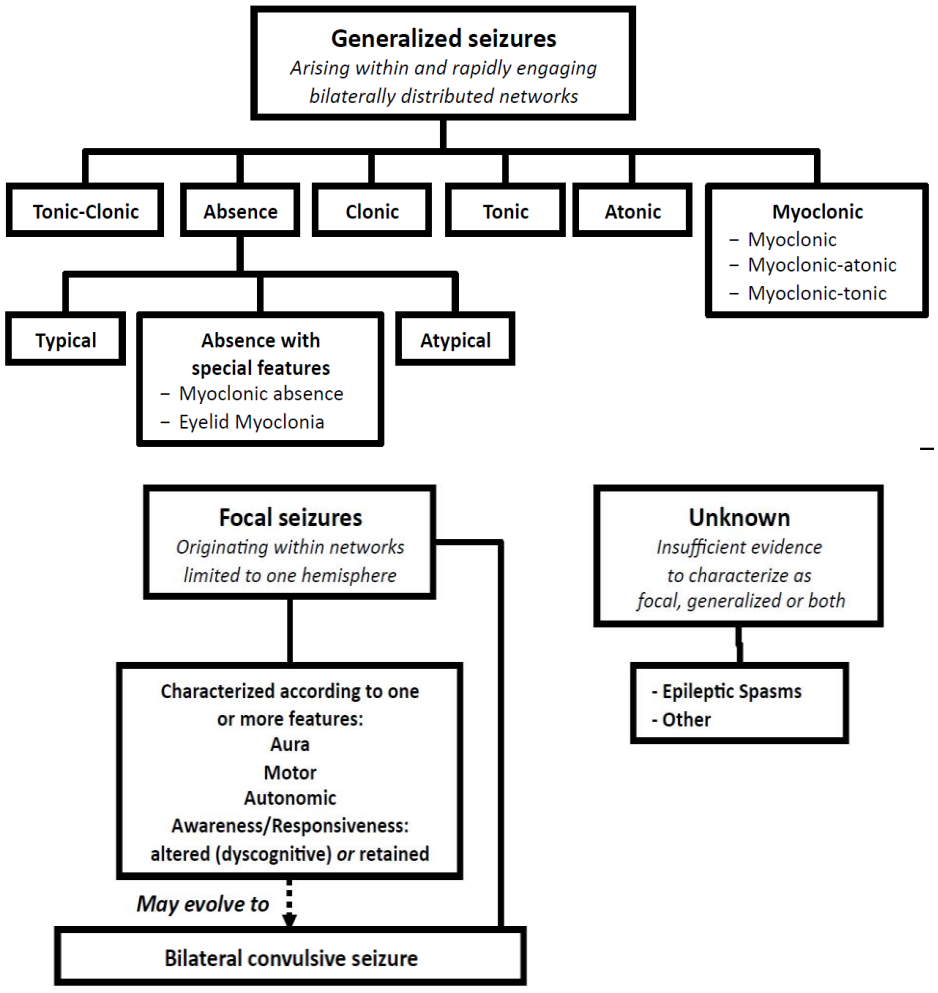


Figure 1.1: ILAE Proposal for Revised Terminology for Organization of Seizures and Epilepsies 2010

1.2. Juvenile myoclonic epilepsy (JME)

Juvenile myoclonic epilepsy (JME) was first described by Frenchman Théodore Herpin in 1867 (Eadie et al 2002) and then several patients with JME were reported and described in an article by Janz and Christian in 1957 (Janz et al 1957). Therefore JME is also known as Janz syndrome. Later the term juvenile myoclonic epilepsy was included in the International League Against Epilepsy classification of epileptic syndromes in 1989 (ILAE 1989).

JME is an age-dependent disorder with onset typically during adolescence (12-15 yrs). It is characterized by irregular myoclonic jerks (MJ) occurring early in the morning, although many patients experience generalized tonic-clonic seizures (GTCS) and absence seizures as well. Myoclonus manifests itself as sudden jerks in the muscles, usually involves bilateral and proximal upper extremity. It can also afflict distal muscles, such as the hands or the lower limbs (Moschetta et al 2011). Having interpreted as a manifestation of myoclonus nervousness or restlessness, it is the seizure that brings the patient to the doctor or hospital. Convulsive seizures usually come a few months after the onset of myoclonus, but in some it may take several years (Asconapé et al 1984).

1.2.1. Epidemiology

The incidence of JME has been estimated to be 1 per 100,000 individuals. The prevalence of JME has been estimated to be 5-10% of all epilepsies and 18% of idiopathic generalized epilepsies (IGE) (Jallon et al 2005). The literature suggests that JME typically appears in the second decade of life. However, the age of onset of JME spans a wide range from about 8-36 years, with peak onset between 12 and 18 years (Delgado-Escueta et al 1984). Those with the onset of JME outside the 8-36 year age bracket are uncommon and should be carefully evaluated for other diagnoses.

1.2.2. Clinical features

JME is clinically a heterogeneous disorder characterized by myoclonic jerks mainly on awakening and is often associated to generalized tonic-clonic seizures (GTCS: 80-97% cases) and typical absence seizures (12-54% cases) (Montalenti et al 2001). The myoclonic jerks are characterized by single or repetitive, bilateral, abrupt, symmetric, arrhythmic, involuntary movements, predominantly involving the shoulders and arms; however, they can be unilateral. Neurologic examination and neuroimaging results are normal in JME patients. Typical EEG features of JME consist of generalized discharges of single or multiple spike and slow wave of frequency of 3–5 Hz, although occasional complexes as slow as 2 Hz or as fast as 7 Hz may be evident.

Precipitating factors of epileptic seizures are varied, and the most common one is sleep deprivation, reported in 58.3-89.5% of cases. Other seizure triggering factors that have been identified include fatigue (73.7%), photosensitivity (36.8%), menses (24.1%), mental concentration (22.8%), and stress, excitement, or frustration (12.3%). Alcohol ingestion was also found to be a precipitating factor in 51.2% of cases (Alfradique et al 2007)

1.2.3. JME genetics

The etiology of JME is largely genetic and several genes have been identified. Genetic studies establish JME as a complex trait, associated with extensive genetic and phenotypic heterogeneity. Genetic heterogeneity could perhaps explain dissimilarities of the phenotypic expression of JME, as well as the conflicting linkage results from different populations. Ethnic background of the population, by contributing to heterogeneity in susceptibility alleles and polygenic effects, may contribute to genotypic and phenotypic heterogeneity (Vijay et al 2003). To date 30 loci linked or associated with JME phenotype have been reported, out of which putative causative genes have been identified at 8 (Delgado-Escueta et al 2013). OMIM (Online Mendelian Inheritance in Man; http:// www.omim.org/) has listed 9 JME loci, of which the causative genes have been identified only for four (Table 1.1).

Table 1.1: JME loci and genes given in OMIM

Cytogenetic Location	Loci name	Gene identified in the loci	Reference
6p12.2	EJM1	EFHC1	Suzuki et al 2004
5q14	EJM2		Elmslie et al 1997
6p21	EJM3		Greenberg et al 1998
5q12-q14	EJM4		Kapoor et al 2007
5q34	EJM5	GABRA1	Cossette et al 2002
2q23.3	EJM6	CACNB4	Escayg et al 2000
1p36.33	EJM7	GABRD	Dibbens et al 2004
3q27.1	EJM8		
2q33-q36	EJM9		Ratnapriya et al 2010

1.3. Objectives of my work

Till date, only a few genes for Juvenile myoclonic epilepsy have been mapped and identified in large families. In the present study my aim is to identify disease-causing genes at two genomic loci linked to Juvenile myoclonic epilepsy in two different large families from south India. These loci are mapped to chromosome 2q (EJM9) and 5p region. To find the causative mutation in both disease-linked intervals, next generation sequencing experiment was conducted. To identify the disease causing mutation in these two families, I examined all the protein coding genes in these disease-linked loci.

Chapters 2 and 3 of my thesis present the results of this analysis.

Chapter 2

Analysis of the 2q33-q36 locus, EJM9

Summary: In this chapter, a study aimed at identifying a potential JME-causing gene in a four-generational, multi-affected family, SCT135, is presented. A locus at 2q33-q36 was identified in this family by Ratnapriya et al 2010. The critical genomic region encompasses 24Mb of sequence length. To examine this region, a whole-exome based sequencing and analysis was carried out in two affected members of SCT135. Among 969 of gene variants detected, three novel rare variants: c.966A>T in *DES*, c.401G>C in *USP37* and c.2908C>T in *TNS1* were observed in the family. Each of these three variants co-segregated with the disease phenotype. However, on examining their presence in apparently normal control individuals, only c.966A>T in *DES* appeared to be a potential causative variant in SCT135.

2.1. Background research data

The family under study, SCT135, is a four generation multi-affected family. Out of the 17 members who had participated in the study, 6 have JME with manifestation of myoclonic seizures and generalized tonic-clonic seizures (GTCS). Whereas, few affected individuals also exhibited absence and febrile seizures.

2.1.1. Whole-genome based linkage mapping

Using 382 microsatellite markers from the ABI Prism Linkage Mapping Set $V_{2.5}$, genome wide linkage analysis was performed. Allele sizes were defined by GeneMapper™ (Applied Biosystems, CA, USA) and their Mendelian segregation was checked in the family. Linkage analysis was performed by two- and multi-point lod score calculations and highest two-point lod score of 2.29 was obtained at recombination fraction (θ) = 0 for the microsatellite marker D2S2248 at 2q35. This region was further fine mapped with additional microsatellite markers and the highest lod score of 3.32 was obtained for the marker D2S2248. No significant evidence of linkage was found for markers elsewhere in the genome. The boundary for the disease-linked haplotype was defined by recombination events. This linked region, which falls between D2S116 and D2S2390, is of 24Mb sequence length and harbors 158 annotated protein coding genes.

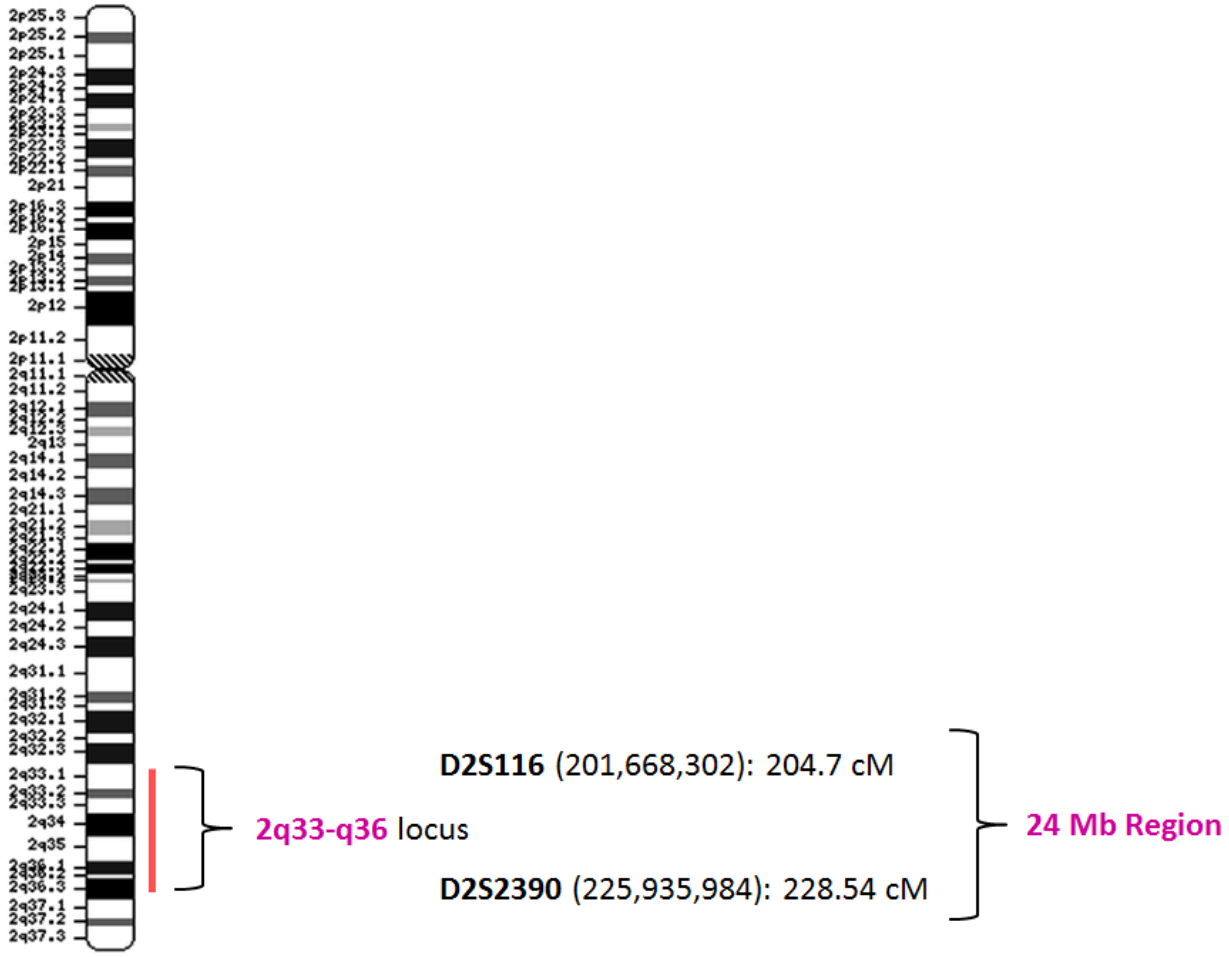


Figure 2.1: Position of the EJM9 locus (2q33-q36) on chromosome 2.

Objective of current study: Based on the evidence of linkage of the locus 2q33-q36 with JME in the family SCT135, the aim of my work is to analyze this region using whole exome based sequencing and identify the disease-causing gene in this family.

2.2. Materials and Methods

2.2.1. Whole-exome sequencing experiment

Five micrograms of genomic DNA was fragmented (sonication at 55 pulses ON at 30s ON and 30s OFF) (Bioruptor-Diagenode, NJ, USA) and purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). The target peak for base pair size was 150-400 bp. The sheared DNA was analyzed for size distribution using Agilent DNA 1000 Bioanalyzer (Agilent Technologies, CA, USA). Successively, the sheared DNA fragments were used to construct DNA libraries using Agilent's SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library. The constructed library was reformed by a series of steps using different enzymes to repair ends and make blunt-ended 5' phosphorylated fragments, add a single nucleotide A overhang and ligate 60bp sequence adaptors to fragment ends. Each step was followed by a purification step using Agencourt AMPure XP beads. After ligation, the adapter-ligated fragments were enriched by PCR and concentrated using a vacuum concentrator (Eppendorf, Hamburg, Germany). The library was then hybridized to SureSelect™ biotinylated RNA baits at 65°C for 24-72 hours. Hybridized library fragments were isolated by magnetic capture using Dynal M-280 streptavidin coated beads (Invitrogen, CA, USA) followed by purification of the capture library-bead solution using AMPure XP beads. PCR amplification was carried out to enrich the captured library and the amplified products were purified using AMPure XP beads. The Amplified Capture DNA was analyzed using the high sensitivity bioanalyzer chip (Agilent) which shows a peak in the size range of 300-400 nucleotides. The SureSelect Human exome kit is designed to enrich total of 51 Mb region in genome. The sequencing was carried out for the captured libraries with Illumina Genome Analyzer IIx platform (GAIIx) obtaining the 72 bp paired-end reads.

2.2.2. Next-generation sequencing analysis

The whole-exome FASTAQ sequencing reads were aligned to human genome reference (hg19/GRch37) using BWA v-0.6.0 (Li and Durbin 2009). The reads showing at least 70% of bases with a minimum Phred score of 20, obtained by SeqQC v-2.0 (http:// genotypic.co.in/Products/7/Seq-QC.aspx), were used for alignment. Using SAMtools v-0.1.7a (Li et al 2009), duplicate reads arising possibly from PCR artifacts, were removed. The variant

calling was performed by SAMtools at a Phred like SNP quality score of 20. The variants identified were annotated by SNPeff and filtered against the dbSNP131. Novel variants were further examined in updated databases such as dbSNP139 (http:// www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes (http://browser.1000genomes.org/index.html), Ensemble (http://asia.ensembl.org/index.html) and EVS datasets (http:// evs.gs.washington.edu/EVS/). In order to obtain potential variants, which may have gotten missed at high coverage, variants up to 3x read depth were manually examined. Those transcript regions which remained uncovered by the whole-exome sequencing were manually identified and examined by Sanger sequencing.

2.2.3. Sanger-based sequence validation and genetic analysis

All new variants identified in the NGS dataset were validated by Sanger sequencing. Primers were designed spanning the variant-carrying exons/regions. The variants common to both NGS-sequenced samples were analyzed in other family members for their presence/absence. The co-segregating changes were examined in an ethnically matched control set of normal individuals.

2.2.3.1. Polymerase chain reaction

Primers were designed for the exons/regions having the variants identified in the NGS analysis using the Primer3.0 (http://bioinfo.ut.ee/primer3-0.4.0/) and Oligocalc (http://www.basic.northwestern.edu/biotools/OligoCalc.html) softwares and synthesized by Sigma-Aldrich. The PCR conditions were standardized for each primer set and amplification was performed on thermal cycler GeneAmp® PCR System 9700 (Applied Biosystems, Life Technologies). The PCR reaction mixture contained deionized water, 10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$, 0.8mM dNTPs, 0.25µM of each primer, 0.05U/µl Taq Polymerase (NEB) and 100ng/µl DNA in a 20µl volume. The standardized amplification conditions were: Initial denaturation at 95^{o}C for 5 minutes, followed by 40 cycles of denaturation at 95^{o}C for 30 seconds, annealing for 30 seconds and elongation at 72^{o}C for 30 seconds, and a final extension step for 10 minutes at 72^{o}C. The amplified products were electrophoresed on 1.5% agarose gel containing 2µl Ethidium Bromide (2µg/ml) and purified using a Millipore vacuum manifold plate and eluted in 20µl of deionized water.

2.2.3.2. Sanger based sequencing

PCR-amplified products were single strand amplified by cycle sequencing using 1µl of BigDye® Terminator v3.1 Cycle Sequencing reaction mix (Applied Biosystems, CA, USA), 1X sequencing buffer (Applied Biosystems), 0.25µM primer and 3µl of purified PCR product in a 20µl volume. The following cycling conditions were used: initial denaturation at 95°C for 1 minute, followed by 25 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C for 4 minutes, and a final hold at 4°C. It is followed by alcohol precipitation of the amplified product by adding 16 µl of chilled autoclaved deionized water and 64 µl of chilled 95% alcohol to each well of the sequencing plate. The sequencing plate contents were invert mixed and incubated at room temperature for 30 minutes, followed by centrifugation at 2500g for 30 minutes. The precipitated DNA was washed with 150 µl of 70% alcohol followed by a 10 minutes centrifugation at 2000g. The plate was air dried to remove all residual alcohol and the DNA denatured at 95^{o} C in presence of 10µl of formamide per well. These denatured single stranded amplified products were Sanger sequenced using an automated DNA sequencer, DNA Analyzer 3730 (Applied Biosystems). The sequences thus obtained were aligned to the respective reference gene sequences obtained from Genbank database and the variations were identified using SeqMan 5.01 (DNASTAR, Madison, WI).

2.2.4. Bioinformatic analysis

DES protein sequence for multiple species was obtained from NCBI (http:// www.ncbi.nlm.nih.gov/) and aligned by Clustal Omega (http:// www.ebi.ac.uk/Tools/msa/clustalo/). To predict the effect of the variation on the structure/function of the protein, various bioinformatics tools: SIFT (http://sift.jcvi.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and Mutation taster (http:// www.mutationtaster.org/) were used.

2.3. Results

To examine the 2q33-q36 region in detail, whole-exome based sequencing was undertaken. This region spans 24Mb of sequence length and harbors 158 well annotated protein coding genes. In the whole-exome sequencing experiment, a total of 9.59 GB sequence per sample, as

43.74 million paired end reads of 72 bp, were generated from sequencing on the GAIIx where more than 98% of bases were of high quality (Table 2.1 and 2.2) and a total coverage of 98% was obtained. The region of our interest has 2061 coding exons and 563 non-coding exons out of which 120 coding exons and 380 non-coding exons have not been covered by the whole-exome sequencing. These missing exons have been targeted to be covered by Sanger sequencing. Out of the 120 missing coding exons, 95 have been Sanger sequenced. Upon variant analysis a total of 940 variants that were identified were already reported in dbSNP135 or 1000 Genome project datasets and hence, were not carried forward for our study. The 29 novel variants identified were Sanger validated and then checked for their segregation in the family. Three variants, namely, c.966A>T in *DES,* c.401G>C in *USP37* and c.2908C>T in *TNS1*, co-segregated with the JME phenotype in the family. Two of the co-segregating variants, c.401G>C in *USP37* and c.2908C>T in *TNS1*, were present in normal control individuals and were therefore, not taken forward for analysis (table 2.3).

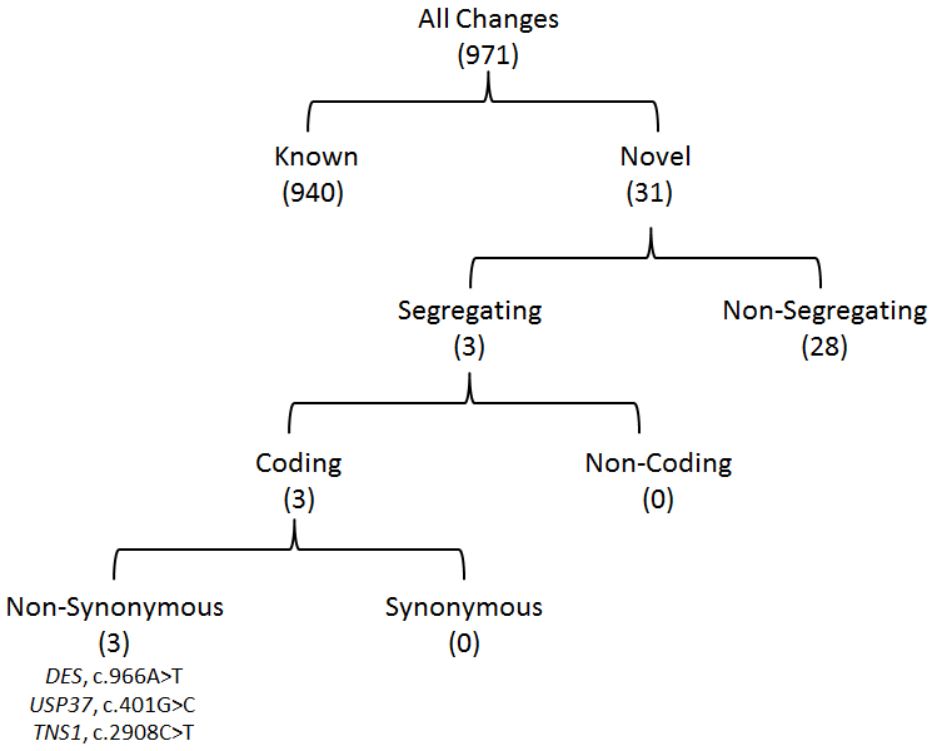


Figure 2.2: Analysis of the novel variants identified: Each category with the number of variants identified is shown.

Table 2.1: Summary statistics for the *Illumina* reads generated

	Sample I		Sample II	
Read length in bases (Single/paired-end)	72	(paired)	72	(paired)
Total reads generated (percentage high quality reads[a])	95985558	(93.54%)	83505302	(93.81%)
Total bases (percentage high quality bases[b])	6910.96Mb	(94.25%)	6012.38Mb	(94.41%)
Reads aligned to genome post filtering (percentage reads aligned [c])	87345896	(99.84%)	75873694	(99.82%)
Reads on targets (percentage reads aligned[d])	70441120	(80.97%)	60801508	(79.99%)
Reads on chromosome 2 exome (percentage reads aligned[e])	1012860	(1.16%)	811040	(1.07%)

[a] Bases with Phred score >20, [b] Reads with more than 70% bases with Phred score >20, [c] Reads were filtered for PCR duplicates, [d] Whole-exome target, [e] Target is exome in the EJM9 locus on chromosome 2. The reference genome is human ref seq, hg19 (GRch37).

Table 2.2: Sequence coverage summary for the EJM9 locus whole-exome sequencing experiment

	Chromosome 2	Whole exome	Chromosome 2	Whole exome
%Total target covered with at least 5X Read Depth	93.81	89.70	94.79	91.37
%Total target covered with at least 10X Read Depth	89.83	83.66	91.51	86.18
%Total target covered with at least 15X Read Depth	85.88	78.33	87.95	81.33
%Total target covered with at least 20X Read Depth	82.26	73.51	84.06	76.50

Coverage is for the EJM9 exome on chromosome 2 and whole-exome region targeted by the whole-exome sequencing experiment.

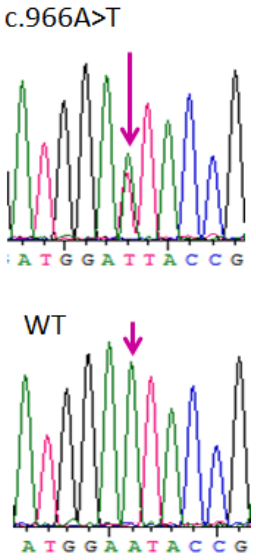
Table 2.3: New gene variants identified in the EJM9 locus by whole-exome sequencing analysis

Gene	Sequence variant	Location	Effect on protein	Family segregation	Allele frequency in controls	Allele frequency in databases
DNAH7	NM_018897.2:c.11586+57G>A	Intron 63	_	not segregating	_	_
DNAH7	NM_018897.2:c.8131A>G	Exon 44	p.2711I>V	not segregating	_	_
HECW2	NM_020760.1:c.2587-40T>C	Intron 11	_	not segregating	_	G=10/A=12996
TRAK2	NM_015049.2:c.1194-21G>A	Intron 11	_	not segregating	_	Frequency not given
ALS2	NM_020919.3:c.3308A>G	Exon 20	p.1103H>R	not segregating	_	_
ALS2	NM_020919.3:c.1114-86G>A	Intron 4	_	not segregating	_	_
CDK15	NM_001261435.1:c.731-70C>T	Intron 7	_	not segregating	_	_
ICA1L	NM_138468.4:c.1209T>A	Exon 12	p.403F>L	not segregating	_	0.003
PARD3B	NM_152526.5:c.1028C>T	Exon 8	p.343A>V	not segregating	_	_
IDH1	NM_005896.2:c.851-24C>T	Intron 7	_	not segregating	_	A=1/G=13005
UNC80	NM_032504.1:c.6718-58G>A	Intron 43	_	not segregating	_	_
ACADL	NM_001608.3:c.1191A>G	Exon 10	p.(=)	not segregating	_	_
XRCC5	NM_021141.3:c.798+12G>A	Intron 7	_	not segregating	_	A=2/G=13004
TNS1	**NM_022648.4:c.2908C>T**	**Exon 20**	**p.970R>W**	**Co-segregating**	**3/95**	_
TNS1	NM_022648.4:c.2887A>G	Exon 19	p.963S>G	not segregating	_	0.002
TNS1	NM_022648.4:c.1313G>A	Exon 17	p.438R>Q	not segregating	_	0.002
CXCR2	NM_001557.3:c.90C>T	Exon 3	p.(=)	not segregating	_	0.0005
CXCR1	NM_000634.2:c.633C>A	Exon 2	p.211F>L	not segregating	_	Frequency not given
C2ORF62	NM_198559.1:c.283T>A	Exon 3	p.95F>I	not segregating	_	Frequency not given
USP37	**NM_020935.2:c.401G>C**	**Exon 6**	**p.134S>T**	**Co-segregating**	**5/382**	**0.0032**
STK36	NM_015690.4:c.3643G>A	Exon 26	p.1215A>T	not segregating	_	_

TTLL4	NM_014640.4:c.880G>A	Exon 3	p.294D>N	not segregating	_	_
PRKAG3	NM_017431.2:*71C>T	Exon 13	_	not segregating	_	_
STK16	NM_001008910.2:c.855C>T	Exon 8	p.(=)	not segregating	_	_
DES	**NM_010043.2:c.966A>T**	**Exon 5**	**p.322E>D**	**Co-segregating**	**0/380**	**_**
OBSL1	NM_015311.2:c.1860A>C	Exon 5	p.(=)	not segregating	_	_
CUL3	NM_003590.4:c.883+88T>C	Intron 6	_	not segregating	_	0.002
DOCK10	NM_014689.2:c.3037-38A>T	Intron 53	_	not segregating	_	_
COL4A4	NM_000092.4:c.4523-64G>T	Intron 46	_	not segregating	_	_
COL4A4	NM_000092.4:c.1547A>T	Exon 22	p.516D>V	not segregating	_	_
SLC19A3	NM_025243.3:c.779A>T	Exon 3	p.260D>V	not segregating	_	_

The variant, c.966A>T in *DES*, was observed to co-segregate with JME and was absent in 384 chromosomes of normal control individuals. The change of 'A' nucleotide to 'T' nucleotide gives rise to an amino acid change of glutamic acid to aspartic acid at 322nd position of desmin protein molecule. To determine the functional importance of E322, its conservation was examined across *DES* from different species using clustal omega to perform multiple sequence alignment. The residue was found to be conserved in Human, Chimpanzee, Olive baboon, cow, sheep, pig, dog, chicken, pigeon, mouse, rat, golden hamster, ferret, frog, dog-fish and zebra-fish (Figure 2.3)

A ***DES*, c.966A>T** **B** ***DES*, p.322E>D**




```
Zebrafish          SKNNEALKQSKLETMEYRHQIQSYTCEIDSL
Dog_Fish           TKNNDALRQAKQEVMEYRHQVQSYTCEIDAL
Frog               KKNNDAMRQSKQEMMEYRHQIQSYTCEIDAL
Dog                NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Mouse              NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Rat                NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Ferret             NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Golden_Hamster     NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Human              NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Chimpanzee         NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Olive_Baboon       NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Cow                NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Sheep              NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Pig                NKNNDALRQAKQEMMEYRHQIQSYTCEIDAL
Chicken            NKNNDALRQAKQEMLEYRHQIQSYTCEIDAL
Pigeon             NKNNDALRQAKQEMLEYRHQIQSYTCEIDAL
                   .***:*::*:* * :*****:********:*
```

Figure 2.3: Novel, rare, co-segregating variant c.966A>T in *DES*: A: Representative sequences of an affected and a normal individual. Arrow marks the nucleotide showing variation. **B:** The corresponding amino acid conservation for the variation.

To check for the effect of the variant p.322E>D on function of the desmin protein molecule, bioinformatic analysis tools: PolyPhen-2, SIFT and Mutation taster, were used. PolyPhen-2 and mutation taster predicted that this variant can have possible damaging effects but SIFT predicted the effect to be neutral.

2.4. Discussion

The aim of the current study was to examine the EJM9 locus using whole-exome sequencing to isolate any possible candidate gene/s causing JME in SCT135.

Upon analysis of all the variants picked in the whole exome analysis three novel, segregating variants, namely, c.966A>T in *DES*, c.401G>C in *USP37* and c.2908C>T in *TNS1* were identified. Two of the disease co-segregating variants, namely, c.401G>C in *USP37* and c.2908C>T in *TNS1*, were present in normal control individuals and were therefore not taken forward for our analysis.

The final variant that was highlighted in our analysis was c.966A>T in *DES*. This variant gives rise to a p.322E>D mutation in the desmin protein. The amino acid Glu322 was found to be evolutionary conserved and the Glu322Asp mutation is predicted to cause deleterious effects on normal protein functionality. These results suggest p.322E>D substitution in Desmin to be the underlying genetic defect, thus leading to epilepsy phenotype in this family.

DES codes for a 470 amino acid, muscle specific, type III intermediate filament. It is important for cell architecture and structure as it connects many cell organelles. This protein was first reported in 1976 (Lazarides et al 1976) in an immunological characterization based study aimed at studying intermediate filaments from muscle cells. Mutations in *DES* gene have been reported to cause human skeletal as well as cardiac myopathy, termed desmin related myopathies (Clemen et al 2013). Before establishing *DES* as the JME causing gene in SCT135, all remaining missing and low coverage exons need to be examined and checked for presence/absence of any other disease-segregating rare variant/s. To gather additional genetic evidence for its involvement in JME causation, it needs to sequenced and analysed in large cohort of JME affected samples.

Chapter 3

Analysis of the locus at 5p15.33-q12.1

Summary: In this chapter, a study aimed at identifying a potential JME causing gene in a multiple generation, multi-affected family, GLH35, is presented. A locus at 5p15.33-q12.1 was identified to be segregating with the JME phenotype in this family (unpublished data). The critical genomic region encompasses 64Mb of sequence length. To examine this region, a whole exome based sequencing and analysis was carried out in two affected members of GLH35. Among a large number (891) of gene variants detected, twelve novel rare variants: c.8952-49C>G in *DNAH5*, c.582C>T in *CDH18*, c.644-71T>G in *CDH18*, c.1394-11G>A in *CDH12*, c.538-30C>T in *IL7R*, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.364-43C>T in *PRKAA1*, c.3749+28T>C in *HEATR7B2*, c.*60G>A in *GZMA*, c.1525G>A in *CDC20B*, c.865-24C>T in *DDX4* were observed in the family. Each of these twelve variants co-segregated with the clinical phenotype. Further, upon checking for allele frequencies in apparently normal individuals, four of the twelve variants had MAF>0.005 and were hence not carried forward for analysis. The remaining eight variants were analyzed for their possible role in causation of JME in GLH35.

3.1. Background research data

The family under study, GLH35, is a three generation multi affected family. Out of the 13 members who had participated in the current study, 7 have JME with manifestation of myoclonic seizures and generalized tonic-clonic seizures (GTCS). A few affected individuals also have absence seizures and febrile seizures.

3.1.1. Whole-genome based linkage mapping

Using 382 microsatellite markers from the ABI Prism Linkage Mapping Set $V_{2.5}$, genome wide linkage analysis was performed. Allele sizes were defined by GeneMapper™ (Applied Biosystems) and their Mendelian segregation was checked in the family. Linkage analysis was performed by two and multipoint lod score calculations and highest two-point lod score of 1.7 was obtained at recombination fraction=0 for D5S426 at 5p13.2. No suggestive evidence of

linkage was found for markers elsewhere in the genome. The boundary for the region that was linked and co-segregated with the disease phenotype was defined by D5S1981 and D5S407. This linked region is of 64 Mb sequence length and harbors 177 annotated protein coding genes.

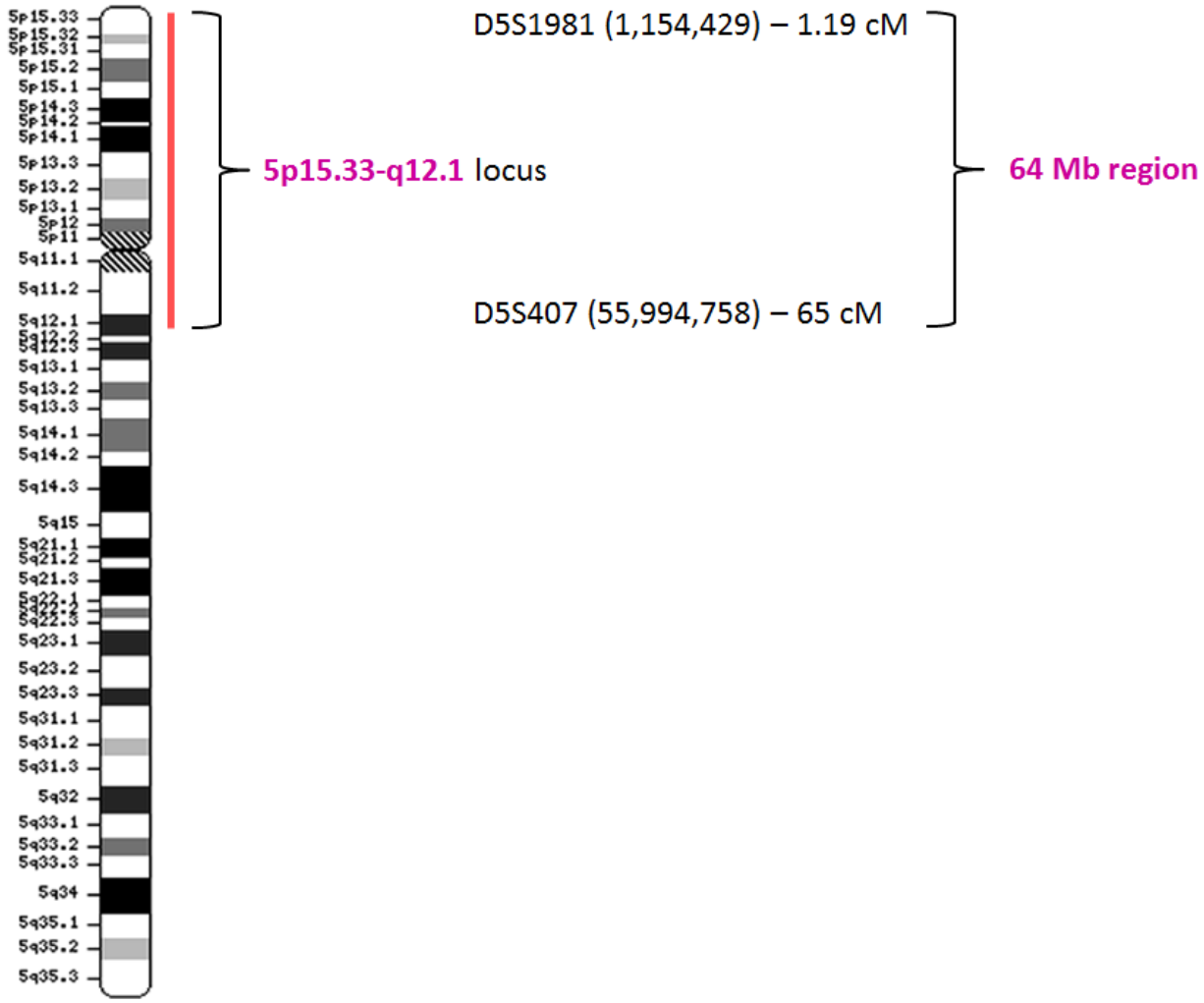


Figure 3.1: Position of the loci 5p15.33-q12.1 on chromosome 5.

Objective of current study: Based on the evidence suggesting the linkage of the locus 5p15.33-q12.1 with JME in the family GLH35, the aim of my work is to analyze this region using whole exome based sequencing and analysis to identify a potential JME causing candidate gene.

3.2. Materials and Methods

3.2.1. Whole-exome sequencing experiment

Five micrograms of genomic DNA was fragmented (sonication at 55 pulses ON at 30s ON and 30s OFF) (Bioruptor-Diagenode, NJ, USA) and purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). The target peak for base pair size was 150-400 bp. The sheared DNA was analyzed for size distribution using Agilent DNA 1000 Bioanalyzer (Agilent Technologies, CA, USA). Successively, the sheared DNA fragments were used to construct DNA libraries using Agilent's SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library. The constructed library was reformed by a series of steps using different enzymes to repair ends and make blunt-ended 5' phosphorylated fragments, add a single nucleotide A overhang and ligate 60bp sequence adaptors to fragment ends. Each step was followed by a purification step using Agencourt AMPure XP beads. After ligation, the adapter-ligated fragments were enriched by PCR and concentrated using a vacuum concentrator (Eppendorf, Hamburg, Germany). The library was then hybridized to SureSelect™ biotinylated RNA baits at 65°C for 24-72 hours. Hybridized library fragments were isolated by magnetic capture using Dynal M-280 streptavidin coated beads (Invitrogen, CA, USA) followed by purification of the capture library-bead solution using AMPure XP beads. PCR amplification was carried out to enrich the captured library and the amplified products were purified using AMPure XP beads. The Amplified Capture DNA was analyzed using the high sensitivity bioanalyzer chip (Agilent) which shows a peak in the size range of 300-400 nucleotides. The SureSelect Human exome kit is designed to enrich total of 51 Mb region in genome. The sequencing was carried out for the captured libraries with Illumina Genome Analyzer IIx platform (GAIIx) obtaining the 72 bp paired-end reads.

3.2.2. Next-generation sequencing analysis

The whole-exome FASTAQ sequencing reads were aligned to human genome reference (hg19/GRch37) using BWA v-0.6.0 (Li and Durbin 2009). The reads showing at least 70% of bases with a minimum Phred score of 20, obtained by SeqQC v-2.0 (http://genotypic .co.in/Products/7/Seq-QC.aspx), were used for alignment. Using SAMtools v-0.1.7a (Li et al 2009), duplicate reads arising possibly from PCR artifacts, were removed. The variant calling

was performed by SAMtools at a Phred like SNP quality score of 20. The variants identified were annotated by SNPeff and filtered against the dbSNP131. Novel variants were further examined in updated databases such as dbSNP139 (http:// www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes (http://browser.1000genomes.org/index.html), Ensemble (http://asia.ensembl.org/index.html) and EVS datasets (http:// evs.gs.washington.edu/EVS/). In order to obtain potential variants, which may have gotten missed at high coverage, variants up to 3x read depth were manually examined. Those transcript regions which remained uncovered by the whole-exome sequencing were manually identified and examined by Sanger sequencing.

3.2.3. Sanger-based sequence validation and genetic analysis

All new variants identified in the NGS dataset were validated by Sanger sequencing. Primers were designed spanning the variant-carrying exons/regions. The variants common to both NGS-sequenced samples were analyzed in other family members for their presence/absence. The co-segregating changes were examined in an ethnically matched control set of normal individuals.

3.2.3.1. Polymerase chain reaction

Primers were designed for the exons/regions having the variants identified in the NGS analysis using the Primer3.0 (http://bioinfo.ut.ee/primer3-0.4.0/) and Oligocalc (http://www.basic.northwestern.edu/biotools/OligoCalc.html) softwares and synthesized by Sigma-Aldrich. The PCR conditions were standardized for each primer set and amplification was performed on thermal cycler GeneAmp® PCR System 9700 (Applied Biosystems, Life Technologies). The PCR reaction mixture contained deionized water, 10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$, 0.8mM dNTPs, 0.25µM of each primer, 0.05U/µl Taq Polymerase (NEB) and 100ng/µl DNA in a 20µl volume. The standardized amplification conditions were: Initial denaturation at 95^{o}C for 5 minutes, followed by 40 cycles of denaturation at 95^{o}C for 30 seconds, annealing for 30 seconds and elongation at 72^{o}C for 30 seconds, and a final extension step for 10 minutes at 72^{o}C. The amplified products were electrophoresed on 1.5% agarose gel containing 2µl Ethidium Bromide (2µg/ml) and purified using a Millipore vacuum manifold plate and eluted in 20µl of deionized water.

3.2.3.2. Sanger based sequencing

PCR-amplified products were single strand amplified by cycle sequencing using 1µl of BigDye® Terminator v3.1 Cycle Sequencing reaction mix (Applied Biosystems), 1X sequencing buffer (Applied Biosystems), 0.25µM primer and 3µl of purified PCR product in a 20µl volume. The following cycling conditions were used: initial denaturation at 95°C for 1 minute, followed by 25 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C for 4 minutes, and a final hold at 4°C. It is followed by alcohol precipitation of the amplified product by adding 16 µl of chilled autoclaved deionized water and 64 µl of chilled 95% alcohol to each well of the sequencing plate. The sequencing plate contents were invert mixed and incubated at room temperature for 30 minutes, followed by centrifugation at 2500g for 30 minutes. The precipitated DNA was washed with 150 µl of 70% alcohol followed by a 10 minutes centrifugation at 2000g. The plate was air dried to remove all residual alcohol and the DNA denatured at 95^{o} C in presence of 10µl of formamide per well. These denatured single stranded amplified products were Sanger sequenced using an automated DNA sequencer, DNA Analyzer 3730 (Applied Biosystems). The sequences thus obtained were aligned to the respective reference gene sequences obtained from Genbank database and the variations were identified using SeqMan 5.01 (DNASTAR, Madison, WI).

3.2.4. Bioinformatic analysis

Nucleotide or protein sequences for *DNAH5, CDH18, CDH12, IL7R, C5ORF42, NUP155, PRKAA1, HEATR7B2, GZMA, CDC20B*, and *DDX4* were obtained from NCBI (http://www.ncbi.nlm.nih.gov/) and aligned by Clustal omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). To predict the effect of the variation on the structure/function of the protein, various bioinformatics tools: SIFT (http://sift.jcvi.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) and Mutation taster (http://www.mutationtaster.org/) were used. Splice site predictions for the intronic changes were made using: Berkeley Drosophila Genome Project - Splice Site Prediction (http:// www.fruitfly.org/seq_tools/splice.html) and Human Splicing Finder v2.4.1 (http:// www.umd.be/HSF/).

3.3. Results

To examine the 5p15.33-q12.1 region in detail, whole-exome based sequencing was undertaken. This region spans 64Mb of sequence length and encodes for 177 protein coding genes. In the whole-exome sequencing experiment, a total of 9.01 GB sequence, as 41.58 million paired end reads of 72bp, was generated from sequencing on the GAIIx where more than 98% of bases were of high quality (Table 3.1 and 3.2) and a total coverage of 98% was obtained. The region of our interest has 2255 coding exons and 570 non-coding exons out of which 136 coding exons and 415 non-coding exons have not been covered by the whole-exome sequencing. These missing exons shall be examined by Sanger sequencing. A total of 855 variants that were identified were already reported in dbSNP135 or 1000 Genome project datasets and hence were not carried forward for further study. The 36 novel variants identified were validated and then checked for their segregation in the family.

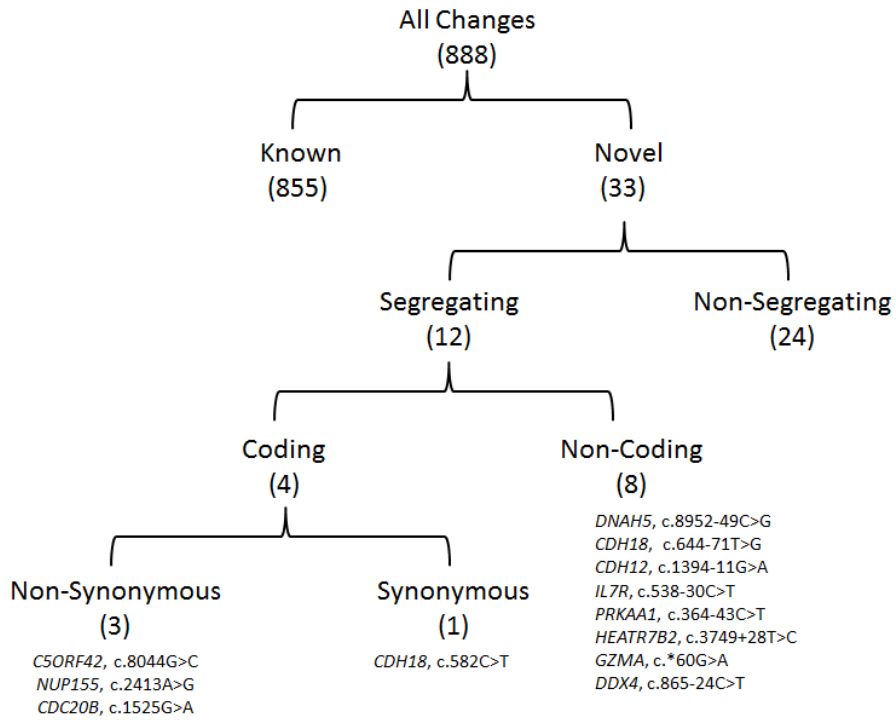


Figure 3.2: Analysis of the novel variants identified: Each category with the number of variants identified is shown.

Twelve variants: c.8952-49C>G in *DNAH5*, c.582C>T in *CDH18*, c.644-71T>G in *CDH18*, c.1394-11G>A in *CDH12*, c.538-30C>T in *IL7R*, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.364-43C>T in *PRKAA1*, c.3749+28T>C in *HEATR7B2*, c.*60G>A in *GZMA*, c.1525G>A in *CDC20B*, c.865-24C>T in *DDX4*, co-segregated with the JME phenotype in the family. Four of the co-segregating variants, namely c.582C>T in *CDH18*, c.644-71T>G in *CDH18*, c.364-43C>T in *PRKAA1*, c.3749+28T>C in *HEATR7B2*, were present in normal control individuals and were therefore not taken forward for our analysis. The remaining eight variants, namely, c.8952-49C>G in *DNAH5*, c.1394-11G>A in *CDH12*, c.538-30C>T in *IL7R*, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.*60G>A in *GZMA*, c.1525G>A in *CDC20B*, c.865-24C>T in *DDX4*, were observed to co-segregate with JME and were present at MAF < 0.005 in 192 chromosomes of normal control-individuals examined. In order to further characterize the remaining 8 variants, bioinformatic analysis was carried out to check the conservation status (at DNA or protein level) and the effect of the variant on the protein structure/function. Out of the three non-synonymous variants, namely, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.1525G>A in *CDC20B*, only *CDC20B* was conserved across species and showed deleterious effects on performing mutation prediction analysis using PolyPhen-2, SIFT and mutation taster. None of the analysis done on the intronic variants showed any evidence of them being disease causing.

Table 3.1: Summary statistics for the *Illumina* reads generated

Summary statistics	Sample I		Sample II	
Read length in bases (Single/paired-end)	72	(paired)	72	(paired)
Total reads generated (percentage high quality reads[a])	91303488	(93.45%)	87326104	(91.69%)
Total bases (percentage high quality bases[b])	6573.85Mb	(94.25%)	6287.47Mb	(94.41%)
Reads aligned to genome post filtering (percentage reads aligned[c])	82955930	(99.76%)	76753838	(99.78%)
Reads on targets (percentage reads aligned[d])	67096500	(80.69%)	63833426	(82.99%)
Reads on chromosome 5 exome (percentage reads aligned[e])	688950	(0.83%)	711628	(0.93%)

[a] Bases with Phred score >20, [b] Reads with more than 70% bases with Phred score >20, [c] Reads were filtered for PCR duplicates, [d] Whole-exome target, [e] Target is exome in the 5p15.33-q12.1 locus on chromosome 5. The reference genome is hg19 (GRch37).

Table 3.2: Sequence coverage summary for the 5p15.33-q12.1 locus whole-exome sequencing experiment

Coverage summary	Chromosome 5	Whole exome	Chromosome 5	Whole exome
%Total Target covered with at least 5X Read Depth	95.43	94.22	95.61	94.01
%Total Target covered with at least 10X Read Depth	91.93	89.68	93.57	89.86
%Total Target covered with at least 15X Read Depth	88.02	85.33	91.49	85.97
%Total Target covered with at least 20X Read Depth	83.85	80.94	88.69	81.91

Coverage is for the 5p15.33-q12.1 exome on chromosome 5 and whole-exome region targeted by the whole-exome sequencing experiment.

Table 3.3: New gene variants identified in the 5p15.33-q12.1 locus by whole-exome sequencing analysis

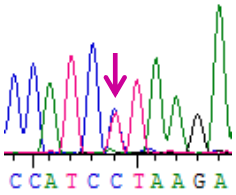
Gene	Sequence Variant	Location	Effect on protein	Family segregation	Allele frequency in controls	Allele frequency in databases
EXOC3	NM_007277.4:c.144+59T>C	Intron 2	_	not segregating	_	_
SLC9A3	NM_004174.2:c.1816C>T	Exon 12	p.606R>W	not segregating	_	_
TRIP13	NM_004237.3:c.535+39A>G	Intron 5	_	not segregating	_	1 in 13,005
CLPTM1L	NM_030782.3:c.1371+29G>A	Intron 14	_	not segregating	_	freq not given
SLC6A3	NM_001044.4:c.1805A>G	Exon 14	p.602E>W	not segregating	_	0.006
MED10	NM_032286.2:*47G>A	Exon 4	_	not segregating	_	_
MARCH6	NM_005885.3:c.2506+4C>T	Exon 24	_	not segregating	_	_
MARCH6	NM_005885.3:c.2642+71A>G	Exon 25	_	not segregating	_	_
ROPN1L	NM_031916.4:c.269G>A	Exon 3	p.90R>Q	not segregating	_	_
DNAH5	**NM_001369.2:c.8952-49C>G**	**Intron 53**	**_**	**Co-segregating**	**1/192**	_
CDH18	**NM_004934.3:c.582C>T**	**Exon 5**	**p.(=)**	**Co-segregating**	**3/192**	**0.006**
CDH18	**NM_004934.3:c.644-71T>G**	**Intron 5**	**_**	**Co-segregating**	**3/192**	**0.002**
CDH12	**NM_004061.3:c.1394-11G>A**	**Intron 11**	**_**	**Co-segregating**	**0/192**	_
DROSHA	NM_013235.4:c.1091G>A	Exon 7	p.364R>H	not segregating	_	_
C1QTNF3	NM_181435.5:c.571-87G>T	Intron 3	_	not segregating	_	_
RAI14	NM_015577.2:c.37-50G>A	Exon 3	_	not segregating	_	0.001
IL7R	**NM_002185.3:c.538-30C>T**	**Intron 4**	**_**	**Co-segregating**	**0/192**	_
IL7R	NM_002185.3:c.1092T>G	Exon 8	p.364D>E	not segregating	_	freq not given
NIPBL	NM_133433.3:c.6589+9A>T	Intron 38	_	not segregating	_	_
C5orf42	**NM_023073.3:c.8044G>C**	**Exon 41**	**p.2682G>R**	**Co-segregating**	**0/384**	_

NUP155	**NM_153485.2:c.2413A>G**	**Exon 22**	**p.805I>V**	**Co-segregating**	**0/384**	**1 in 13,005**
RICTOR	NM_152756.3:c.4634-72T>A	Intron 34	_	not segregating	_	_
DAB2	NM_001343.3:c.2248-26T>C	Intron 13	_	not segregating	_	_
DAB2	NM_001343.3:c.625-3A>C	Intron 8	_	not segregating	_	_
PRKAA1	**NM_206907.3:c.364-43C>T**	**Intron 3**	**_**	**Co-segregating**	**3/192**	_
HEATR7B2	**NM_173489.4:c.3749+28T>C**	**Intron 34**	**_**	**Co-segregating**	**2/192**	**0.002**
ITGA1	NM_181501.1:c.1455+7T>C	Intron 12	_	not segregating	_	0.003
ITGA1	NM_181501.1:c.1857+95CCTC/-	Intron 14	_	not segregating	_	_
ITGA2	NM_002203.3:c.1594A>C	Exon 13	p.532I>L	not segregating	_	freq not given
GZMA	**NM_006144.3:c.*60G>A**	**Exon 5**	**_**	**Co-segregating**	**0/192**	**0.006**
CDC20B	**NM_001170402.1:c.1525G>A**	**Exon 12**	**p.509A>T**	**Co-segregating**	**0/384**	**0.001**
DDX4	**NM_024415.2:c.865-24C>T**	**Intron 13**	**_**	**Co-segregating**	**0/192**	**0.001**
IL31RA	NM_139017.5:c.773-26G>A	Intron 6	_	not segregating	_	_

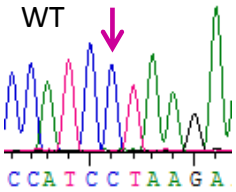
***IL7R*, c.538-30C>T**

A

c.538-30C>T



WT



B

c.538-30C

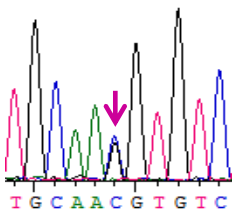
```
Chicken       TTCTATGCCTATTAGAAAG
Rehsus        TTCC-CATCC-TAAGAATG
Rat           TTTG-CTT-C-TAATAATG
Dog           TTTC-AATCC-TAAAAATG
Mouse         TTTA-CTTCT-----AATA
Pig           TTTTCCATCTTAAAGAATG
Human         TTCC-CATCC-TAAGAATG
Chimp         TTCC-CATCC-TAAGAATG
Cow           TTTC-CACCT-TAAAAATG
Horse         TTTC-CATCC-TAAATATG
                    **          :*:.
```

Conservation of c.538-30C in IL7R. This nucleotide is conserved in 6 out of 10 species.

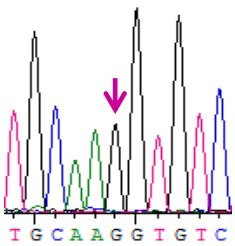
***C5ORF42*,**

A

c.8044G>



WT



B

p.2682G

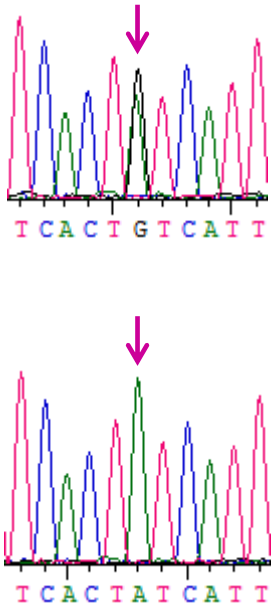
```
Zebrafish             DPVTLSVLAGIRAPRYS—
Desert Jerboas        ESLLWTLLQGVPAAQP---
Cow                   DYLLWEVLQEVPTARP---
Pig                   LYPLWELLQEVSETDPAPR
Degu                  GSLLRTLLQDVPA------
Squirrel              NYLLWKVLQDDPAAR----
Naked Mole Rat        GHLLRNLLQDVPA------
Human                 DYLLWKRLQGVSA------
Chimpanzee            DYLLWKRLQGVSA------
Macaque               DYLLWKRLQGVSA------
Horse                 DYLLWEVLHDVSV------
Dog                   DDLLWELLQDVSP------
Ferret                DDPPWELLQDVSA------
Vole                  QSLLWTLLQDVPTACPTPS
Brown Rat             QSLLWTLLQDASPACPTPS
Mouse                 QSLLWTLLQNASPACPTPS
Golden Hamster        KSMLLTLLQDVHTACPTP-
Painted Turtle        DNLTWNLLHEDVSIIHSTG
Chicken               ESVTWDIELEDARAFPSTR
Wild Duck             DSVTWNVVHEDARTFPSSG
```

Conservation of p.2682G in C5ORF42. This amino acid residue is conserved in 3 out of 9 species.

***NUP155*, c.2413A>G**

A

c.2413A>G



B

p.805I

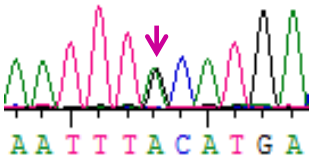
```
Zebrafish          KLLCDHQFSLILSEMPKEF
Frog               KLLCEHQFSLVVSDLQKEL
Bat                KLLCEHQFTVIVGELQKEF
Cat                KLLCEHQFTVIVGELQKEF
Pig                RLLCEHQFTVIVGELQKEF
Cow                KLLCEHQFTVIVGELQKEF
Horse              KLLCEHQFTVIVGELQK—
Dog                KLLCEHQFTVIVGELQKEF
Rat                KLLCEHQFTVIVGELQKEF
Mouse              KLLCEHQFSVIVGELQKEF
Chimpanzee         KLLCEHQFTVIVAELQKEL
Macaque            KLLCEHQFTVIVAELQKEL
Human              KLLCEHQFTIIVAELQKEL
Bonobo             KLLCEHQFTVIVAELQKEL
Alligator          KLLCEHQFNVVVGELQKEL
Turtle             KLLCEHQFNVVVGELQKEF
Sparrow            KLLCEHQFSVAVGELQKEL
Chicken            KLLCEHQFSVVVGELQKEL
                    :***:***.: :.:: *
```

Conservation of p.805I in NUP155. This amino acid residue is not conserved in any of the 18 species examined.

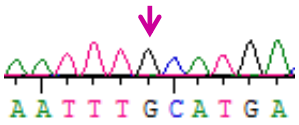
***GZMA*, c.*60G>A**

A

c.*60G>A



WT



B

c.*60G

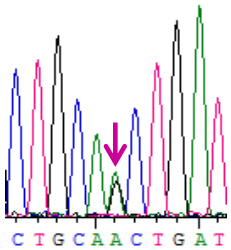
```
Zebrafish          TGACTGG-GGTGTAGCTGT
Frog               ATTCTGGT------GATGT
Turtle             ACTCAGT----------AGC
Chicken            ATGCT-TTTGAAAATAGGT
Cow                TGTCTTGAAGTATAGTTGA
Mouse              AATCAACTTGAATG-----
Rat                AATCAACTTAAAGAGCTGC
Dog                CTTCAAAT---------T
Horse              AATCAGTTTGTATGACTGT
Pig                AACCAATTTGTATCATTGT
Macaque            AATCAATTTGCATGACTGT
Human              AATCAATTTGCATGACTGT
Chimp              AATCAATTTGCATGACTGT
                      *:
```

Conservation of c.*60G in the GZMA gene. This nucleotide is conserved in 6 out of 10 species analyzed.

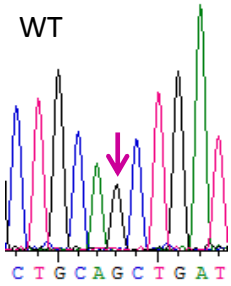
***CDC20B*, c.1525G>A**

A

c.1525G>A



WT



B

p.509A

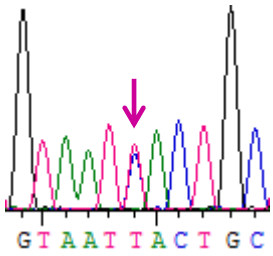
Finch	DHSRFIFVA**V**EGMSCLWKQ
Mouse	DQTRLFSAA**A**DGTACVWKC
Guinea Pig	DQTRVFSAA**A**DGTACVWKC
Cow	DQMRVFSAA**A**DGTACIWNC
Human	DQTWVFSAA**A**DGTASVWNC
Chimpanzee	DQTRVFSAA**A**DGTASVWNC
Rehsus Macaque	DQTRVFSAA**A**DGTASVWNC
Pig	DQTKVFSAA**A**DGIACVWNC
Dog	DQTRVFSAA**A**DGTACVWNC
Cat	DQTRVFSAA**A**DGTACVWNC
Horse	DQTRVFSAA**A**DGTACVWNC
Ground Tit	AQNSLIFVA**V**YGMSCL---
Chicken	DQRRLFSVA**A**DGIACLWKC
Frog	DQRRIFSAA**A**NGTANIWKY
Chinese Alligator	DCSRFISLA**A**DGVACVWKY
Chinese Turtle	DHCRIFSLA**A**DGLACVWKY
	. *. * : :

Conservation of p.509A in CDC20B. This amino acid residue is conserved in 14 out of 16 species.

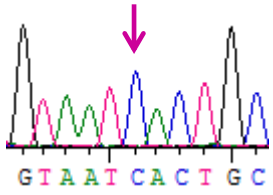
***DDX4*, c.865-24C>T**

A

c.865-24C>T



WT



B

c.865-24C

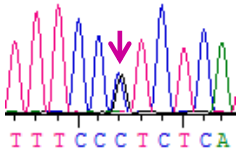
Rat	-------TT**G**ACTTCT---
Mouse	AATTGTT--**G**ACTTCT---
Human	GAATGTAAT**C**ACTGCTTTT
Chimp	GAATGTAAT**C**ACTGCTTTT
Macaque	GAATGTAAT**C**ACTGCTTTT
	*** **
Cat	GAATTTAAT**C**ACTGCCCCC
Cow	GAATTTAAC**C**ACTGCC—C
Horse	GAATTTA--**-**-----ATTC
Human	GAATGTAAT**C**ACTG-CTTT
Dog	GAATTTAAT**C**ATATTCCTC
	**** ** .
Turtle	TGTTAAAGA**C**TTTCTCTT-
Chicken	TGTACAATA**C**ATTATTTTT
Frog	CT---TAAA**A**ACTGACATT
Human	-----TAAT**C**ACTGCTTTT
Zebrafish	-----CAAA**C**TTTGTTG-T
	* :.: *

Conservation of c.865-24C in the DDX4 gene. This nucleotide is conserved in 9 out of 13 species.

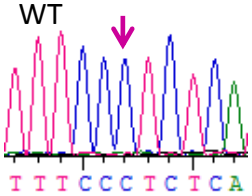
***DNAH5*, c.8952-49C>G**

A

c.8952-49C>G



WT



B

c.8952-49C

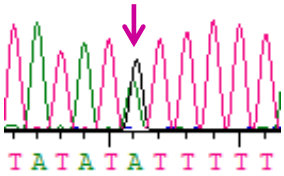
```
Chicken     TTTTTGTTCTACCCAGAAG
Frog        GGGTTGGCCTACCAAACTT
Chimp       GTTCTTTCCCTCTCATGAA
Mouse       --CGTGGCCTCCACATGGA
Rat         --TGTGGCCTTCTTATGGA
Dog         GATCTTTCCTTCCCATCAA
Sheep       GATCTTTCCTTCCCAAGAA
Human       GTTCTTTCCCTCTCATGAA
Macaque     GTTCTTTCCCTCTCATGAA
                    *   *   *   *
```

Conservation of c.8952-49C in the DNAH5 gene. This nucleotide is conserved in 3 out of 9 species analyzed.

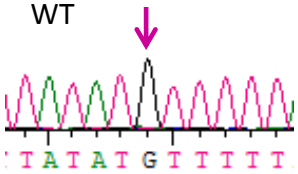
***CDH12*, c.1394-11G>A**

A

c.1394-11G>A



WT



B

c.1394-11G

```
Zebrafish   TGTTTGTGTGGTTTATCTA
Turtle      TCTTTCTAACATTTTTATA
Mouse       TTTTTTTATGTTTTAACA
Rat         TC--TTTATGTTTTAACA
Chicken     TCCTA-AAAATTATTTACA
Dog         TCTTTCTATGTTTTCAACA
Cow         TCCTTGTATGTTTTAACA
Pig         TCCTTTCATGTTTTTAACA
Human       TCTTTATATGTTTTTAACA
Chimp       TCTTTATATGTTTTTAACA
Macaque     TCTTTATATGTTTTTAACA
                    *   :   .:   *:* :. *
```

Conservation of c.1394-11G in the CDH12 gene. This nucleotide is conserved in 9 out of 11 species analyzed.

Figure 3.3: Novel, rare, co-segregating variant: A: Representative sequences of an affected individual and a normal individual. Arrow marks the nucleotide showing variation. **B:** The corresponding amino acid conservation for the variation.

3.4. Discussion

The present study was aimed at elucidating a plausible JME causing gene at the 5p15.33-q12.1 locus, identified in GLH35 by genome wide linkage analysis followed by whole exome based sequencing.

On analyzing all the variants picked in the whole exome analysis twelve novel, segregating variants: c.8952-49C>G in *DNAH5*, c.582C>T in *CDH18*, c.644-71T>G in *CDH18*, c.1394-11G>A in *CDH12*, c.538-30C>T in *IL7R*, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.364-43C>T in *PRKAA1*, c.3749+28T>C in *HEATR7B2*, c.*60G>A in *GZMA*, c.1525G>A in *CDC20B*, c.865-24C>T in *DDX4*, were identified. Upon checking the allele frequency for all these twelve variants in apparently normal control individuals, four of the variants were eliminated due to their high minor allele frequency (MAF>0.005). Conservation and mutation analysis was carried out for the remaining eight variants, namely, c.8952-49C>G in *DNAH5*, c.1394-11G>A in *CDH12*, c.538-30C>T in *IL7R*, c.8044G>C in *C5ORF42*, c.2413A>G in *NUP155*, c.*60G>A in *GZMA*, c.1525G>A in *CDC20B*, c.865-24C>T in *DDX4*. The only variation which stands out at this point of the study is c.1525G>A in *CDC20B* which gives rise to a residue change of p.509A>T. This residue is highly conserved across species and bioinformatic prediction studies show that this residue change has a deleterious effect on the *CDC20B* protein.

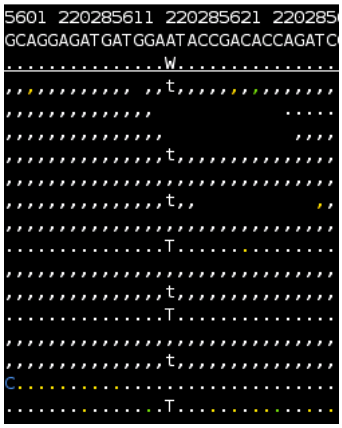
CDC20B is a 519 amino acid cell division cycle protein 20 homolog B protein. It belongs to the WD repeat family and has 7 WD repeats. Not much literature is available for this protein. In order to show that this is the disease causing factor in the family GLH 35, all the missing exons need to be covered to eliminate the presence of any other candidate variant.

Appendix I

NGS read alignments showing rare, co-segregating variants

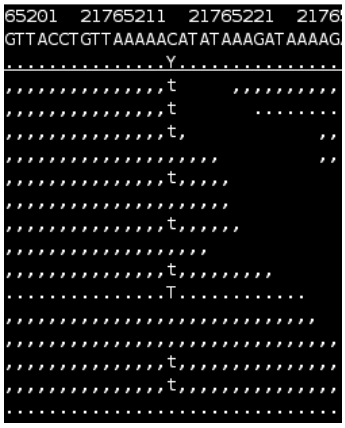
(A) In locus 2q33-q36 (EJM9)

***DES*, c.966A>T**

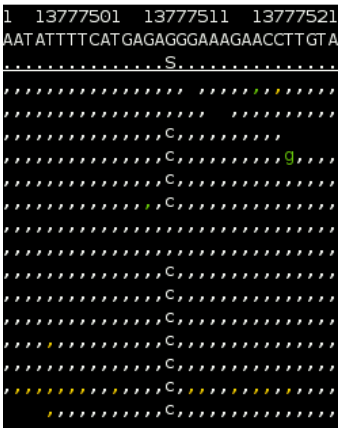


(B) In locus 5p15.33-q12.1

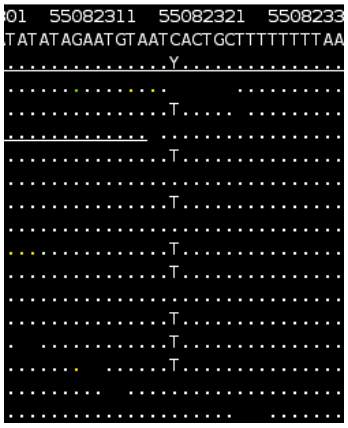
***CDH12*, c.1394-11G>A**



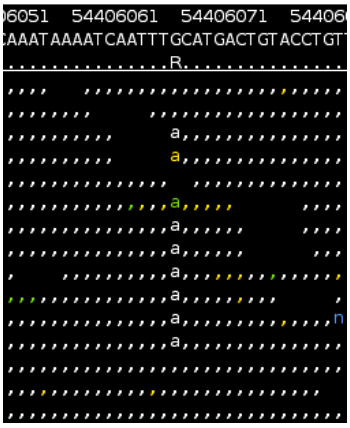
***DNAH5*, c.8952-49C>G**



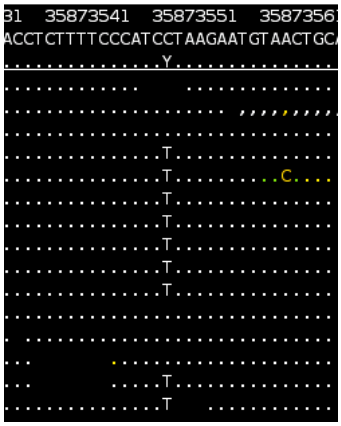
***DDX4*, c.865-24C>T**



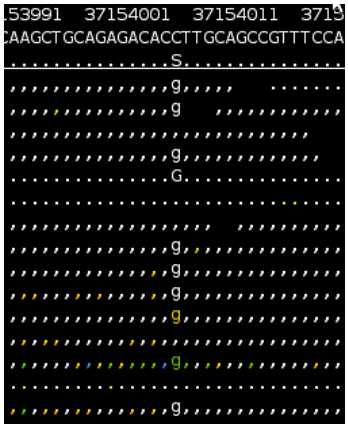
***GZMA*, c.*60G>A**



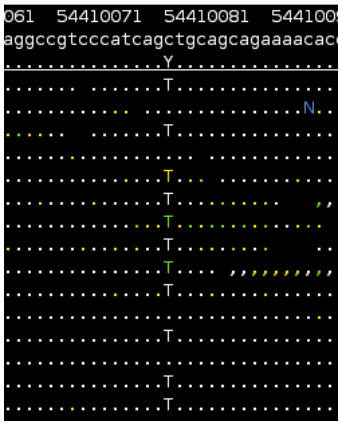
***IL7R*, c.538-30C>T**



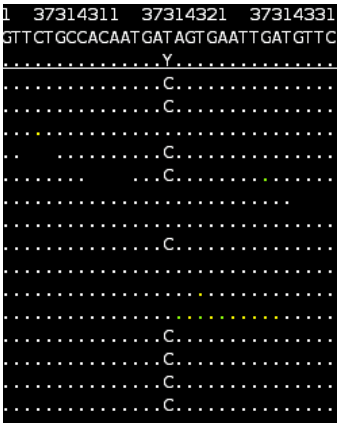
***C5ORF42*, c.8044G>C**



***CDC20B*, c.1525G>A**



***NUP155*, c.2413A>G**



Appendix II

Primer sequences

(i) Primers for Sanger validation of variants identified by the whole-exome sequencing analysis of the EJM9 locus.

OligoName	5'<-----Sequence----->3'	Length
ACADL-Ex10-F	ctggttttgaagcacctc	18
ACADL-Ex10-R	agattgcacacctcctttcc	20
SLC19A3-Ex2-F	catgcaaaacccagcagag	19
SLC19A3-Ex2-R	tgcttacctccaaaggttgc	20
DNAH7-Ex44-F	ggtgggactgaggagaaaag	20
DNAH7-Ex44-R	acccatttgcacagaccttc	20
STK16-Ex7-F	tgcagaaccaactcagcatc	20
STK16-Ex7-R	gatggaatgggtggtgattc	20
OBSL1-Ex5-F	tggtgttccacggttctg	18
OBSL1-Ex5-R	gcaggatgagtctgtgctg	19
COL4A4-Ex21-F	tctgccattgatcctctgtc	20
COL4A4-Ex21-R	tgtcacctgctgtcgtgtaag	21
CXCR2-Ex1-F	cctatcccagtttcttgagtgg	22
CXCR2-Ex1-R	aggttcagcaggtagacatcag	22
HECW2-Ex9-F	cgttatgggtgagggatgag	20
HECW2-Ex9-R	ggaatgggtctgcctgtc	18
SPATS2L-Ex4-F	ctccccagtggcaggataac	20

OligoName	5'<-----Sequence----->3'	Length
MOB4-Ex6-F	gacacacacttgatggtgctg	21
MOB4-Ex6-R	agactgaactggccttgagc	20
ORC2-3'UTR-a-F	aaggaagaagaggaggcttg	20
ORC2-3'UTR-a-R	aggagaaagctggatcactcac	22
ORC2-3'UTR-b-F	ttttaacactccccgagctg	20
ORC2-3'UTR-b-R	caacagtagccaaggtgagc	20
ORC2-3'UTR-c-F	ttgtctgtgccatgaccttc	20
ORC2-3'UTR-c-R	ggcctcccaaagtgttgtg	19
WDR12-5'UTR-F	agtctactcaacctcgtcttgg	22
WDR12-5'UTR-R	cagaccacaaacatcgactacc	22
CYP20A1-5'UTR-a-F	gccatctcggctcactgtag	20
CYP20A1-5'UTR-a-R	gccagaggaaagagcaatgg	20
CYP20A1-5'UTR-b-F	agtcgcgggcatattctg	18
CYP20A1-5'UTR-b-R	cgctcaccggatagaggtag	20
FASTKD2-Ex7-F	gttccccagacagttcgtg	19
FASTKD2-Ex7-R	cctttccacagcaacacctg	20
RPE-5'UTR-F	agtcaagagccgaggagagg	20

SPATS2L-Ex4-R	aacagcagccccaatgag	18	RPE-5'UTR-R	ggagttacccgtccattacg	20
CCNYL1-5'UTR-F	gggaactctgaccatgattg	20	C2orf67-Ex8-F	gagtgaaacagcattaggtagg	22
CCNYL1-5'UTR-R	ccctctccgaaatccaactc	20	C2orf67-Ex8-R	agggaacataggagggaaatac	22
GPBAR1-5'UTR-I-F	ccgagctggagtagggaaac	20	CPS1-Ex31-F	ggccaggttatctctcatgc	20
GPBAR1-5'UTR-I-R	ggctgtaccacctgcacac	19	CPS1-Ex31-R	tctccaggtgtctgtgatgc	20
GPBAR1-5'UTR-IIa-F	tcctcttccctctcctctgg	20	SPAG16-Ex15-F	gaaaatgcttgtctgaaacg	20
GPBAR1-5'UTR-IIa-R	ccacccagccagctttatc	19	SPAG16-Ex15-R	aaattcacctcattgccaggac	22
GPBAR1-5'UTR-IIb-F	gctggctggaagaccactac	20	TNS1-Ex3-F	ggcttctgccctcttctcac	20
GPBAR1-5'UTR-IIb-R	gcagtaggctcaggaagaagc	21	TNS1-Ex3-R	gctgatgtcccctacacttc	20
GPBAR1-Ex1-F	ccctgctctttgccagtc	18	ARPC2-5'UTR-Ex1-F	gcatcacacatacgcaagc	19
GPBAR1-Ex1-R	gacactgctttggctgcttg	20	ARPC2-5'UTR-Ex1-R	gacaacattggggattagctg	21
AAMP-Ex4-F	gggccagataccagaatcac	20	CTDSP1-Ex7-F	ctgggcaacagagcaagac	19
AAMP-Ex4-R	ggggactttccaggtagcag	20	CTDSP1-Ex7-R	cgaagaaggggaggaggtc	19
PAX3-5'UTR-Ex1-F	ccaatcagcgcgtgtctc	18	PRKAG3-Ex4-F	actacctggggcgactctc	19
PAX3-5'UTR-Ex1-R	cctggaagcaccaaaggag	19	PRKAG3-Ex4-R	cctctgtggctgggaactc	19
PAX3-Ex2-F	tgaaaggaggtctgggtctg	20	ABCB6-Ex17-F	ccttcctgccttgtgtatgc	20
PAX3-Ex2-R	cagatgtcagccgttaccc	19	ABCB6-Ex17-R	tctcttaccgtcccctctcc	20
DNAH7-Ex62-F	ggcagctatgtgaatgacttcc	22	ATG9A-5'UTR-I-F	agcagcgaagaggacaacc	19
DNAH7-Ex62-R	agttctgggcaatgaagc	18	ATG9A-5'UTR-I-R	ctcccaacagcggacaac	18
HECW2-Ex11-F	gttccccaaacaatggacag	20	ATG9A-Ex1-F	caaggaggctggtagtggac	20
HECW2-Ex11-R	ccagtatcccacagatcaag	20	ATG9A-Ex1-R	aggtgctgtgggataggaac	20
TRAK2-Ex11-F	gctgagattgaggggactatg	21	STK16-5'UTR-I-F	acccgtccacccctacac	18
TRAK2-Ex11-R	gctggaatgtgcgtatgtg	19	STK16-5'UTR-I-R	ggtgagtgggctcttccag	19

ALS2-Ex4-F	ctccctccctttttactgtg	20
ALS2-Ex4-R	gtttagggctgaggtgcttg	20
CDK15-Ex8-F	ctgccaagaaggccacatc	19
CDK15-Ex8-R	aagcaaccctcgtgagtagg	20
IDH1-Ex6-F	taggcagttggaccctgaac	20
IDH1-Ex6-R	actgagcagccaagggaac	19
UNC80-Ex44-F	tagccagtcttgcagcag	18
UNC80-Ex44-R	gatggtagccaactcccttg	20
XRCC5-Ex7-F	gccatctcctgctgtttctc	20
XRCC5-Ex7-R	tctcctagcctgccttttacac	22
PRKAG3-UTR3'-I-F	agacgctctctccctcagtc	20
PRKAG3-UTR3'-I-R	cagttttcacagggctgctc	20
CUL3-Ex6-F	aggtggttgaaagggaactc	20
CUL3-Ex6-R	ctcactgggctcaccaatac	20
DOCK10-Ex54-F	tcatgccactgcactctagc	20
DOCK10-Ex54-R	attcttgggggattgaggac	20
DOCK10-Ex28-F	atgagaggccacagagaagg	20
DOCK10-Ex28-R	cttgcatccaccaacatgg	19
COL4A4-Ex46-F	cagggaggtgtcagaaatgg	20
COL4A4-Ex46-R	gtgaatgagccagggttttc	20
WDR69-3'UTR-F	gcgttgactgaaggaagctg	20
WDR69-3'UTR-R	gcactgggattacaggtgtg	20

SPEG-Ex20-F	agaggcgtggttaggaggag	20
SPEG-Ex20-R	accagggagcactcattctc	20
SPEG-Ex21-F	tctgtccacctgtcccagtc	20
SPEG-Ex21-R	ccgcaagtaggagaaagcac	20
ACCN4-Ex5-F	gaccacccttctccaactcc	20
ACCN4-Ex5-R	cacactcacacacccacaag	20
OBSL1-Ex8-F	cagttccacccttcataccg	20
OBSL1-Ex8-R	cctcctctccatccttggtc	20
PAX3-3'UTR-a-F	aaggcaatggtttcacatgg	20
PAX3-3'UTR-a-R	ctggaaaaacgtcacacacc	20
CUL3-3'UTR-a-F	gcctttactccctttgagag	20
CUL3-3'UTR-a-R	caccaccctatacaatccac	20
CUL3-Ex2-F	tctgctccaggtaagattgg	20
CUL3-Ex2-R	gtccagtgtagagcttttctcc	22
COL4A4-Ex26-F	aatcgggataggcaacatcc	20
COL4A4-Ex26-R	gctttccttctctggcaaac	20
AGFG1-Ex10-F	cctgtgtccatgtgttgttc	20
AGFG1-Ex10-R	gatgaagcaggctgtgtctg	20
PID1-3'UTR-a-F	attggtcacggagttcaagg	20
PID1-3'UTR-a-R	ggctgtgctgatttgcatac	20

(ii) Primers for covering the exons in the EJM9 locus missed by the whole exome sequencing experiment.

OligoName	5'<-----Sequence----->3'	Length
BZW1-Trns1,2-Ex1-F	ttctggctctttcctcttcg	20
BZW1-Trns1,2-Ex1-R	cggtaacgtgtctcccaatc	20
PPIL3-Ex3-F	ctgtgagaggacacccaaaac	21
PPIL3-Ex3-R	aaaaggccaggtgcagtg	18
TMEM237-Trns1-Ex1-F	gactcgttggtcctggagag	20
TMEM237-Trns1-Ex1-R	cccttagtgattcccagctc	20
ALS2-Ex15-F	ggcacctacatgggcttaac	20
ALS2-Ex15-R	tctctgctctgtgctgaacg	20
SUMO1-Ex1-F	gaaggagctgacaaaactgc	20
SUMO1-Ex1-R	agaagtgggacgacatgagg	20
FAM117B-Ex1a-F	cttcgtcaccccgtcttg	18
FAM117B-Ex1a-R	gctggggctcctcttctc	18
FAM117B-Ex1b-F	acaacggtggctgctgtg	18
FAM117B-Ex1b-R	ctcatcagctcaagcccaag	20
ABI2-Ex1-F	taggagacgccggaagtg	18
ABI2-Ex1-R	attcccacccattcaccag	19
RAPH1-Ex13-F	caccaccaacttctgcatcc	20
RAPH1-Ex13-R	ggcactgaactgtggaagg	19
PARD3B-Ex1-F	ggcagtttcgctttggtg	18
PARD3B-Ex1-R	ttctgcatcccctgagtgtc	20

OligoName	5'<-----Sequence----->3'	Length
CCDC108-Ex6-R	gaacagcccctcgaagaag	19
CCDC108-Ex15-F	cccgtggggatataggattg	20
CCDC108-Ex15-R	aagggcaagtttggggtatc	20
CCDC108-Ex32-F	gtagtggatgaggggagtcg	20
CCDC108-Ex32-R	ggtggagtacctctgcttgc	20
CCDC108-Ex33-F	tctgcgtgcccaggtaag	18
CCDC108-Ex33-R	tgctccctgccatccttc	18
IHH-Ex1-F	gggccggcctatttattg	18
IHH-Ex1-R	gtaattgggggtgagctcct	20
C2orf24-Ex1-F	acgcttgacacctcccttc	19
C2orf24-Ex1-R	ccctacgcccattcctatac	20
FAM134A-Ex1-F	ctgtgtaggcgcagtgtcag	20
FAM134A-Ex1-R	gaaaggggaatggtgctctc	20
ZFAND2B-Ex1-F	gtcttccgactcagccttct	20
ZFAND2B-Ex1-R	cttcccaagcccattacctc	20
ABCB6-Ex5-F	cctttcacatcctggtgctg	20
ABCB6-Ex5-R	cggggtctgttctcttcctc	20
TUBA4A-Ex1-F	ccagcgtgtctgctcaaaac	20
TUBA4A-Ex1-R	aaaaccctcgcacctcctg	19
PTPRN-Ex1-F	gaagagggctacagggatgg	20

PARD3B-Ex4-F	ctggggaaagatgagagttg	20	PTPRN-Ex1-R	ctcccccaacccatattctc	20
PARD3B-Ex4-R	ccttcctctgcctcaatttc	20	PTPRN-Ex19-F	gtggttggagagaccgagaa	20
NDUFS1-5'UTR-F	cgggtccaagttgtccttc	19	PTPRN-Ex19-R	gtggaggagaaggagccagt	20
NDUFS1-5'UTR-R	cctaagtcatcggacactgg	20	PTPRN-Ex22-F	gttttctgggtcctgtcctc	20
ADAM23-Ex1-F	acccctggactcctctgc	18	PTPRN-Ex22-R	acctgtgctctgccctcaag	20
ADAM23-Ex1-R	cgagaagggtggaaagacag	20	RESP18-Ex1-F	tagaggtcgagcggaggtt	19
FZD5-Ex-1a-F	ctggaatccgagccctaac	19	RESP18-Ex1-R	cgcctgtactcccagtccta	20
FZD5-Ex-1a-R	cggttgtaatccatgcagag	20	RESP18-Ex2-F	gagtacaggcggtggttctt	20
FZD5-Ex-1b-F	ctgcgcttcttcctatgctc	20	RESP18-Ex2-R	ctctttccgctagacgctgt	20
FZD5-Ex-1b-R	acaggtagcaggctgacagg	20	RESP18-Ex7-F	actggttgccctggtgct	18
FZD5-Ex-1c-F	atccgcagcgtcatcaag	18	RESP18-Ex7-R	agttgtgctcccctcatcc	19
FZD5-Ex-1c-R	ccctcttccctctctccaag	20	DNPEP-Ex1-F	acgaggaagcttgacgagg	19
UNC80-Ex1-F	acagtgggaggtgctgaaag	20	DNPEP-Ex1-R	ggtttcgctttgggtcagg	19
UNC80-Ex1-R	gtgcagagggcttgttttcc	20	DES-Ex1a-F	caggacagcgggatcttg	18
UNC80-Ex60-F	gacaaacccagcacaaggtg	20	DES-Ex1a-R	gcgcaccttctcgatgtagt	20
UNC80-Ex60-R	taagacggggcgagaagag	19	DES-Ex1a-F	gctgctggacttctcactgg	20
ACADL-Ex1-F	gtatttgggggctccatagc	20	DES-Ex1a-R	acaggtggaggacccttttct	20
ACADL-Ex1-R	gctgacacccctttttcctc	20	SPEG-Ex1-F	cccccagacttgtctccta	19
SPAG16-Ex1-F	tatcttgtccgctcccagag	20	SPEG-Ex1-R	ggcggaaccagtatccagta	20
SPAG16-Ex1-R	tggctctcaggaagactgtg	20	SPEG-Ex3-F	ctggggtgtacaaagagcag	20
BARD1-Ex1-F	ccctgcgagtccctattttg	20	SPEG-Ex3-R	atgggtggaggctgactg	18
BARD1-Ex1-R	gggaacggaaggaggaaac	19	SPEG-Ex4a-F	acccaattcctgtcacaagc	20
ATIC-Ex1-F	gtggagtggcctcactttg	18	SPEG-Ex4a-R	acttgggctgctccagaga	19

ATIC-Ex1-R	ggacgctggctttcaatc	19	SPEG-Ex4b-F	gacaagctgcagttcttcgag	21
FN1-Ex1-F	accttcttggaggcgacaac	20	SPEG-Ex4b-R	attggtcccgacgcctaac	19
FN1-Ex1-R	cacaaaacttcagccccaac	20	SPEG-Ex4c-F	cctccacccccaagacat	18
FN1-Ex42-F	tatgtggtgtttgcgctgtg	20	SPEG-Ex4c-R	gcagtcttgctctccacaaa	20
FN1-Ex42-R	agcagttgtatgccaacagg	20	SPEG-Ex6-F	aggtttggtctcctgtgtgg	20
MREG-Ex1-F	gtgccctgggattttgag	18	SPEG-Ex6-R	tcaaggctgagagtgggaag	20
MREG-Ex1-R	acctccccaactcacacaag	20	SPEG-Ex8-F	tgctcccattcaaaccctct	20
March4-Ex1-F	cccacaacacagatccactg	20	SPEG-Ex8-R	gaagcccaccaagattccat	20
March4-Ex1-R	taccactgtccaagctgctg	20	SPEG-Ex30a-F	gttccctgaccctctgcat	19
SMARCAL1-Ex1-F	gaaatggggtggagaggaac	20	SPEG-Ex30a-R	catactcgccctcacccag	19
SMARCAL1-Ex1-R	ctgcaagccaccagtgaac	19	SPEG-Ex30b-F	ctccctctcaggaccagga	19
IGFBP2-Ex1-F	ggggaagggagtggtctc	18	SPEG-Ex30b-R	taggtgtggtggcagaaggt	20
IGFBP2-Ex1-R	cccctaaaaccgtccgtaag	20	SPEG-Ex30c-F	gagtctccttccctgtctgc	20
IGFBP2-Ex2-F	tcatcattacggtccaggtg	20	SPEG-Ex30c-R	gtatatgccatcctcctcctc	21
IGFBP2-Ex2-R	tactgactgccccaaaggtc	20	SPEG-Ex30d-F	tcgaggccaagttcaagc	18
IGFBP5-Ex1-F	cctcttggcccctttatcc	19	SPEG-Ex30d-R	ctcctcccacagcctctctc	20
IGFBP5-Ex1-R	aaggaccctccccgactac	19	SPEG-Ex41-F	agactcactgtccccattcc	20
TNS1-Ex15-F	ctcactaagcgtgccctcac	20	SPEG-Ex41-R	gtaagagcccagccagatgt	20
TNS1-Ex15-R	cacaaagctggctgctgac	19	GMPPA-Ex10-F	gtggtcggatggatggag	18
TNS1-Ex18-F	tccaggaagctgtgttccag	20	GMPPA-Ex10-R	gcgttcacccttctctgttc	20
TNS1-Ex18-R	gccataggcaacagatccag	20	ASIC4-Ex1-F	agaatgagctgaggaccctg	20
TNS1-Ex19-F	caccagatgatgggtccac	19	ASIC4-Ex1-R	ttgcacacaatctcgatcgg	20
TNS1-Ex19-R	cgggaaccacagatccag	18	ASIC4-Ex7-F	tgctgggtgagactggtgt	19

RUFY4-Ex8-F	tctcgggaatgttgttcagg	20	ASIC4-Ex7-R	ccctgaacctgactttccag	20
RUFY4-Ex8-R	tgggaggttctcacaagacc	20	CHPF-Ex1-F	tgctggaggggaatcgag	18
AAMP-Ex9-F	gatggggaaaggggtttg	18	CHPF-Ex1-R	cttccttccggagcctgac	19
AAMP-Ex9-R	tctggggtagatgctcctg	19	CHPF-Ex2-F	ctaggacccgctacatcagc	20
PNKD-Ex1-Trns1-F	gcggagagaacccaaactc	19	CHPF-Ex2-R	ccaacatccctttgcctctc	20
PNKD-Ex1-Trns1-R	cctgacctcctgctatcgtc	20	CHPF-Ex4-F	gctaccgacgctttgatcc	19
PNKD-Ex1-Trns2-F	ctgtggaccccgatcagc	18	CHPF-Ex4-R	acgtgggccttgacaggt	18
PNKD-Ex1-Trns2-R	gtctcctcgatccctcttcc	20	CHPF-3'UTRa-F	gctgttcctccacttctcca	20
PNKD-Ex6-Trns1-F	cctggagatgctgtggtaaag	21	CHPF-3'UTRa-R	cctccttccccaacaactct	20
PNKD-Ex6-Trns1-R	agatccacctgctgataccc	20	TMEM198-Ex2-F	agcgggtgctagagacacag	20
TMBIM1-Ex1-F	ctggacaaggctggaagtg	19	TMEM198-Ex2-R	agtcacttggcccccaac	18
TMBIM1-Ex1-R	cccattctcctgtggtgtg	19	OBSL1-5'UTR-F	cagtctgggctcttgtcctc	20
C2ORF62-Ex10-F	taggggctagaaggctccag	20	OBSL1-5'UTR-R	aggaccacgcacttgagc	18
C2ORF62-Ex10-R	aggtacgacggtaacggttc	20	OBSL1-Ex1a-F	ctgtgcgggtggtaagtg	18
CTDSP1-Ex1-Trns1-F	gggaaggaaactccatgttg	20	OBSL1-Ex1a-R	acacgtagacgccggaatc	19
CTDSP1-Ex1-Trns1-R	ttcaaactctctccctgctc	20	OBSL1-Ex1b-F	acgaagtgtgggacagcag	19
CTDSP1-Ex1-Trns3-F	ggtagacccgaagcacgtc	19	OBSL1-Ex1b-R	ccacctctcctccagtcttg	20
CTDSP1-Ex1-Trns3-R	aacacccctcggcacttc	18	OBSL1-Ex2-F	gcatgaagagcacacagca	19
CYP27A1-Ex1-F	gcccagagttcagaccaagc	20	OBSL1-Ex2-R	ctgagatgcggacaggaatc	20
CYP27A1-Ex1-R	gctgtcctagacgtggaatctc	22	OBSL1-Ex12-F	tctcagtaccttccccgtca	20
PRKAG3-Ex6-F	tctcagcacaaggacactgg	20	OBSL1-Ex12-R	ggagccagaagcagcaaag	19
PRKAG3-Ex6-R	cccaccatcaccaacagc	18	OBSL1-Ex17-F	gggaacacgagacacgcata	20
WNT 6-Ex1-F	ccggctctgatttcttctcc	20	OBSL1-Ex17-R	cggagagtaccgccacagta	20

WNT 6-Ex1-R	gaacaccccagtctgtcctg	20	OBSL1-Ex18-F	atgtggaggctgggcact	18
WNT6-Ex3-F	ttgcctgagccccacttc	18	OBSL1-Ex18-R	gggaaagaacagggacgag	19
WNT6-Ex3-R	actcctgccaacacttcctc	20	OBSL1-Ex19-F	agcctcgtccctgttctttc	20
WNT6-Ex4a-F	cacctcccattcccaatc	18	OBSL1-Ex19-R	ttagccctctatgccaccag	20
WNT6-Ex4a-R	aggcagttctcttcgagctg	20	OBSL1-Ex20-F	tggtggcatagagggctaag	20
WNT6-Ex4b-F	gattcgcccgacttctgc	18	OBSL1-Ex20-R	cagttccagggctttccag	19
WNT6-Ex4b-R	gggagcccagtatccagag	19	INHA-Ex1-F	gactggggaagactggatga	20
WNT6-3'UTR-F	aggggcttgagaggaacg	18	INHA-Ex1-R	cctgcaaacccatgctgt	18
WNT6-3'UTR-R	gaatgccaaggggagatagc	20	STK11IP-Ex1,2-F	gttttccggtcgtcccttg	19
WNT10A-Ex1-F	gagtcggagctgtgtgtcg	19	STK11IP-Ex1,2-R	atgactctcagggcgtgtct	20
WNT10A-Ex1-R	tcttccagggtcctctaccc	20	STK11IP-Ex14-F	gcttgctcagttctgggttc	20
WNT10A-Ex4a-F	ggagtgggtttcagaagcag	20	STK11IP-Ex14-R	ccactcacaaacatcccaag	20
WNT10A-Ex4a-R	gtctggcgcaggatgttg	18	SGPP2-Ex1-F	gcaaggtggaggcagaca	18
WNT10A-Ex4b-F	tctcccgacttctgcgag	18	SGPP2-Ex1-R	gcatcctggttactggaagg	20
WNT10A-Ex4b-R	tccattcattccccacctcc	20	FARSB-Ex1-F	cgggacttcagggtcagta	19
CDK5R2-Ex1a-F	agctgcccatcaggagtg	18	FARSB-Ex1-R	cggagccaaaaccttcag	18
CDK5R2-Ex1a-R	gaaggttctcgcggttgc	18	MOGAT1-Ex1-F	agcctctgccttttcctctc	20
CDK5R2-Ex1b-F	gtccgccaagaagaagaaag	20	MOGAT1-Ex1-R	ctctttgacctgcctgctct	20
CDK5R2-Ex1b-R	gtttgcaggcgtaatgaagg	20	AP1S3-Ex1-F	agggaggagaaaggggaaag	20
CDK5R2-Ex1c-F	ggcaacgagatctcctaccc	20	AP1S3-Ex1-R	ccggcacagactaagcactc	20
CDK5R2-Ex1c-R	ggaatgggctgaagggaagg	20	WDFY1-Ex1-F	tctcccagccacagactttc	20
FEV-Ex1-F	gatgggacgataagaggggc	20	WDFY1-Ex1-R	agctaaggggcgcagagt	18
FEV-Ex1-R	cactctcttcccatgcctga	20	MRPL44-Ex1-F	cgcaagcgtagcctcaag	18

FEV-Ex2-F	cggggccctttgtcaag	18
FEV-Ex2-R	cacactgctcccacctact	19
FEV-Ex3a-F	tttcccgccagcactctt	18
FEV-Ex3a-R	aagctgggactggggtaga	19
FEV-Ex3b-F	gcctctccaaactcaacctc	20
FEV-Ex3b-R	tgaatggggcttctaggagc	20
CRYBA2-Ex1-F	gctgtgtgtggctcgaaac	19
CRYBA2-Ex1-R	atgttggcaggtctctccag	20
CRYBA2-Ex2-F	gtgtcaggggaagggtttg	19
CRYBA2-Ex2-R	atgcagggcttgaacagc	18
CCDC108-Ex4-F	cctcgtttcctgcctcttc	19
CCDC108-Ex4-R	ctcctggttgctctgtcctg	20
CCDC108-Ex6-F	tgcccagcatccctgact	18

MRPL44-Ex1-R	atcagccgcagacacgac	18
SERPINE2-Trns4-Ex1-F	ccctgaccctgaacctga	18
SERPINE2-Trns4-Ex1-R	cctgaaggtggagtgctgtc	20
FAM124B-Trns2-Ex2-F	ttgggctcaggaatgtcac	19
FAM124B-Trns2-Ex2-R	agctcggccaacatagtgag	20
CUL3-Trns1-Ex1-F	tcactctccggctctcct	18
CUL3-Trns1-Ex1-R	ctgttgggggacttcagc	18
CUL3-Trns3-Ex1-F	actgccattcctcagatgct	20
CUL3-Trns3-Ex1-R	tccttacctccctccaatcc	20
DOCK10-Ex1-F	cgggtggatagagaaggttg	20
DOCK10-Ex1-R	cattaaagcccctgcacatc	20

(iii) Primers for Sanger validation of variants identified by the whole-exome sequencing analysis of the 5p15.33-q12.1 locus.

Primer Name	5'<-----Sequence----->3'	Length
ZDHHC11-3'UTR-F	agccacctgcttaactgtgc	20
ZDHHC11-3'UTR-R	tgggctaggctgaaaaactc	20
Med10-Ex4-F	tggagtgttgatgtgaagagc	21
Med10-Ex4-R	gcactctgaaagccagttgac	21
DNAH5-Ex54-F	gttccaggttgacttttcag	20

Primer Name	5'<-----Sequence----->3'	Length
ITGA2-Ex13-F	aggaactgtgctctctgtcttc	22
ITGA2-Ex13-R	ggggacatcctcaaaaatg	19
PPAP2A-Ex3-F	gctgggattacaggcatgag	20
PPAP2A-Ex3-R	aacactcacctgccttcc	18
IL31RA-Ex7-F	gattccagttccttgaccac	20

DNAH5-Ex54-R	aatacccatccccaatagcac	21	IL31RA-Ex7-R	gaagggaaggaggatggaag	20
CDH18-Ex3-F	tgctaaatgccatcacacac	20	PLEKHG4B-Ex11-F	caggaaacatccccagaag	19
CDH18-Ex3-R	gcaacttactgtagcaaacg	20	PLEKHG4B-Ex11-R	actgctaggacgagaggag	19
CDH18-Ex4-F	tttggggcatgggataag	18	ZDHHC11-5'UTR-F	agtgtggcctctttctgacg	20
CDH18-Ex4-R	tcagaagacaatagctggagtg	22	ZDHHC11-5'UTR-R	ccgttcactctggagatgc	19
CDH12-Ex8-F	ccctggtatgatggctaagtc	21	SLC12A7-Ex18-F	ctggaagcaggaggacaacc	20
CDH12-Ex8-R	tgcggtatcacctccacatac	21	SLC12A7-Ex18-R	gcagagggggacagtgag	18
IL7R-Ex5-F	ttgctgttgactcctttacg	20	TERT-Ex13-F	ggcaggcagatgacacagag	20
IL7R-Ex5-R	acttgctcccacactttgac	20	TERT-Ex13-R	caggagttccaaggtgaagc	20
C5orf42-Ex40-F	cccagaaatcaagaagctg	19	NDUFS6-Ex1-F	ctgggatgaaaacgggtgac	20
C5orf42-Ex40-R	ctgaggttgacaaaccctagc	21	NDUFS6-Ex1-R	agcgacagcacaaccttacc	20
NUP155-Ex22-F	atgggggctaggaagaaaac	20	NSUN2-EX9-F	gaaggtggaaggatggtg	18
NUP155-Ex22-R	ccctccaaagccaaaatactc	21	NSUN2-EX9-R	agatggatggtggtggtg	18
PRKAA1-Ex4-F	cccccagaactcataatcctc	21	DNAH5-Ex52-F	gaggctgatgctgaaacacc	20
PRKAA1-Ex4-R	aggggcttttgcataccac	19	DNAH5-Ex52-R	ctccgtggtgaaagcactg	19
HEATR7B2-Ex34-F	tagtccaggggtgggaaag	19	FAM134B-3'UTR-F	caagaccaggaagcagaagc	20
HEATR7B2-Ex34-R	tgaggtgagtggtgtgtgtg	20	FAM134B-3'UTR-R	agagatggcagtcaatgg	18
GZMA-Ex5-F	gtcaaggttggtcttaactgc	21	MYO10-Ex27-F	aagctcaagggcaccgtag	19
GZMA-Ex5-R	acgcacaaatgactctggtg	20	MYO10-Ex27-R	accaggcaacttccagatcc	20
CDC20B-Ex12-F	cactgagggagaaaactgtcc	21	PRDM9-Ex3-F	agccactcgaccagctttc	19
CDC20B-Ex12-R	tgggaagcagagcaagtaaag	21	PRDM9-Ex3-R	aggctgaggcaggagaatc	19
DDX4-Ex13-F	cttctgagtggaggcatgttac	22	PDZD2-5'UTR-F	tgcttggtgcctggaaag	18
DDX4-Ex13-R	ccaccccaggaatagttcag	20	PDZD2-5'UTR-R	gcattgtcctgggtgatg	18

SLC6A18-Ex5-F	cagccactctgaccacaagg	20	C1QTNF3-Ex4-F	gggtttctcaggtccttg	18
SLC6A18-Ex5-R	ctcacccctgaccaccag	18	C1QTNF3-Ex4-R	ctgaagtgggttgccagag	19
EXOC3-Ex1-F	gagttcctcatcctagttcagc	22	NIPBL-Ex37-F	cctcagactggctgctaatcc	21
EXOC3-Ex1-R	ctgcctccttgcacattacc	20	NIPBL-Ex37-R	gctcaaggttacccaggaag	20
SLC9A3-Ex16-F	gtgtggggtcatggatgtg	19	C5ORF42-Ex20-F	tcccttccttcactgagtcc	20
SLC9A3-Ex16-R	ggtccagggaggagagagac	20	C5ORF42-Ex20-R	ggataggttcagcccaggag	20
SLC9A3-Ex12-F	cgtggggagtcagcctaag	19	GDNF-5'UTR-Trns-2-F	gtctccaagtccctgctaac	20
SLC9A3-Ex12-R	atcaggcacggaggtcac	18	GDNF-5'UTR-Trns-2-R	ggtagttcccacccttcgtc	20
TRIP13-Ex5-F	gctccctcttctcatgtagg	20	RICTOR-Ex5-F	gagcagcagggctacgag	18
TRIP13-Ex5-R	gagctggatctgcttcacag	20	RICTOR-Ex5-R	ggttgtgtgtgtgtgtgtgtg	21
SLC12A7-Ex9-F	gaattggagcctgcttgc	18	DAB2-Ex1-F	gcctagcaagttttcgatcc	20
SLC12A7-Ex9-R	ctgcttgggagactcaggtg	20	DAB2-Ex1-R	taacctcccacagacacctg	20
CLPTM1L-Ex14-F	gcatagacgtggcagctc	18	TTC33-Ex3-F	ccgtccagcaaaatccac	18
CLPTM1L-Ex14-R	gaagacgcccttgctcac	18	TTC33-Ex3-R	acaggaggctgagacaggag	20
SLC6A3-Ex13-F	tgagggtgctggtaggtgag	20	MRPS30-Ex1-F	tacagcaagagggcaaggac	20
SLC6A3-Ex13-R	gcacacgtctccactgtcac	20	MRPS30-Ex1-R	gcttctcgtctaccgactcc	20
March6-Ex24-F	ggttggcagattcagttgtg	20	ISL1-Ex4-F	cctgctttgtgtgctgagg	19
March6-Ex24-R	gcatatggagtgagacatcc	20	ISL1-Ex4-R	gtcttctccggctgcttg	18
March6-Ex25-F	ctgccccagtccattcttc	19	DHX29-Ex14-F	aaaggtggtgtcccctgttc	20
March6-Ex25-R	ccttgctaatcccacctc	18	DHX29-Ex14-R	acttgccctatcctgagcac	20
ROPN1L-Ex3-F	cctgtcatgcatttggag	18	SLC12A7-Ex3-F	cgcaggaggagatggacag	19
ROPN1L-Ex3-R	ataataggcctggagcag	18	SLC12A7-Ex3-R	cagggcagctcttctgac	18
DROSHA-Ex5-F	cacggtgatttggcatgg	18	LIFR-Ex19-F	ccgcagatgaagctggag	18

DROSHA-Ex5-R	caggtagagcccagagatgc	20
PDZD2-Ex5-F	ccctggcctttgaggtaac	19
PDZD2-Ex5-R	gcactggtgtcttttcccatc	21
RAI14-Ex2-F	cgggagcaggctttaatttg	20
RAI14-Ex2-R	tcaacatgggagtggagagg	20
SPEF2-Ex2-F	gctgtcatcataggttgtgtg	21
SPEF2-Ex2-R	tgcactccagtctcagtaacag	22
IL7R-Ex8-F	tgtgtctctctggtgccatc	20
IL7R-Ex8-R	cccattcttgccactctcc	19
RICTOR-Ex35-F	atgctgagatgctgatcctc	20
RICTOR-Ex35-R	gtcagtgggtgatataggtg	20
DAB2-Ex13-F	atacactgccatctgcaagc	20
DAB2-Ex13-R	aaatccacatccccaagg	18
DAB2-Ex8-F	cttcctgacctaaacagc	18
DAB2-Ex8-R	acctcctccttctaatcc	18
ITGA1-Ex12-F	ggagaaagcagttgtggtag	20
ITGA1-Ex12-R	gccctgttcccagttgttac	20

LIFR-Ex19-R	ggaccaccctcctccattag	20
DHX29-Ex11-F	gcactgtccttggaatcagc	20
DHX29-Ex11-R	cttcggggttgggtacagac	20
ANKRD55-Ex2-F	ctgaggaagttgacctgacc	20
ANKRD55-Ex2-R	tagcctgctctttccgtacc	20
ANKRD55-Ex4-F	agggggagatggagagattc	20
ANKRD55-Ex4-R	gccacagagggatggattag	20
ITGA2-Ex26-F	ttcccagacccctacaagtg	20
ITGA2-Ex26-R	ggagaaaagcagcgtcctg	19
FYB-Ex2-F	gcttgaccaccagagacc	18
FYB-Ex2-R	cgcttttggacctgactcc	19
BRD9-Ex3-F	gccagggtctctgtgaggag	20
BRD9-Ex3-R	ctccacctccaccttcttcc	20
ITGA1-Ex14-F	ggcaagactataaggaaagag	21
ITGA1-Ex14-R	tatgcacactacatatacac	20

Appendix III

Buffers and reagents

Tris-acetate EDTA (TAE) buffer (50X):

- Tris base (242g)
- Glacial acetic acid (57.1ml)
- EDTA (18.61g)
- Distilled water to 1000ml
- Sterilized by autoclaving

Tris-EDTA (TE) buffer:

- 1M Tris-HCl, pH 8.0 (1ml; final concentration = 10mM)
- 0.5M EDTA (200µl; final concentration = 1mM)
- Distilled water to 100ml
- Sterilized by autoclaving

6x DNA loading buffer

- 0.2% Bromophenol blue
- 0.2% Xylene Cyanol
- 60% Glycerol

References

Alfradique I, Vasconcelos MM (2007) Juvenile myoclonic epilepsy. *Arq Neuropsiquiatr* 65:1266-1271.

Asconapé J, Penry JK (1984) Some clinical and EEG aspects of benign juvenile myoclonic epilepsy. *Epilepsia* 25:108-114.

Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, Boas WE, Engel J, French J, Glauser TA, Mathern GW, Moshe SL, Nordli D, Plouin P, and Scheffer IE (2010) Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia* 51:676-685.

Bozzi Y, Casarosa S, and Caleo M (2012) Epilepsy as a neurodevelopmental disorder. *Frontiers in Psychiatry* 3:19.

Christoph SI, Harald H, Sergei VS, Rolf S (2013) Desminopathies: pathology and mechanisms. *Acta Neuropathol* 125:47-75.

Clemen CS, Herrmann H, Strelkov SV and Schroder R (2013) Desminopathies: pathology and mechanisms. Acta Neuropathol 125:47–75.

Commission on Classification and Terminology of the International League Against Epilepsy (1981) Proposal for revised clinical and electrographic classification of epileptic seizures. *Epilepsia* 22:489-501.

Commission on Classification and Terminology of the International League Against Epilepsy (1989) Proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia* 30:389-399.

Cossette P, Liu L and Brisebois K (2002) Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. *Nat Genet* 31:184-189.

Delgado-Escueta AV, Enrile-Bacsal F (1984) Juvenile myoclonic epilepsy of Janz. *Neurology* 34:285–294.

Delgado-Escueta AV, Koeleman PC, Bailey JN, Medina MT, Durón RM (2013) The quest for Juvenile Myoclonic Epilepsy genes. *Epilepsy Behav* 28:52-57.

Dibbens LM, Feng HJ and Richards MC (2004) GABRD encoding a protein for extra- or peri-synaptic GABAA receptors is a susceptibility locus for generalized epilepsies. *Hum Mol Genet* 13:1315–1319.

Eadie MJ (2002) The epileptology of Théodore Herpin (1799-1865). *Epilepsia* 43:1256-1261.

Elmslie FV, Rees M, Williamson MP, Kerr M, Kjeldsen MJ, Pang KA, Sundqvist A, Friis ML, Chadwick D, Richens A, Covanis A, Santos M, Arzimanoglou A, Panayiotopoulos CP, Curtis D, Whitehouse WP, Gardiner RM (1997) Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. *Hum Molec Genet* 6:1329-1334.

Escayg A, De Waard M and Lee DD (2000) Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. *Am J Hum Genet* 66:1531–1539.

Greenberg DA, Delgado-Escueta AV, Widelitz H, Sparkes RS, Treiman L, Maldonado HM, Park MS, Terasaki PI (1988) Juvenile myoclonic epilepsy (JME) may be linked to the BF and HLA loci on human chromosome 6. *Am J Med Genet* 31:185-192.

Jallon P, Latour P (2005) Epidemiology of idiopathic generalized epilepsies. *Epilepsia* 46:10-4.

Janz D, Christian W (1957) Impulsive petit mal. *Dtsch Z Nervenheilkd* 176:346-386.

Kapoor A, Ratnapriya R, Kuruttukulam G, Anand A (2007) A novel genetic locus for juvenile myoclonic epilepsy at chromosome 5q12-q14. *Hum Genet* 121:655-662.

Lazarides E, Hubbard BD (1976) Immunological characterization of the subunit of the 100 A filaments from muscle cells. *Proc. Nati. Acad. Sci.* 73:4344-43481.

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25:2078-9.

Magiorkinis E, Sidiropoulou K and Diamantis A (2011) Hallmarks in the History of Epilepsy: From Antiquity Till the Twentieth Century. Novel Aspects on Epilepsy; Prof. Humberto Foyaca-Sibat (Ed.), ISBN: 978-953-307-678-2, InTech, Available from: http://www.intechopen.com/books/novel-aspects-onepilepsy/

Montalenti E, Daniele I, Andrea R, Bruno B, Paolo B (2001) Clinical features, EEG findings and diagnostic pitfalls in juvenile myoclonic epilepsy: a series of 63 patients. *J Neurol Sci* 184:65-70.

Moschetta S, Fiore LA, Fuentes D, Gois J, Valente KD (2011) Personality traits in Patients with juvenile myoclonic epilepsy. *Epilepsy Behav* 21:473-477.

Ratnapriya R, Vijai , Kadandale JS, Iyer RS, Radhakrishnan K and Anand A (2010) A locus for juvenile myoclonic epilepsy maps at 2q33-q36. *Hum Genet* 128:123-130.

Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, Nishida M, Numata T, Medina MT, Takeuchi T, Morita R, Bai D, Ganesh S, Sugimoto Y, Inazawa J, Bailey JN, Ochoa A, Jara-Prado A, Rasmussen A, Ramos-Peek J, Cordova S, Rubio-Donnadieu F, Inoue Y, Osawa M, Kaneko S, Oguni H, Mori Y and Yamakawa K (2004) Mutations in EFHC1 cause juvenile myoclonic epilepsy. *Nat Genet* 36:842–849.

Vijay J, Cherian PJ, Sylaja PN, Anand A, Radhakrishnan K (2003) Clinical characteristics of a South Indian cohort of juvenile myoclonic epilepsy probands. *Seizure* 12:490-496.

Websites used for bioinformatics analysis

1000genomes – http://www.browser.1000genomes.org/index.html

Ensembl – http://www.asia.ensembl.org/index.html

Entrez gene, SNP and Genbank database: http://www.ncbi.nlm.nih.gov/

Exome Variant Server – http://www.evs.gs.washington.edu/EVS/

NCBI Mapviewer – http://www.ncbi.nlm.nih.gov/mapview/

OligoCalc: http://www.basic.northwestern.edu/biotools/OligoCalc.html

Primer3: http://frodo.wi.mit.edu/

Clustal omega: http://www.ebi.ac.uk/Tools/msa/clustalo/

SIFT: http://sift.jcvi.org/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/

Mutation taster: http://www.mutationtaster.org/

Berkeley Drosophila Genome Project - Splice Site Prediction: http://www.fruitfly.org/seq_tools/splice.html

Human Splicing Finder v2.4.1: http:// www.umd.be/HSF/

Softwares used for data analysis

BWA v-0.6.0

SeqQC v-2.0

SAMtools v-0.1.7a

SeqMan™ II – DNA Star